Research on Facial Expression Recognition

Zijie Sheng

The Chinese University of Hong Kong, Shenzhen, China

Keywords: facial expression recognition; feature extraction; facial expression features.

Abstract: With the rapid development of machine learning and artificial intelligence and the widespread popularity of intelligent devices, facial recognition technology is experiencing unprecedented development. Among this recent growth, facial expression recognition (FER) has been a hot research topic and has received extensive attention by various fields over recent years. This paper introduces the basic steps of face recognition in detail, briefly reviews the research progress of scholars at home and abroad, and compares their research progress and methods. Through the above comparison and analysis, we can uncover the current research difficulties and discuss them in depth. Finally, trends in the future development of facial expression recognition are addressed, which can provide great reference material for researchers attempting to master facial expression recognition technology.

1. Introduction

In daily life, facial expressions play a very important role in communication as they convey important non-verbal information during human communication. In some ways, facial expressions are more effective than verbal communication. With the rapid development of computer technology, facial expression recognition research has become very active. Nowadays, facial expression recognition refers to a technology that uses computers to automatically analyze facial expression changes; that is, to give computers the human cognitive thinking to analyze human emotions, to create better levels of human-computer interaction. This paper summarizes recent research on facial expression recognition, summarizes the improvement of classical algorithms in this field by scholars at home and abroad, and analyses the advantages and disadvantages of these methods by comparing them to other methods.

2. Overview of Facial Expression Recognition System

Facial expression recognition (FER) system includes three main parts: image preprocessing, expression feature extraction, and expression classification. In order to build a FER system, image preprocessing is first needed, that is, preprocessing the collected samples to reduce the influence of sample noise on the expression recognition. This process mainly involves image illumination compensation, scale normalization, and face location detection. The purpose of facial feature extraction is to extract feature information that can express the essence of facial expression from preprocessed images. At the same time, in order to avoid the high number of image cones and affect the operation of the algorithm, feature dimension reduction and feature decomposition are also needed. Facial expression classification is used to analyze the relationship between extracted features and classify the input facial expression images into corresponding categories. The general process of facial expression recognition is shown in Figure 1.

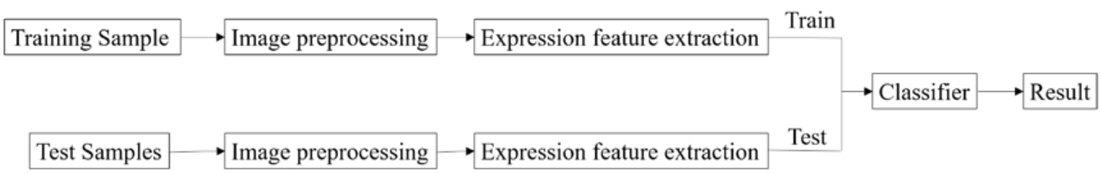

Fig. 1 General process of facial expression recognition

The purpose of image preprocessing is to eliminate the interference caused by many objective conditions including the intensity of illumination, to avoid the influence of these interference factors on image extraction and classification, and to ensure the speed and accuracy of recognition. The main purpose of feature extraction is to obtain attribute information that can represent image categories. Classification and recognition are a process based on the feature information of the input image. Since image preprocessing involved in the research of facial expression recognition is already a relatively mature technology, this paper will not introduce it in detail in this respect.

3. Facial Feature Extraction

Facial expression feature extraction is the most important part of a facial expression recognition system. Effective facial expression feature extraction can effectively improve the performance of a system's recognition ability. In facial expression feature extraction, in order to avoid the high dimensions of the image and affect the operation of the algorithm, it is generally necessary to reduce the dimensions of features and decompose the features.

3.1 Extraction of Primitive Expression Features

Facial expression feature extraction is mainly based on a mathematical method, which relies on the computer to organize and process facial expression image data, remove non-expressive noise, and extract features that can represent facial expressions. According to the different states of an expression, the image can be divided into static images or a dynamic image sequence. Because the feature extraction methods of these two images are different, the expression feature extraction algorithms can be divided into static image feature extraction and dynamic image sequence feature extraction algorithms. Next, we will introduce several classical feature extraction algorithms, and summarize the improvements made by scholars at home and abroad on the basis of these algorithms and their innovative applications.

