

# 人脸表情识别的研究进展

蒋 斌<sup>1</sup> 贾克斌<sup>1</sup> 杨国胜<sup>2</sup>

(北京工业大学电子信息与控制工程学院 北京 100124)<sup>1</sup> (中央民族大学信息工程学院 北京 100081)<sup>2</sup>

**摘 要** 人脸表情识别是人机交互、机器学习、智能控制和图像处理等领域涉及的重要研究方向,目前已成为国内外研究的热点。从人脸表情识别的特征提取和特征分类两方面出发,总结了国内外近几年人脸表情识别的进展状况。在特征提取阶段,根据所处理的图像的属性,分别从静态图像和动态图像两个方面总结人脸表情的特征提取算法,前者包括整体法和局部法,后者分为模型法、光流法和几何法。在分类器的设计上,以贝叶斯网络和距离度量两条理论主线,贯穿主要的方法。最后结合国内外最新的研究成果和应用领域,展望了人脸表情识别的发展。

**关键词** 人脸表情识别,特征提取,特征分类

中图分类号 TP391.4 文献标识码 A

## Research Advance of Facial Expression Recognition

JIANG Bin<sup>1</sup> JIA Ke-bin<sup>1</sup> YANG Guo-sheng<sup>2</sup>

(School of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124, China)<sup>1</sup>

(Information Engineering College, Minzu University of China, Beijing 100081, China)<sup>2</sup>

**Abstract** In recent years, facial expression recognition has become a hot research direction in human computer interaction, machine learning, intelligent control and image processing. According to feature extraction and feature classification, recent developments of facial expression recognition were presented. From static images and image sequences, the methods of feature extraction can be divided into two categories. The former includes holistic methods and local methods, the later includes template-based methods, geometry-based methods and optical flow methods. In the classifier design, the main methods of feature classification can be categorized by Bayesian Network methods and Distance Metric methods. Finally, combining the latest productions and applications at home and abroad, the expectation of the development of facial expression recognition was given.

**Keywords** Facial expression recognition, Feature extraction, Feature classification

## 1 引言

人脸表情是传播人类情感信息与协调人际关系的重要方式。据心理学家 A. Mehrabian<sup>[1]</sup> 的研究表明,在人类的日常交流中,通过语言传递的信息仅占信息总量的 7%,而通过人脸表情传递的信息却达到信息总量的 55%。通过识别,人脸表情可以获取大量有价值的信息。因此基于人类视觉特征,采用计算机技术,对人脸表情特征进行归类的人脸表情识别成为人机交互、情感计算、智能控制、机器视觉、图像处理与模式识别等领域的重要研究课题。

20 世纪 70 年代美国心理学家 Ekman 和 Friesen<sup>[2]</sup> 对现代人脸表情识别做出了开创性的工作。Ekman 定义了人类的 6 种基本表情:高兴(Happy)、生气(Angry)、吃惊(Surprise)、恐惧(Fear)、厌恶(Disgust)和悲伤(Sad),确定了识别对象的类别;其次是建立了面部动作编码系统(Facial Action Coding System, FACS),使研究者按照系统划分的一系列人

脸动作单元(Action Unit, AU)来描述人脸面部动作,通过人脸运动和表情的关系,进而检测人脸面部细微表情。1978 年, Suwa<sup>[3]</sup> 等人对一段人脸视频动画进行了人脸表情识别的最初尝试,一系列的研究在人脸表情视频序列上展开。到上世纪 90 年代,随着图像处理与模式识别技术的发展,使得人脸表情识别的计算机自动化处理成为可能, K. Mase 和 A. Pentland<sup>[4]</sup> 是其中的先驱者。二人首先使用光流来判断肌肉运动的主要方向,然后提取局部空间中的光流值,组成表情特征向量,最后利用表情特征向量构建人脸表情识别系统。该系统可以识别高兴、生气、厌恶和惊奇 4 种表情,识别率接近 80%。1997 年,哈尔滨工业大学的高文教授<sup>[5]</sup> 领导的团队将人脸表情识别的研究成果引入我国。2003 年,北京科技大学的王志良教授<sup>[6]</sup> 领导的团队,将人脸表情识别算法应用于机器人的情感控制研究中,并发表了 2002 年以来人脸表情识别发展情况的综述。2004 年,东南大学的郑文明博士<sup>[7]</sup> 在面部表情识别方面,提出了基于核典型相关分析、偏最小二乘回归

