

人脸表情识别研究进展

黄 建 李文书 高玉娟
(浙江理工大学信息学院 杭州 310018)

摘 要 人脸表情识别(Facial Expression Recognition, FER)是计算机视觉、机器学习、人工智能等领域的重要研究方向,目前已经成为国内外学者的研究热点。介绍了 FER 系统流程,总结了表情特征提取和表情分类的常用方法以及近年来国内外学者对这些方法的改进,并对这些方法的优缺点进行比较。最后,对目前 FER 研究的难点问题进行了分析,并对 FER 未来的发展方向进行展望。

关键词 人脸表情识别,表情特征提取,表情分类

中图法分类号 TP391.4 文献标识码 A

Research Advance of Facial Expression Recognition

HUANG Jian LI Wen-shu GAO Yu-juan

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract Facial expression recognition (FER) is an active research topic in the fields of computer vision, machine learning and artificial intelligence, and has become a hot research. In this paper, we introduced the FER system processes, summarized the common method of facial expression feature extraction and facial expression classification, as well as the improvement of these methods proposed by domestic and foreign scholars in recent years, and compared the advantages and disadvantages of these methods. Finally, the current difficult problems of FER studies were analyzed, and the future development direction of FER was presented.

Keywords Facial expression recognition, Facial expression feature extraction, Facial expression classification

1 引言

在日常生活中,面部表情在交流中起到了非常重要的作用,是人类交流中一种重要的非语言信息,不仅传达了情绪,也传递了交流者丰富的情感信息,在某些方面比语言交流更加有效。随着计算机技术的快速发展,人脸表情识别技术的研究已经变得非常活跃。现如今,人脸表情识别是指使用计算机自动分析人脸面部表情变化的一种技术,即赋予计算机人类认知思维去分析人类的情绪,从而创建一个更好的人机交互的环境。本文主要对近年来有关人脸表情识别的研究进行总结,归纳国内外学者对该领域内经典算法的改进,并分析其方法的优缺点以及与其他方法的比较。

2 人脸表情识别系统概述

人脸表情识别(FER)系统主要包括 3 个部分:图像预处理、表情特征提取和表情分类。建立 FER 系统,首先需要进行图像预处理,即对采集的样本进行预处理操作,减小样本噪声对表情识别的影响,该过程主要涉及到图像的光照补偿、尺度归一化及人脸检测定位等操作。表情特征提取的目的是从经过预处理的图像中提取出能够表示表情本质的特征信息。同时,为了避免图像维数过高,影响算法的运行,还需要特征降维、特征分解等操作。表情分类是分析提取出的特征之间的关系,将输入的人脸表情图像分类到相应的类别中。人脸

表情识别的一般过程如图 1 所示。

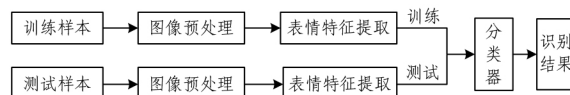


图 1 人脸表情识别的一般过程

图像预处理是为了消除一些光照强弱等多种客观条件造成的干扰,避免这些干扰因素对图像提取和分类识别造成影响,同时也是为了确保识别的速度和准确率;特征提取主要是为了获取能够表征图像类别的属性信息;分类识别是一个根据输入图像的特征信息进行类别判别的过程。由于人脸表情识别研究中涉及到的图像预处理一般都采用比较成熟的技术,因此本文在这方面不再详细介绍。

3 表情特征提取

表情特征提取是面部表情识别系统中最重要的一部分,有效表情特征提取能够有效提高系统识别的性能。在人脸表情特征提取中,为了避免图像维数过高从而影响算法的运行,一般还需要特征降维、特征分解等操作。

3.1 原始表情特征提取

表情特征提取主要是采用数学方法,依靠计算机对人脸表情图像进行数据的组织和处理,去除非表情的噪声,并提取能够表征人脸表情的特征。根据表情发生时的状态的不同,

可以将图像分为静态图像和动态图像序列,因这两种图像的特征提取方法不一样,所以可以将表情特征提取算法分为静态图像的特征提取算法和动态图像序列的特征提取算法。下面将介绍几种经典的特征提取算法,并总结国内外学者在这些算法基础上进行的改进和这些算法的创新应用。