3.1.1 Static Image Expression Feature Extraction

Static images show the expression state of a single image when facial expressions occur, and intuitively show the facial deformation and texture features generated by facial muscle movement when facial expressions occur. The commonly used methods of static image expression feature extraction include texture feature extraction and geometric feature extraction.

(1) Texture Feature Extraction

Texture feature extraction refers to the use of the underlying information of facial expression images, highlighting the changes of local expression information. Representative methods of texture feature extraction are Local Binary Pattern (LBP) and Gabor Wavelet Transform.

Local binary pattern is a method of comparing local features, which can represent local texture features by local grayscale changes in images. Ojala et al. proposed the most primitive LBP operator, and then a large number of improved algorithms [1]. The advantage of an LBP operator is that it can effectively overcome the problems of image displacement, rotation, and uneven illumination. It has fast operation speed and can comprehensively describe the texture information of the image. However, LBP is sensitive to noise, and only considers the difference symbols between the central and neighboring pixels. It does not consider the difference magnitude across larger parts of the data information. In order to extract LBP features more fully, Guo et al. proposed a complete local binary pattern (CLBP) algorithm [2]. The features extracted by CLBP are comprehensive and have strong discriminant ability. But while CLBP extracts more comprehensive information, it inevitably also extracts interference information, which not only increases the features dimension, but also increases the time complexity, so the recognition rate is affected to a certain extent. In order to solve this problem, Guo et al. proposed a Fisher criterion-based local binary pattern (FCL-LBP) algorithm and a discriminant full local binary pattern (disCLBP) algorithm [3]. By maximizing the class spacing and minimizing the class spacing, Guo et al. screened out the more robust texture features. Zhou et al. improved the disCLBP algorithm to screen out the unique features for each expression, to obtain more discriminative features than common features, which is more conducive to facial expression recognition [4].

The Gabor function was formally proposed by Dr. Dennis Gabor in 1946, and then widely used in signal analysis and image processing. The advantage of the Gabor wavelet is that it can detect multi-scale and multi-direction texture changes, and is insensitive to light changes. It has good adaptability and can describe the texture features of images well. However, Gabor filter banks usually extract the features of the whole image, which cannot obtain the finer local features, and the features are also redundant. To solve this problem, Liu et al. proposed two Gabor multi-directional feature fusion rules to fuse features of the same scale and different directions together [6]. Zhou et al. proposed a feature extraction method based on local Gabor features [7], which is used to extract local Gabor features for radial coding.

(2) Geometric Feature Extraction

The geometric feature extraction method is mainly used to extract facial feature points, through the analysis of the geometric relationships between feature points. There are three commonly used methods in geometric feature extraction.

Active Shape Models (ASM) is a feature extraction method based on statistical models proposed by Cootes et al. [8] in 1995. It is mainly used to extract feature points of facial contours. This model mainly uses global shape models to match the initial shape of the face, and then establishes a local texture model to obtain the contour features of the target more accurately. If the initial matching facial model has a large displacement, the accuracy of the result will be affected using a local texture model because of illumination, background noise, and other reasons. Peng et al. proposed an improved ASM algorithm based on the local texture model [9] to solve the local texture problem of ASM. Based on the ASM algorithm, Hou Jie extracts and locates 118 feature points from eyebrows, eyes, nose, and mouth to form contour shape data for each part, thus realizing face feature point location [10].

Active Appearance Models (AAM) [11] are evolved from the basis of ASM. The model not only considers global shape information, but also local texture feature information. The AAM algorithm uses both shape and texture information to build the appearance model, and then fits the model to the target, which can accurately represent the shape and texture changes of the target image. Saatci et al. proposed a gender and expression recognition method based on the active appearances model. The method uses an AAM cascade SVM classifier and achieved a high recognition rate [12].