到稿日期:2010-05-07 返修日期:2010-12-04 本文受国家自然科学基金项目(30970780)资助。

蒋 斌(1983-),男,博士生,主要研究方向为 人脸表情识别和多媒体信息处理, E-mail:bj2009@emails.bjut.edu.cn;贾克斌(1962-),男,博士,教授,主要研究方向为基于内容的图像检索与处理技术;杨国胜(1963-),男,博士后,教授,CCF 高级会员,主要研究方向为多传感器多目标数据融合技术、目标识别、图像处理技术和机器人控制。

等多种识别方法,并负责开发了自动面部表情识别系统。2006年,国家自然科学基金对人脸表情识别的相关研究正式立项。直至今日,项目数有总体增长的趋势。国内的清华大学、中国科学院、北京航空航天大学、北京交通大学、北京科技大学、哈尔滨工业大学、东南大学、上海交通大学、西北工业大学、华中师范大学等多所高校和科研机构参与了人脸表情识别相关课题的研究。虽然人脸表情识别的商业应用还处于起步阶段,但是国内外研究机构和企业都在不同的领域进行尝试和研究,部分成果已经取得了专利。因此人脸表情识别的研究具有很大的开发潜力。

人脸表情识别的研究存在着不少难点。首先,关于表情的准确定义仍然存在争论,一些人脸动作是否属于表情,还有待学者们讨论和判断;其次,表情的变化不易捕捉,表情本身具有一定的复杂性;最后,表情识别的鲁棒性算法还有待研究。虽然国内外的研究者<sup>[8-12]</sup>分别在不同时期对人脸表情识别的发展状况进行了总结,但是人脸表情识别的发展十分迅速;因此作者根据近几年国内外的研究成果,介绍人脸表情识别研究的进展状况。

## 2 人脸表情识别

### 2.1 人脸表情识别系统概述

人脸表情识别系统如图1所示,主要包括人脸图像的获取、人脸检测、特征提取和特征分类。人脸图像的获取可以依靠摄像设备实时采集,经过光照补偿和几何归一化处理,以静态图像或动态图像的形式保存在图像库中。



图1 人脸表情识别系统框图

目前国际上常用的专业人脸表情图像库有:

(1)日本 ATR 人类信息处理研究实验室和日本九州大学心理学系建立的日本女性人脸表情数据库(Japan Female Facial Expression, JAFFE)<sup>[13]</sup>,包括 10 名日本女性共 213 幅图像,每人有 6 种基本表情和中性表情,每种表情有 2~4 幅图像;

(2)美国卡内基梅隆大学机器人研究所和心理学系建立的 Cohn-Kanade(CK)人脸表情数据库<sup>[14]</sup>,从各种人群中提取了包含各种动作单元的表情图像序列。该数据库已于近期推出第 2 版,新的 CK+ 数据库<sup>[15]</sup>增加了人脸位置移动的图像序列。

(3)荷兰代尔夫特理工大学的 MMI(Man Machine Interaction)数据库<sup>[16]</sup>,目前已经采集了超过 50 个人的视频和图像,主要包括反映人脸表情的各种动作单元。

国内的人脸表情图像库有:

(1)清华大学的人脸表情视频数据库<sup>[17]</sup>,包括了 70 个人的 1000 段脸部表情视频,涵盖了常见的 8 类情感类表情和中文语音发音的说话类表情;

(2)北京航空航天大学(BHU(Beihang University))人脸表情数据库<sup>[18]</sup>是一个较为全面的人脸表情数据库,包含 3 类人脸表情:单一表情、混合表情和复杂表情;

(3)中国科学院的 CAS-PEAL(Chinese Academy of Sci-

ence-Pose, Expression, Accessory and Lighting)人脸数据库<sup>[19]</sup>包含了 1040 个人的 6 种面部表情和动作,包括中性、闭眼、皱眉、微笑、惊奇和张嘴。

部分人脸图像库中也包含若干表情图像,比如 AR 图像库<sup>[20]</sup>包含中性、微笑、生气和尖叫 4 种表情;Yale 人脸图像库<sup>[21]</sup>包含中性、高兴、悲伤、欲睡、惊奇和眨眼 6 种人脸的表情和动作。另外,纽约州立宾汉姆顿大学建立的 BU-3DFE (Binghamton University 3D Facial Expression)三维人脸表情图像库<sup>[22]</sup>,包含了 100 个人的 2500 个人脸表情模型。