3.1.1 静态图像的表情特征提取

静态图像呈现的是表情发生时单幅图像的表情状态,直观地显示了表情发生时人脸肌肉运动所产生的面部形变和纹理特征。常用的静态图像表情特征提取方法有纹理特征提取和几何特征提取。

(1) 纹理特征提取

纹理特征是指使用人脸面部图像的所有特征,反映了人脸表情图像底层的信息,突出体现了局部表情信息的变化。纹理特征提取方法中具有代表性的方法有局部二值模式(Local Binary Pattern, LBP)、Gabor 小波变换。

局部二值模式是一种进行局部特征比较的方法,通过图像的局部灰度变化关系来表征局部纹理特征。Ojala 等人提出最原始的 LBP 算子,之后出现了大量的改进算法^[1]。LBP 算子的优点是能有效克服图像位移、旋转及光照不均等问题,运算速度快,能够有效描述图像的纹理信息。但 LBP 对噪声敏感,且只考虑了中心像素和邻域像素的差值符号特征,没有考虑差值幅度,丢失了部分数据信息。为了使 LBP 特征提取更充分,Guo 等人提出了完全局部二值模式(CLBP)算法^[2],CLBP 提取的特征比较全面且具有较强的鉴别能力。但 CLBP 在提取比较全面的信息时,不可避免地提取了干扰信息,这不仅增加了特征维数,也增加了时间复杂度,因而识别率受到一定的影响。针对这一问题,Guo 等人又提出了基于 Fisher 准则的局部二值模式(FCL-LBP)算法和具有判别力的完全局部二值模式(disCLBP)算法^[3],其通过使类间距最大、类内距最小,筛选出更具有鲁棒性的纹理特征。Zhou 等人改进了 disCLBP 算法,即针对每种表情分别筛选出其特有的特征,从而得到比共同特征更具有辨别性的表情特征,从而更利于表情识别^[4]。

Gabor 函数由 Dennis Gabor 博士在 1946 年正式提出,随后被广泛应用于信号分析和图像处理领域。Gabor 小波的优点是能够检测多尺度、多方向的纹理变化,而且对光照变化不敏感,具有较好的适应性,可以很好地描述图像的纹理特征^[5]。但是 Gabor 滤波器组通常提取整张图像的特征,无法获取局部细微的特征,特征也存在冗余。针对这一问题,Liu 等人提出了两个 Gabor 多方向特征融合规则,以将相同尺度不同方向的特征融合到一起^[6]。Zhong 等人提出一种基于局部 Gabor 特征的特征提取方法^[7],用以提取局部 Gabor 特征进行径向编码。

(2) 几何特征提取

几何特征的提取方法主要是提取人脸面部的特征点,通过分析特征点之间的几何关系,实现人脸表情的判别。几何特征提取方法中较为常用的方法有以下 3 种。

主动形状模型(Active Shape Models, ASM)是由 Cootes 等人^[8]于 1995 年提出的一种基于统计模型的特征提取方法,主要用于提取表情轮廓的特征点。这一模型主要利用全局形状模型来匹配人脸初始形状,再建立局部纹理模型,从而较为精确地获取目标的轮廓特征。如果初始匹配的人脸模型位置偏移较大,利用局部纹理模型会因为光照、背景等原因而使最

终结果的准确性受到影响。Peng 等针对 ASM 局部纹理问题,提出基于局部纹理模型的改进 ASM 算法^[9]。侯婕基于 ASM 算法,针对人脸眉毛、眼睛、鼻子和嘴巴部分提取定位出 118 个特征点形成各部位的轮廓形状数据信息,从而实现人脸特征点定位^[10]。

主动外观模型(Active Appearance Models, AAM)^[11]是在 ASM 的基础上演变来的,该模型不仅考虑全局形状信息,而且考虑了局部纹理特征信息。AAM 算法同时利用形状和纹理信息建立外观模型,然后再利用该模型对目标进行拟合,能够准确地生成形状及纹理变化的目标图像。Saatci 等提出基于主动外观模型的性别和表情识别方法其采用 AAM 级联 SVM 分类器的方法得到了很高的识别率^[12]。