Scale-invariant Feature Transform (SIFT) is a new image local feature descriptor [13], which was proposed by David Lowe in 1999 and perfected in 2004[14]. Berretti et al. proposed using the SIFT descriptor to describe the depth and texture of an image, and then using SVM to classify [15].

In order to obtain better extraction results, some researchers use a combination of multiple feature extraction methods to extract facial features. Cai Leyi uses Gabor and SIFT feature extraction to improve the accuracy of matching, avoiding the shortcomings of local convergence, but also reducing the speed of algorithm [16].

Huang et al. proposed a new method for facial expression recognition based on AM-SIFT and adaptive region weighting [17]. This method not only improves the recognition rate of positive facial expressions, but also has good robustness to non-positive facial expressions.

3.1.2 Feature Extraction from Dynamic Image Sequences

Different from static images, dynamic image sequencing reflects the continuous process of facial expression movement, mainly using temporal and spatial information, combined with motion changes to recognize facial expressions. The main methods of expression feature extraction in dynamic image sequences are the optical flow method, model method, and feature point tracking method.

(1) Optical Flow Method

The optical flow method reflects the important information of inter-frame motion in dynamic image sequence. In facial expression recognition, the optical flow method is often used to extract the expression features of dynamic images. Yacoob and Davis used the optical flow field and gradient field between continuous frames to represent the temporal and spatial changes of the image [18]. They tracked the facial expression area in the dynamic image sequence, and then analyzed the changes of facial expressions' spatial characteristics and the corresponding movement of facial muscles, so as to classify facial expressions.

The advantage of this algorithm is that it can highlight the deformation of facial muscles and reflect trends in facial movements. However, the disadvantages include a vulnerability to the influence of uneven illumination, which leads to a reduction in accuracy of feature extraction results. To solve this problem, Brox et al. used the uniform brightness and gradient model [19] to pre-process the data, and Papenberg et al. used Hessian matrixes to counteract the change of illumination [20].

(2) Model Method

The model method in facial expression recognition refers to the statistical method for developing a parameterized description of the expression information of a dynamic image sequence. The feature extraction method based on a model that can extract the geometric deformation and texture changes of the human face, but the establishment of the model cannot be completed automatically and needs manual assistance. In addition, the algorithm is computationally complex. He Jun et al. proposed a new unsupervised feature extraction and recognition method based on Deep Belief Nets [21]. Gou Jiali proposed a local principal texture pattern (LPTP) [22], which combines LBP and LDP as a single feature. This method can make up for the deficiencies of LBP and LDP which are sensitive to light and noise, but it only considers spatial feature changes and ignores temporal changes.

(3) Characteristic Point Tracking Method

The tracking method based on feature points usually chooses feature points in areas where the grayscale values of the image change greatly, such as the corners of the mouth, the corners of the eyes and so on. The placement of feature points in these regions is tracked to obtain the displacement and deformation information of facial features. Tie et al. proposed a method for automatically extracting 26 feature points from face models in video sequences and tracking them through multiple particle filters [23]. Tslakanidou et al. proposed a 3D facial feature tracking method based on ASM [24]. This method tracks 81 feature points of the face and realizes the recognition of some composite motion units.

3.2 Feature Dimension Reduction Method

After extracting the original features, facial expression images often have large spatial dimensions. It is necessary to reduce the dimensions of these features. This greatly reduces the operation times of the algorithms, and improves classification and recognition. The main task of feature dimension reduction is to optimize the dimensional reduction of extracted features without affecting the information gained from the original data. Next, we will introduce some classical feature dimension reduction methods and the improvements made by scholars in this field over recent years.

Principal Component Analysis (PCA) is a common feature dimension reduction method. This method can effectively extract regions that contribute greatly to expression recognition in images, and can effectively eliminate redundancy and reduce input dimensions. However, it does not take into account the distinction between different types of data, and the benefits of PCA decrease significantly when the dimensions are too high, Ying et al. proposed a discriminant analysis algorithm based on support vector machine [25]. This algorithm is based on Fisher and SVM. It can produce have the greatest class separation in facial expression data with a limited number of samples, and then reduce the dimensionality, which overcomes the limitation of PCA algorithm. In contrast, the PCA algorithm only uses second-order statistics of samples, and does not take into account higher-order statistics. Radulovic et al. proposed Independent Component Analysis (ICA). ICA can be regarded as a generalization of PCA [26]. It not only uses second-order statistical analysis, but also uses higher-order statistical analysis.