如图 2 所示,由于专业图像库的背景较为单一,人脸检测算法又已经发展成为独立的研究方向,因此人脸表情识别算法的发展主要体现在特征提取和特征分类之上。



图2 JAFFE 和 CK 数据库的部分图像

### 2.2 特征提取<sup>1)</sup>

表情特征提取主要是采用数学方法,依靠计算机技术对人脸表情的数字图像进行数据的组织和处理,提取表情特征,去除非表情噪声的方法。在某些情况下,特征提取算法提取了图像的主要特征,客观上降低了图像的维数,因此这些特征提取算法也具有特征降维的作用。人脸表情的产生是一个很复杂的过程,如果不考虑心理和环境因素,呈现在观察者面前的就是单纯的肌肉运动,以及由此带来的面部形体和纹理的变化。静态图像呈现的是表情发生时单幅图像的表情状态,动态图像呈现的是表情在多幅图像之间的运动过程。因此根据表情发生时的状态和处理对象来区分,表情特征提取算法大体分为基于静态图像的特征提取方法和基于动态图像的特征提取方法。其中基于静态图像的特征提取算法可分为整体法和局部法,基于动态图像的特征提取算法又分为光流法、模型法和几何法。

#### 2.2.1 基于静态图像的特征提取方法

##### (1) 整体法

人脸表情依靠肌肉的运动来体现。人脸表情静态图像直观地显示了表情发生时人脸肌肉运动所产生的面部形体和纹理的变化。从整体上看,这种变化造成了面部器官的明显形变,会对人脸图像的全局信息带来影响,因此出现了从整体角度考虑表情特征的人脸表情识别算法。

整体法中的经典算法包括主元分析法(Principal Component Analysis, PCA)、独立分量分析法(Independent Component Analysis, ICA)和线性判别分析法(Linear Discriminant Analysis, LDA)。其中 PCA 算法是一种无监督方法,其优点是不需要待处理样本的类别信息,同时可以有效地处理高维数据。缺点是该方法处理数据时产生的协方差矩阵维数过高,影响了算法的运行。作为 PCA 算法的延伸,二维主元分

1) 部分算法的表情识别率如表 1 所列。

析算法(2DPCA)在一定程度上解决了这一问题。ICA 算法也属于无监督方法,可以提取像素间隐藏的信息,并且适合非

高斯分布的数据处理,但是算法的实时性有待提高。LDA 算法属于有监督方法,其特点是充分地保留了样本的类别结构。

表1 部分表情识别算法的结果

	参考文献	图像来源	特征提取	特征分类	数据量	训练/测试	识别率
静态 图像	[24]	JAFFE	FastICA	HMM	213 幅	70/143	95.80% <sup>①</sup>
	[26]	JAFFE	SVDA	SVDA	210 幅	70/140	91.43% <sup>①</sup>
	[29]	JAFFE	Gabor 小波	SFRCS	213 幅	交叉验证	96.71% <sup>①</sup> 85.92% <sup>②</sup>
		MMI			235 幅		93.61% <sup>①</sup>
	[30]	JAFFE	Gabor 小波	KCCA	183 幅	交叉验证	98.36% <sup>①</sup> 77.05% <sup>②</sup>
	[31]	JAFFE	CBP	CNN	196 幅	交叉验证	87.50%
		CK			1820 幅		88.95%
	[32]	AR	SMOM	ESTM	121 幅	60/61	96.40%
		Yale			75 幅	交叉验证	94.70%
	动态 图像	[34]	实时采集	光流法	SVM		
[36]		自主采集	2D+3D ASM	基于规则的 分类法	832 个序列	260/572	83.60% <sup>①</sup>
[39]		FED06	Differential- AAM	k 近邻法	512 个序列	交叉验证	95.00% <sup>①</sup>
[43]		CK	ASM 结合 几何特征	LSVM	492 个序列	交叉验证	89.11% <sup>①</sup>
[48]		CK	Gabor 小波	FBMM	240 个序列	120/120	94.90%
[61]		CK	Zernike 矩	NB			73.20% <sup>②</sup>