尺度不变特征变换(Scale-invariant Feature Transform, SIFT)是由 David Lowe 于 1999 年提出^[13]并在 2004 年得以完善的一种新的图像局部特征描述算子^[14]。Berretti 等人提出利用 SIFT 描绘子来描述图像的深度及纹理,然后再利用 SVM 进行分类^[15]。

为了获得更好的提取效果,一些研究者们采用多种特征提取方法相结合的形式来对表情特征进行提取。蔡乐毅使用 Gabor 特征和 SIFT 特征的提取来提高匹配的准确性,避免了局部收敛的缺点,但同时也降低了算法的运算速度^[16]。Huang 等人提出了基于 AAM-SIFT 和自适应区域加权的面部表情识别的新方法^[17],该方法不仅提高了正面人脸表情的识别率,对非正面的面部表情也有较好的鲁棒性。

3.1.2 动态图像序列的表情特征提取

与静态图像不同,动态图像序列反映的是人脸面部表情运动的连续过程,主要是利用时空信息,结合运动变化来识别表情。动态图像序列表情特征提取的主要方法有光流法、模型法和特征点跟踪法。

(1) 光流法

光流法反映了动态图像序列中帧间运动的重要信息,在人脸表情识别中经常采用光流法提取动态图像的表情特征。Yacoob 和 Davis 等人使用连续帧之间的光流场和梯度场来表示图像的时空变化^[18],实现动态图像序列中人脸表情区域的跟踪,再分析人脸表情区域特征的变化,以及对人脸肌肉的运动,从而对人脸面部表情进行分类。

该算法的优点是能够突出人脸形变并反映人脸运动趋势,而缺点也比较明显,即容易受到光照不均的影响,从而导致特征提取结果的准确性降低。针对这一问题,Brox 等人在构建数据项时同时采用亮度和梯度一致模型^[19],Papenberg 等人利用 Hessian 矩阵来抵抗光照变化^[20]。

(2) 模型法

人脸表情识别中的模型法指的是对动态图像序列的表情信息进行参数化描述的统计方法。基于模型的特征提取方法能够提取人脸几何形变特征和纹理变化信息,但建立模型无法自动完成,需要手动辅助;另外,算法运算量较大且计算复杂。何俊等人提出一种新的基于 DBN(Deep Belief Net)模型无监督的表情特征提取和识别的方法^[21]。苟佳丽提出一种将 LBP 和 LDP 融合为单一特征即局部主要纹理特征模型(Local Principal Texture Pattern, LPTP)^[22],该方法能较好地弥补 LBP 和 LDP 对光照、噪声敏感的缺陷,但它仅考虑了空域上的特征变化,忽略了时域上的变化。

(3) 特征点跟踪法

基于特征点的跟踪方法通常是将特征点选取在图像灰度

变化比较大的区域,如嘴角、眼角等部位。跟踪这些区域中特征点的位移变化,得到面部特征的位移和形变信息。Tie 等人提出一种自动从视频序列中的人脸面部模型上提取 26 个特征点并通过多个粒子滤波器来进行跟踪的方法^[23]。Tslakanidou 等人提出了基于 ASM 的三维人脸特征跟踪方法^[24],该方法对人脸 81 个特征点进行跟踪建模,实现了对部分复合运动单元的识别。

3.2 特征降维方法

经过提取原始特征的人脸表情图像经常会有着较大的空间维数,对这些图像进行降维是十分有必要的,这不仅可以使特征的维数降低,大大缩减算法的运算时间,同时也能提高分类识别的效果。特征降维的主要任务是在不影响原始数据信息的基础上对抽取的数据特征进行降维优化。接下来将介绍几种经典的特征降维方法以及近年学者在该方面的改进。

主元分析法(Principal Component Analysis,PCA)是一种常用的特征降维方法,该方法可以有效提取图像中对表情识别贡献较大的区域,可以有效消除冗余、降低维数;但它没有考虑不同类别数据之间的区分性,而且当维数过高时,PCA 的效果下降明显。Ying 等人提出了一种基于支持向量机的鉴别分析算法^[25],该算法以 Fisher 和 SVM 为基础,能够在有限样本的情况下使表情数据具有最大的类间分离性,然后再进行降维,克服了 PCA 算法的局限性;但 PCA 算法仅仅利用了样本的二阶统计信息,没有考虑到高阶统计量。Radulovic 等人提出了独立成分分析(Independent Component Analysis,ICA),ICA 可以看作是 PCA 的推广^[26],它不仅利用了二阶统计特性的分析方法,还利用了高阶统计特性的分析方法。