Linear Discriminant Analysis (LDA), also known as Fisher Linear Discriminant (FLD), is a classical algorithm for pattern recognition. In 1996, Belhuneur introduced pattern recognition through artificial intelligence. The LDA algorithm is an effective feature extraction method, which can maximize the inter-class and intra-class scatter matrices of samples. Wang et al. proposed an improved LDA algorithm, Local Linear Discriminant Analysis (LLDA) [27]. This method can overcome the limitations of traditional LDA and reduces the complexity of the algorithm. Experiments on JAFFE and CK show that it is more effective than LDA.

Locally Linear Embedding (LLE) is an unsupervised dimension reduction method for non-linear data proposed by Mclvor et al. It can reduce the dimensionality of data while retaining the original data's topological structure [28]. Zhao et al. proposed a new LLE-based supervised manifold learning algorithm [29], called Discriminant Kernel Locally Linear Embedding (DKLLE). Experiments on the CK dataset show that this algorithm is superior to the LLE algorithm.

3.3 Eigen Decomposition Method

Facial expression images contain abundant facial expression information. Facial expression recognition is to distinguish different expressions by using different information extracted from different facial features. Favorable information for one expression's recognition may interfere with another expression's recognition. One solution is to decompose the facial image into several subspaces for analysis and processing. M.A.O. Vasilescu et al. proposed a high order singular value decomposition (HOSVD) face recognition algorithm [30], and Wang et al. improved the algorithm [31]. HOSVD decomposes the facial images into a personal subspace and an expression subspace using third-order tensors, to realize facial expression recognition and synthesis. Zhu et al. proposed a popular expression classification algorithm [32], which is based on the popular learning algorithm LPP (Locality Preserving Projections) and HOSVD. Combining the feature extraction ability of the LPP algorithm with the pattern decomposition ability of HOSVD, the advantages of both are complementary. LPP solves the problem of extracting image feature values for HOSVD, while HOSVD solves the problem of interference between personal patterns and expression patterns when preforming expression recognition.

4. Expression Classification

Facial expression classification is an indispensable step in the process of facial expression recognition. Its purpose is to determine the category of facial features that the test samples fall into. Commonly used expression classification methods include: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN), and AdaBoost with Hidden Markov Models (HMM).

The purpose of SVMs are to find the optimal classification surface under limited samples and maximize the class spacing. However, in the case of a large number of samples, the computation matrix uses a lot of memory, which will affect the efficiency of the algorithm. SVMs have many advantages in solving non-linear and high-dimensional data problems with a small number of samples [33]. The KNN algorithm is a mature classification method, an extension of the nearest neighbor algorithm, and one of the commonly used machine learning methods. ANN is a mathematical model based on the human brain neural network in terms of information processing, and forms different networks according to different connection modes. Adaboost classification is an iterative algorithm. The idea behind it is to train a series of weak classifiers on the training sample, and then combine these weak classifiers to form a strong classifier [34]. HMM is a kind of Markov chain. Its state cannot be observed directly, but it can be observed through the observation vector sequence. Each observation vector shows various states through some probability density distribution, and each observation vector is generated by a state sequence.

Each classification method has some shortcomings, but these classification methods have been improved by recent research. Zhan et al. proposed the Fuzzy Deep Hidden Markov algorithm [35]. This algorithm reduces the conditional independence assumption of the classical HMM algorithm and improves the recognition rate and robustness of the algorithm. Wang Jianyun et al. proposed a method of expression recognition based on local parallel depth neural networks [36]. They planned and trained several parallel networks and obtained better expression classification and discrimination.