注:①表示测试集中的人不在训练集中出现;②表示测试集中的人在训练集中出现。

近年来,研究者们针对人脸表情产生的整体变化,从两方面入手来发展整体法。一方面是在经典算法的基础上进行改进;另一方面则是从统计模式识别中寻找思路,构造新的特征提取算法,进而从图像中提取可能的表情特征。

文献[23-25]采用 FastICA 算法提取表情特征,该方法不但继承了 ICA 算法能够提取像素间隐藏信息的特点,而且可以通过迭代,快速地完成对表情特征的分离。文献[26]提出了支持向量鉴别分析(Support Vector Discriminate Analysis, SVDA)算法,该算法以 Fisher 线性判别分析和支持向量机(Support Vector Machine, SVM)为基础,能够在小样本数据情况下,使表情数据具有最大的类间分离性,而且不需要构建 SVM 算法所需要的决策函数。实验证明了该算法的识别率高于 PCA 和 LDA。文献[27]依靠二维离散余弦变换,通过频域空间对人脸图像进行映射,结合神经网络实现对表情特征的分类。文献[28]则进一步指出二维离散余弦变换可以在较低的特征维数下,实现较高的表情识别率。实验表明,该算法在一定程度上优于 2DPCA 算法。

## (2) 局部法

静态图像上的人脸表情不仅有整体的变化,也存在局部的变化。面部肌肉的纹理、皱褶等局部形变所蕴含的信息,有助于精确地判断表情的属性。局部法的经典方法是 Gabor 小波法和 LBP(Local Binary Pattern)算子法。Gabor 小波在特征提取方面具有良好的视觉特性和生物学背景。但是该算法需要计算不同尺度和方向的小波核函数,由此产生的高维特征向量既不利于算法的实时处理,又在一定程度上造成了信息冗余。和 Gabor 小波相比, LBP 算子法可以更有效地提取人脸表情特征。但是 LBP 算子产生的高维直方图影响了算法的实时性,而且该算法产生的二值数据容易受到噪声的干扰。近年来这两种算法在人脸表情识别领域仍有不同程度的应用,出现了相关的改进算法以及考虑人脸纹理细节特征的算法。

文献[29]以 Gabor 小波等多种特征提取算法为手段,结合新的分类器对静态图像展开实验。文献[30]首先人工标记了 34 个人脸特征点;然后将特征点的 Gabor 小波系数表示成标记图向量;最后计算标记图向量和表情语义向量之间的 KCCA(Kernel Canonical Correlation Analysis)系数,以此实现对人脸表情的分类。针对传统 LBP 算子法存在的直方图维数大、算法抗噪性差的缺点,文献[31]提出了 CBP(Centralized Binary Pattern)算子法,通过比较环形邻域的近邻点对,降低了直方图的维数。针对符号函数的修改,又增强了算法的抗噪性,使 CBP 算子法取得了较高的识别率。文献[32]提出了基于形状和纹理的人脸表情识别算法,该算法通过构建 SMOM(Spatially Maximum Occurrence Model),获取了图像强度分布,并通过 ESTM 算法(Elastic Shape-Texture Matching),测量了图像的形状和纹理信息,最终实现了人脸表情识别。

## 2.2.2 基于动态图像的特征提取方法

动态图像与静态图像的不同之处在于:动态图像反映了人脸表情发生的过程。因此动态图像的表情特征主要表现在人脸的持续形变和面部不同区域的肌肉运动上。目前基于动态图像的特征提取方法主要分为光流法、模型法和几何法。

### (1) 光流法

光流法是反映动态图像中不同帧之间相应物体灰度变化的方法。早期的人脸表情识别算法多采用光流法提取动态图像的表情特征,这主要在于光流法具有突出人脸形变、反映人脸运动趋势的优点。因此该算法依旧是近年来研究动态图像表情识别的重要方法。

文献[33]首先采用连续帧之间的光流场和梯度场,分别表示图像的时空变化,实现每帧人脸图像的表情区域跟踪;然后通过特征区域运动方向的变化,表示人脸肌肉的运动,进而对应不同的表情。文献[34]在综合了人脸生物学信息以后,首先使用改进的 Ratio Template 算法实现对人脸区域的检