线性判别分析(Linear Discriminant Analysis,LDA),也称作 Fisher 线性判别(Fisher Linear Discriminant,FLD),是模式识别的经典算法,1996 年由 Belhumeur 引入模式识别和人工智能领域。LDA 算法是一种较为有效的特征提取方法,能够使样本的类间散布矩阵最大和类内散布矩阵最小。Wang 等人提出了一种改进 LDA 算法——局部线性判别分析(Local Linear Discriminant Analysis,LLDA)^[27],该方法能克服传统 LDA 的局限性,而且降低了算法的复杂度,在 JAFFE 和 CK 上的实验均表明其效果要优于 LDA。

局部线性嵌入(Locally Linear Embedding,LLE)是 McIvor 等人提出的一种非线性数据的无监督降维方法,能在实现数据降维的同时保留原有数据的拓扑结构^[28]。Zhao 等人提出了一种新的基于 LLE 的有监督的流形学习算法^[29],被称为核判别局部线性嵌入(Discriminant Kernel Locally Linear Embedding,DKLLE),在 CK 数据库上的实验表明该算法优于 LLE 算法。

3.3 特征分解方法

人脸表情图像包含了丰富的表情信息,人脸表情识别是利用各种表情特征之间的差异信息来区分各种表情。对一种表情识别有利的信息可能会对另一种表情的识别带来干扰,解决方法之一是把人脸图像分解到多个子空间里进行分析处理。M. A. O. Vasilescu 等人提出了高阶奇异值分解(Higher Order Singular Value Decomposition,HOSVD)的人脸识别算法^[30],Wang 等人对该算法进行了改进^[31],并用三阶张量将人脸图像分解到个人子空间和表情子空间,从而实现人脸表情的识别和合成。Zhu 等人提出一种基于流行的表情分类算法^[32],该算法基于 LPP(Locality Preserving Projections)流行

学习算法和 HOSVD,结合 LPP 算法的特征提取能力和 HOSVD 的模式分解能力,使这两者的优势得到了互补,其中 LPP 解决了 HOSVD 中图像特征值的提取问题,而 HOSVD 解决了表情识别中个人模式和表情模式互相干扰的问题。

4 表情分类

表情分类是人脸表情识别过程中不可或缺的步骤,其目的是判断测试样本的表情特征所属的类别。常用的表情分类方法有:支持向量机(Support Vector Machine,SVM)、K 最近邻算法(K-Nearest Neighbor,KNN)、人工神经网络算法(Artificial Neural Networks,ANN)、AdaBoost 算法和隐马尔科夫模型(Hidden Markov Model,HMM)等。

SVM 的目的是找到有限样本下的最优分类面,使类间距最大化。但由于在大样本的情况下矩阵的计算需要耗费大量的内存,会影响算法的运行效率,SVM 在解决小样本、非线性和高维数据问题上有很多的优势^[33]。KNN 算法是一种比较成熟的分类方法,是最近邻算法的一种扩展,也是常用的机器学习方法之一。ANN 是从信息处理角度对脑神经网络进行抽象而建立的一种数学模型,按照不同的连接方式组成不同的网络。Adaboost 分类方法是迭代算法,其思想是通过在训练样本集上训练一系列弱分类器,然后把把这些弱分类器组合起来形成一个强分类器^[34]。HMM 是 Markov 链的一种,它的状态不能直接观测到,但可以通过观测向量序列观测到,每个观测向量都通过某些概率密度分布表现为各种状态,并且每一个观测向量都由一个状态序列产生。

每种分类方法都有一些缺点,但这些分类方法都得到了改进。Zhan 等人提出了模糊深隐马尔科夫算法^[35],该算法降低了经典 HMM 算法的条件独立假设,提高了算法的识别率和鲁棒性。王剑云等人提出一种局部并行深度神经网络的表情识别方法^[36],其规划训练多个并行网络,获得了更好的表情分类判别性。