In addition, in order to achieve better classification and recognition, some researchers use a combination of various classification methods to classify facial features. Xu Wenhui et al. combined KNN with SVM, integrating nearest neighbor information into SVM [37], and defined a local SVM classifier. Wang et al. proposed a new expression recognition algorithm based on Fuzzy Support Vector Machines (FSVM) and KNN, which divides the input space into different regions and combines the characteristics of FSVM and KNN. Different classification methods are used according to their strengths, which achieves results in good recognition accuracy and effectively reduces the computational complexity.

In conclusion, through the introduction of the basic knowledge making up facial expression recognition, recent research achievements, and the latest technologies, we can see that the research of facial expression recognition has made a lot of progress, but still has a long way to go when compared with human-level recognition, meaning it has not met the requirements needed for practical application. At present, it seems that advancements in the following areas of research will greatly improve the performance of a FER system.

(1) Face detection is a necessary part of FER systems, but, at present, face detection still has great limitations, especially when confounded by occlusion and complex backgrounds.

(2) Micro-expressions are expressions that human beings inadvertently reveal, which can truly reflect their inner emotional changes. They have important applications in psychology, such as in lie detection technology. However, due to the short occurrence time of micro-expressions and the insignificant changes in features, it is necessary to classify expressions more accurately. Research into micro-expressions should also become a focus of domestic and foreign scholars.

(3) At present, facial expression recognition is mainly studied using two-dimensional images, but two-dimensional images are limited. When compared with two-dimensional images, the amount of information in 3D images is significantly greater than that in two-dimensional images, so research into 3D expression recognition could prove to be more promising.

References

[1] Ojala T, Pietikainen M, Harwood D.A comparative study of texture measures with classification based on featured distributions [J]. Pattern Recognition,1996,29(1):51-59.

[2] Guo Z, Zhang D.A completed modeling of local binary pattern operator for texture classification [J]. IEEE Transactions on Image Processing,2010,19(6):1657-1663.

[3] Guo Y, Zhao G, Pietikainen M. Discriminative features for texture description [J]. Pattern Recognition,2012,45(10):3834-3843.

[4] Zhou Yuxuan, Wu Qin, Liang Jiuzhen, et al. Discriminant Complete Local Binary Pattern Face Expression Recognition [J/OL]. Computer Engineering and Application.

[5] Gabor D. Theory of communications. IEEE Journal of the Institution of Electrical Engineers-Part II[J]. Radio and Communication Engineers,1946,93(26):429-457.

[6] Liu Shuai-shi, Tian Yan-tao, Wan Chuan. Facial Expression Recognition Method Based on Gabor Multi-orientation Features Fusion and Block Histogram [J]. Acta Automatica Sinica, 2011, 37(12): 1455-1463.

[7] Zhong Sizhi. Research on Face Expression Recognition Algorithms [J]. Shanghai: East China Normal University, 2015.

[8] Cootes T, Taylor C, Cooper D, et al. Active shape model-their training and application [J]. Computer Vision and image Understanding,1995,61(1):38-59.

[9] Peng Cheng, Liu Shuai-shi, Wan Chuan. An active shape model for facial expression recognition based on a local texture model [J]. CAA1 Transactions on Intelligent Systems,2011,6(3).

[10] Hou Jie. Research on Face Expression Computing Technology [D]. Suzhou: Suzhou University, 2014.

[11] Edwards G, Taylor C, Cootes T. Interpreting face images using active appearance models [C]//Automatic Face and (Gesture Recognition.1998:300-305.

[12] Saatci Y. Town C. Cascaded Classification of Gender and Facial Expression using Active Appearance Models[C]//Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition.2006:393-398.

[13] Lowe DG. Object recognition from local scale-invariant features [C]//International Conference on Computer Vision. Greeee: Kerkyra,1999:1150-1157.

[14] Lowe D G. Distinctive image features from scale-invariant key points[J]. International journal of Computer Vision,2004,60(2):91-110.

[15] Berretti S, Bimbo A D, Pala P, et al. A set of selected SIFT features for 3D facial expression recognition[C]//20th International Conference on Pattern Recognition(ICPR).2010:4125-4128.