测;然后使用光流法计算人脸的多渠道灰度模型,完成对人脸区域的跟踪;最后用 SVM 算法实现了表情分类。

以上算法在突出光流法优点的同时,结合不同的算法,实现了对动态图像的表情特征提取。但是光流法也存在一定的局限性。该算法成立的前提是灰度守恒假设和光流场平滑性假设,这就要求使用光流法的动态图像应该满足灰度守恒和图像平滑的条件。所以当动态图像不满足上述条件时,光流法的效果就会受到很大影响。针对这一问题,文献[35]首先计算了前向后向光流方程的 Hessian 矩阵,然后将 Hessian 矩阵的条件数与 Lucas-Kanade 光流法中的加权矩阵相结合,有效地消除了局部领域中不可靠的约束点,完成了对 Lucas-Kanade 光流法的改进,从而更好地实现了基于动态图像的表情特征提取。

## (2) 模型法

人脸表情识别中的模型法是指对动态图像的表情信息进行参数化描述的统计方法。常用算法主要包括主动形状模型法(Active Shape Model, ASM)和主动外观模型法(Active Appearance Model, AAM)。两种算法都可分为形状模型和表观模型两部分。就表观模型而言,ASM 反映的是图像的局部纹理信息,而 AAM 反映的是图像的全局纹理信息。模型法可以通过动态图像的形状和纹理信息来反映人脸表情的整体和局部特征。但是在一般情况下,该方法需要依靠人工标注的方式实现人脸特征点的初始化,从而在一定程度上影响了算法的自动化程度。模型法的另一个缺点在于模型的构建方式较为复杂。对模型法而言,基于 ASM 和 AAM 的算法改进是其近期发展的主要特点。

文献[36]提出了基于 ASM 的三维人脸特征跟踪方法。该方法对人脸 81 个特征点进行跟踪建模,实现了对部分复合动作单元的识别。文献[37]借助图像的地形特征模型来识别人脸动作和表情;利用 AAM 和人工标记的方法跟踪人脸特征点,并按照特征点取得人脸表情区域;通过计算人脸表情区域的地形直方图来获得地形特征,从而实现表情识别。文献[38]提出了基于二维表观特征和三维形状特征的 AAM 算法,在人脸位置发生偏移的环境下,实现了对表情特征的提取。文献[39]提出了 Differential-AAM 和流形学习相结合的人脸表情识别算法。首先求出表情图像与中性图像之间的差分图像,计算此差分图像的 AAM 参数;然后使用流形学习算法将 AAM 参数嵌入到平滑且连续的特征空间内;最后通过 Directed Hausdorff Distance 和 k 近邻法,实现了表情的相似性测量。实验证明该算法比 AAM 算法的识别率提高了 20%。文献[40]提出了传统的 ASM 方法,即通过提取相邻两个特征点垂线方向上的点来训练模型,其目标是在特征点附近提取灰度信息差异显著的点来进行训练。但该方法不能充分提取某些特殊点(如眉尖、眼角、嘴角)的灰度向量。针对这一问题,文献[40]提出了基于 ASM 和 AAM 的联合模型的表情识别算法。不过该算法在动态图像上的识别效果还有待验证。

## (3) 几何法

在表情特征提取方法中,研究者考虑到表情的产生与表达在很大程度上是依靠面部器官的变化来反映的。人脸的主要器官及其褶皱部分都会成为表情特征集中的区域。因此在

面部器官区域标记特征点,计算特征点之间的距离和特征点所在曲线的曲率,就成为了采用几何形式提取人脸表情的方法。常规方案是将该方法应用在静态图像的特征提取上,而且一般的观点认为几何法只能勾勒出人脸的整体形变,在局部信息的提取上还存在着不足。然而近几年,这一方法却在动态图像的特征提取上得到广泛应用。原因有二:首先,动态图像通过人脸特征点在每帧中的变化反映了表情的完整过程,为几何计算提供了丰富的信息;其次,从动态图像中提取的几何特征对应着人脸的不同表情区域,这些特征适合用于动作单元的识别。

文献[41]使用形变网格对不同表情的人脸进行网格化表示,将第一帧与该序列表情最大帧之间的网格节点坐标变化作为几何特征,实现对表情的识别。文献[42]首先完成了人脸特征点的初始化;然后使用辅助粒子滤波器以及基于颜色的观察值模型,实现对不同帧之间特征点的跟踪;最后通过特征点的几何特征来判断动作单元类别。文献[43]采用 ASM 提取人脸图像的 68 个特征点,通过计算图像序列的几何特征,实现了基于 SVM 的表情分类。文献[44]用几何特征和暂时性特征提取方法,结合 SVM 分别对图像序列对应的表情进行识别。