此外,为了能够达到更好的分类识别效果,一些研究者们采用多种分类方法相结合的形式来对表情特征进行分类。徐文晖等人将 KNN 与 SVM 结合起来,将近邻信息集成到 SVM 中^[37],提出了局部 SVM 分类器。Wang 等人基于模糊支持向量机(FSVM)和 KNN 提出了一种新的表情识别算法^[38],该算法将区域划分为不同的类型,并结合 FSVM 和 KNN 的特点,根据不同的类型使用不同的分类方法,达到了很好的识别精度,并且有效地降低了计算的复杂度。

结束语 通过文中对人脸表情识别相关知识以及近几年国内外学者的研究成果及最新相关技术的介绍,可以看出人脸表情识别领域的研究取得了很多进展,但与真人识别相比仍然有一段距离,还没有达到实际应用的要求。目前看来,以下几个方面的研究将会大幅度提升 FER 系统的性能。

(1)人脸检测是 FER 系统必要的环节,但目前人脸检测仍然有很大的局限性,特别是在有遮挡条件和复杂背景的条件下,将会大大提升识别的难度并降低识别的精确度。

(2)微表情是人类在不经意间流露出来的表情,能够真实地反映对象的内心情感变化,在心理学上有很重要的应用,比如测谎技术的需求。但由于微表情发生的时间短且特征变化不明显,需要对表情分类更加精确,微表情的研究也将是国内外学者的研究重点。

(3)目前,人脸表情识别技术主要是在二维图像上进行研

究,但二维图像毕竟比较局限,相比较来说,3D 的信息量要远超过 2D,三维表情识别的研究将更有前景。

参 考 文 献

- [1] Ojala T, Pietikäinen 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, Pietikäinen M. Discriminative features for texture description [J]. Pattern Recognition, 2012, 45(10): 3834-3843
- [4] 周宇旋, 吴秦, 梁久祯, 等. 判别性完全局部二值模式人脸表情识别[J/OL]. 计算机工程与应用, <http://www.cnki.net/kcms/detail/11.2127.TP.20151012.1535.004.html>
- [5] Gabor D. Theory of communications. IEEE Journal of the Institution of Electrical Engineers-Part III[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] 钟思志. 人脸面部表情识别算法研究[J]. 上海: 华东师范大学, 2015
- [8] Cootes T, Taylor C, Cooper D, et al. Active shape models-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]. CAAI Transactions on Intelligent Systems, 2011, 6(3)
- [10] 侯婕. 人脸表情计算技术研究[D]. 苏州: 苏州大学, 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] Saatei 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 D G. Object recognition from local scale-invariant features [C] // International Conference on Computer Vision. Greece: 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] 蔡乐毅. 基于主动形状模型的人脸识别算法的研究[D]. 杭州: 浙江工业大学, 2013
- [17] Ren Fu-ji, Huang Zhong. Facial Expression Recognition Based on AAM-SIFT and Adaptive Regional Weighting [J]. IEEE Transactions on Electrical and Electronic Engineering, 2015, 10: 713-722
- [18] Yacoob Y, Davis L. 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: Springer 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] 何俊, 蔡建锋, 房灵芝, 等. 基于 LBP/VAR 与 DBN 模型的人脸表情识别[J]. 计算机应用研究, 2016, 33(8): 2509-2513
- [22] 苟佳丽. 基于视频的人脸表情识别技术研究 [D]. 成都: 西南交通大学, 2014
- [23] Tie Y, Guan L. A deformable 3D facial expression model for dynamic human emotional state recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 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): 1763-1777
- [25] Ying Zi-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 power lines [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 11th International Conference on Signal Processing. Beijing, 2012: 1226-1229
- [28] McIvor R T, 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 algorithm 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] 王文成. 基于局部特征分析的人脸表情识别问题研究[D]. 济南: 山东大学, 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): 641-650
- [36] 王剑云, 李小霞. 一种基于深度学习的表情识别方法[J]. 计算机与现代化, 2015(1): 84-87
- [37] 徐文晖, 孙正兴. 面向视频序列表情分类的 LSVM 算法[J]. 计算机辅助设计与图形学学报, 2009, 21(4): 542-548
- [38] Wang Xiao-hu, Liu An, Zhang Shi-qing. New facial expression recognition based on FSVM and KNN [J]. Optik, 2015(126): 3132-3134