[16] Cai Leyi. Research on Face Recognition Algorithms Based on Active Shape Model [D]. Hangzhou: Zhejiang University of Technology, 2013.

[17] Ren Fu-ji, Huang Zhong. Facial Expression Recognition Based on AAM-SIFT and Adaptive Regional Weighting [J]. IEEJ Transactions on Electrical and Electronic Engineering,2015,10:713-722.

[18] Yacoob Y, Davis I. Recognizing Human Facial Expressions from Long Image Sequences Using Optical Flow [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1996,18 (6):636-642.

[19] Brox T, Bruhn A, Papenberg N, et al. High Accuracy Optical Flow Estimation Based on a Theory for Warping[M]. Berlin: Spring Berlin Heidelberg,2004:25-36.

[20] Papenberg N, Bruhn A, Brox T, et al. Highly accurate optic flow computation with theoretically justified warping [J]. Inter. J. Computer Vision,2006,67(2):141-158.

[21] He Jun, Cai Jianfeng, Fang Ganoderma lucidum, et al. Face expression recognition based on LBP/VAR and DBN model [J]. Computer application research, 2016, 33 (8): 2509-2513.

[22] Gou Jiali. Research on Video-based facial expression recognition technology [D]. Chengdu: Southwest Jiaotong University, 2014.

[23] Tie Y, (Guan I.A deformable 3D facial expression model for dynamic human emotional state recognition[J]. IEEE Transactions on Circuits and Systems for Video Technoloy,2013,23(1):142-157.

[24] Tslakanidou F, Malassiotis S. Real-time 2D+3D facial action and expression recognition [J]. Pattern Recognition,2010,43(6):1763-177.

[25] Ying Z-lu, Tang Jing-hai, Li Jing-wen, et al. Support vector discriminant analysis and its application to facial expression recognition[J]. Acta Electronica Sinica,2008,36(4):725-730.

[26] Radulovic J, Rankovic V. Feedforward neural network and adaptive network-based fuzzy inference system in study of powerlines[J]. Expert Systems with Applications,2010,37(1):165-70.

[27] Wang Zhan, Ruan Qiu-qi, An Guo-yun. Facial expression recognition based on tensor local linear discriminant analysis[C]//IEEE 11h International Conference on Signal Processing. Bei-jing, 2012: 1226-1229.

[28] Mclvor RT, Humphreys P K.A case-based reasoning approach to the make or buy decision[J]. Integrated Manufacturing Systems,2000(5):295-310.

[29] Zhao X, Zhang S. Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding[J]. EURASIP Journal on Advances in Signal Processing,2012(1):1-9.

[30] Vasilescu M A O, Terzopoulos D. Multilinear analysis of image ensembles: tensor faces[J]. ECCV,2002(1):447-460.

[31] Wang H, Ahuja N. Facial expression decomposition[C]//International Conference on Computer Vision.2003:958-965.

[32] Zhu Ming-han, Luo Da-yong. Expression decomposition algo-rithm based on appearance manifolds[J]. Computer Engineering and Applications,2008,44(23):203-205.

[33] Cortes C, Vapnik V. Support-Vector Networks [J]. Machine Learning,1995,20(3):273-297.

[34] Wang Wencheng. Facial expression recognition based on local feature analysis [D]. Jinan: Shandong University: 2011.

[35] Zhan Y Z, Cheng K Y, Chen Y B, et al. A New Classifier for FaCial Expression Recognition: Fuzzy Buried Markov Model [J].Journal of Computer Science and Technology, 2010, 25 (3): 641-Six hundred and fifty.

[36] Wang Jianyun, Li Xiaoxia. An expression recognition method based on in-depth learning [J]. Computer and Modernization, 2015 (1): 84-87.

[37] Xu Wenhui, Sun Zhengxing. LSVM algorithm for expression classification of video sequences [J]. Journal of Computer Aided Design and Graphics, 2009, 21(4): 542-548.

[38] Wang Xiao-hu, Liu An, Zhang Sh-qing. New facial expression recognition based on FSVM and KNN [J]. Optik, 2015 (126):3132-3134.