目前几何法在动态图像中的研究主要以单纯的应用为主,相应改进算法的研究还有待发展。

## 2.3 特征分类

特征分类的目的是判断特征所对应的表情类别。在人脸表情识别中,表情的类别分为两部分:基本表情和动作单元。前者一般适用于所有的处理对象,后者主要适用于动态图像。研究者按照特征分类算法的时空特点,通常将特征分类算法分为两类:基于空间分析的方法和基于时空相结合的方法。前者主要包括近邻法、基于专家规则的方法、神经网络算法、贝叶斯分类法、SVM 算法以及 AdaBoost 算法。后者主要是隐马尔可夫模型(Hidden Markov Model, HMM)算法。近年来,贝叶斯分类法、HMM 算法、近邻法和 SVM 算法是表情特征分类的主要算法。

本文从上述算法的特点入手,将主要的特征分类方法分为基于贝叶斯网络的分类方法和基于距离度量的分类方法。

### 2.3.1 基于贝叶斯网络的方法

贝叶斯网络是以贝叶斯公式为基础、基于概率推理的图形化网络。从人脸表情识别的角度出发,概率推理的作用就是从已知表情信息中推断出未知表情的概率信息的过程。基于贝叶斯网络的方法包括各种贝叶斯网络分类算法和 HMM 算法。

文献[45,46]就是采用贝叶斯网络对表情进行分类的方法。在文献[46]中,研究者分别采用了朴素贝叶斯(Naive Bayes, NB)分类器、树增强型简单贝叶斯分类器(Tree Augmented Naive Bayes Classifier, TAN)和 HMM 实现表情特征分类。对于这 3 种算法而言,NB 算法具有算法简单、对缺失数据不敏感的优点。NB 算法的缺点在于其所需要的属性之间相互独立的假设与实际应用不太相符,因此在一定程度上影响了算法的识别率。TAN 算法在 NB 算法中引入了树型结构,从而在保证算法可行的同时,放松了属性之间相互独立的假设。HMM 算法通过计算观察值的后验概率,实现对表

情属性的判断。由于引入了时间变量,HMM算法可以看成一种动态贝叶斯网络。文献[47]就是采用HMM算法识别动作单元,实现了基于刚性和非刚性运动分离以及三维位置估计的人脸表情识别系统。文献[48]则在HMM算法的基础上,提出了FBMM(Fuzzy Buried Markov Model)算法,该算法放宽了经典HMM算法的条件独立假设,提高了算法的识别率和鲁棒性。

### 2.3.2 基于距离度量的分类方法

基于距离度量的分类方法是通过计算样本之间的距离来实现表情分类的。代表算法有近邻法和SVM算法。

近邻法是比较未知样本 $x$ 与所有已知类别的样本之间的欧式距离,通过距离的远近来决策 $x$ 与已知样本是否同类;SVM算法则是通过优化目标函数,寻找到使不同类别样本之间距离最大的分类超平面。以上算法中,近邻法的优点是算法简单,缺点是需要将所有样本存入计算机中,每次决策都要计算待识别样本与所有样本之间的距离,因此算法的存储量和计算量都很大。SVM的优点是算法具有坚实的理论基础,尤其适合小样本数据的分类。但是经典的SVM算法是二类分类算法,因此在多类分类和大规模样本的运算上,还存在着一定的不足。目前近邻法和SVM的改进算法是基于距离度量的表情分类方法的主流。

文献[31]采用了最近邻法对表情特征进行分类,并指出最近邻法的不足之处在于分类正确率的大小依赖于待分类样本的数量。因此作者采用中心最近邻分类器,通过计算已标记点与该类中心的距离来度量样本特征的表情属性。文献[43,49]分别从各自角度提出了对SVM的改进,前者将 $k$ 近邻法与SVM结合起来,把近邻信息集成到SVM的构建中,提出了局部SVM分类器;后者提出的CSVMT模型将SVM和树型模块结合起来,以较低的算法复杂度解决了分类子问题。文献[43,49]都在不同程度上提高了表情特征的分类精度。

## 3 人脸表情识别的展望

伴随着各种理论的发展和实验要求的严格,人脸表情识别算法的发展也呈现出以下趋势:

### (1) 结合机器学习理论,研究新的识别算法

流形学习(Manifold Learning)算法是一种研究高维空间数据在低维空间上的流形表示的学习方法。自2000年在“Science”杂志上出现后,其思想逐渐形成了研究的热潮。该方法出现伊始,就提出人脸是一种典型的流形结构,从而建立了流形学习算法与人脸表情识别之间的紧密联系。文献[25,50-53]就是近几年流形学习算法的新成果在人脸表情识别中的应用。文献[52]提出了一种修正的SLLE(Supervised Locally Linear Embedding,SLLE)算法,该算法考虑了表情之间的差异性,对不同类别样本之间的距离采用不同的加权值,从而更充分地利用了关于表情的先验知识。文献[53]指出目前并没有足够的证据表明单一流形可以确保最优的分类结果,因此采用多流形算法提取表情特征是一项有益的尝试。

半监督学习(Semi-Supervised Learning)是目前机器学习领域重点研究的一种理论。其主要思想是结合少量标记样本和大量未标记样本,在简化对样本标记的工作量的同时,建立

更准确的数据模型,完成学习任务。文献[54]采用最基本的基于高斯模型的半监督分类方法实现了人脸表情识别,并指出半监督学习的研究成果在人脸表情识别中的应用将是表情识别研究的重要方向。目前半监督学习在人脸表情识别中的研究成果还不多,因此该方向也具有很大的研究价值。

### (2) 借鉴与人类情感相关的研究成果,寻找新的研究思路

国内外关于人脸表情识别的研究从20世纪90年代开始进入了一个活跃时期。1997年,美国麻省理工学院媒体实验室的Rosalind W. Picard教授提出了情感计算(Affective Computing)理论。1998年,北京科技大学的王志良教授提出了人工心理的概念。这些与情感相关的研究方向,与人脸表情识别之间存在着密切的联系,对表情识别的发展具有极大的启发和促进作用。

虽然神经科学(Neuro-Science)领域也存在很多关于人类情感的研究,但是很多研究成果与自动识别之间并没有直接的联系,因此目前应用到人脸表情识别的成果依然有限。不过研究者已经做出了尝试。文献[55]注意到了神经科学模拟人脑认知情感结构的成果,在Taylor情感注意神经网络的基础上,对CK数据库中的表情图像序列进行了实验。和BP神经网络相比,该方法取得了较好的结果,其中“高兴”表情的识别率达到了91.80%。

### (3) 研究鲁棒性的表情识别,完成更精确的识别任务

人脸表情识别获取的表情图像一般是背景单一、光照一致、头部无运动、面部无遮挡、不说话的人脸正面图像。研究者考虑在实际生活中,头部偏转或者面部存在遮挡物的情况(比如佩戴眼镜、口罩,或者留长发等)是很常见的。在上述情况发生时,获取的人脸表情信息存在着缺失,因此需要鲁棒性的表情识别算法来完成人脸不完整信息的表情识别任务。

文献[38,56,57]研究了头部偏转时的鲁棒表情识别问题。文献[38]通过2D+3D AAM算法,实现了头部偏移时人脸表情的鲁棒性识别,最高识别率达到91.87%。文献[56]采用AAM跟踪时序图像的特征点,结合人脸的动作单元特征,识别人脸方向变化时细微表情的变化。文献[57]以动态贝叶斯网络为基础,采用一种联合概率人脸动作模型(Unified Probabilistic Facial Action Model),实现了在头部偏转状态下的表情识别。

文献[58-60]研究了面部遮挡条件下的鲁棒表情识别问题。文献[58]借鉴Fisher线性判别和SVM的思想,提出了基于最小类内方差的多类分类器。实验研究了在人脸不同器官被遮挡情况下的表情识别效果。文献[59]通过热像处理(Thermal Image Processing)研究了说话人佩戴眼镜和不佩戴眼镜时的表情识别问题。文献[60]指出目前在遮挡条件下的鲁棒表情识别算法并没有去除面部遮挡,使待识别的面部图像上仍然存在遮挡物的错误信息,不可避免地造成了识别结果的不准确。因此作者提出采用显著性检测和鲁棒主成分分析(Robust Principal Component Analysis,RPCA)来实现面部遮挡检测,进而完成对遮挡区域的重构。

文献[61]考虑在噪声环境下的算法鲁棒性问题,采用Zernike矩提取表情特征,采用NB算法识别表情类别。在噪声和人脸图像大幅度旋转的情况下,实现了较高的识别率。

文献[62]提出鲁棒表情识别不仅要考虑光照、头部运动等干扰,而且要注意到人的因素。作者在“相貌相似的人具有相似表情”的假设前提下,使用高阶奇异值分解算法,实现了在与人无关(Person-independent)环境下的表情识别。

另一方面,当实验对象满足常规条件(无遮挡、无偏移的人脸正面图像)时,算法也会因为不同表情之间的相似性而造成误判。美国心理学家 Paul Ekman 指出生气与厌恶、恐惧和惊奇、恐惧与悲伤都是容易混淆的感情。文献[63]指出:大量的调查表明,不同的观察者对同一个表情的类别标注会有很大不同。

因此上述情况可以说明以下两点:首先,对表情的定义存在问题。目前研究人员通常采用7种表情(6种基本表情+中性表情)和动作单元来区分表情的属性。前者简单却过于泛泛,后者相对准确,却不易掌握。一些人脸动作(比如眨眼、开口、痛苦等)的表情属性不易判断,这是研究者需要考虑的问题。目前相关的研究主要有驾驶员疲劳识别<sup>[64]</sup>和儿童面部痛苦表情的识别<sup>[65]</sup>等。

其次,算法所提取的表情特征的精细程度不够,对该问题的研究主要集中在动作单元的识别以及动态图像所反映的表情强度的变化上。文献[66]提出了一种不同的思路,即采用AAM算法提取人脸特征点,并估计特征点的运动,通过运动放大(Motion Magnification)技术实现对精细表情的识别。综上所述,关于精细表情和混合表情识别<sup>[50]</sup>的研究仍需深入开展。

## 参 考 文 献

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

[2] Ekman P, Friesen W V. Facial Action Coding System: A Technique for the Measurement of Facial Movement[M]. Palo Alto: Consulting Psychologists Press, 1978

[3] Suwa M, Sugie N, Fujimora K. A Preliminary Note on Pattern Recognition of Human Emotional Expression[A]//Proceedings of the 4th International Joint Conference on Pattern Recognition [C]. Kyoto, Japan: Institute of Electrical and Electronics Engineers, 1978; 408-410

[4] Mase K, Pentlad A. Recognition of Facial Expression from Optical Flow[J]. IEICE Transactions, 1991, E74(10): 3474-3483

[5] 高文, 金辉. 面部表情图像的分析与识别[J]. 计算机学报, 1997, 20(9): 782-789

[6] 王志良, 陈峰军, 薛为民. 人脸表情识别方法综述[J]. 计算机应用与软件, 2003, 12: 63-66

[7] 郑文明. 基于核函数的判别分析研究[D]. 南京: 东南大学, 2004

[8] Pantic M, Rothkrantz L. Automatic Analysis of Facial Expressions; The State of the Art[J]. IEEE Transactions on PAMI, 2000, 22(12): 1424-1445

[9] Fasel B, Luetttin J. Automatic Facial Expressions Analysis; A Survey[J]. Pattern Recognition, 2003, 36(1): 259-275

[10] 刘晓旻, 谭华春, 章毓晋. 人脸表情识别研究的新进展[J]. 中国图象图形学报, 2006, 11(10): 1359-1368

[11] 薛雨丽, 毛峡, 郭叶, 等. 人机交互中的人脸表情识别研究进展[J]. 中国图象图形学报, 2009, 14(5): 764-772

[12] 王映辉. 人脸识别——原理、方法与技术[M]. 北京: 科学出版社, 2010

[13] Lyons M, Akamatsu S, Kamachi M, et al. Coding facial expressions with Gabor wavelets[A]//Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition [C]. Nara, Japan, 1998; 200-205

[14] Kanade T, Cohn J, Tian Ying-li. Comprehensive database for facial expression analysis[A]//Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition [C]. Grenoble, France, 2000; 46-53

[15] Lucey P, Cohn J, Kanade T, et al. The extended cohn-kanade dataset(ck+): A complete dataset for action unit and emotion-specified expression [A] // Proceedings of the International Workshop on CVPR for Human Communicative Behavior Analysis [C]. San Francisco, USA, 2010

[16] Pantic M, Valstar M, et al. Web-based Database for Facial Expression Analysis[A]//Proceeding of IEEE International Conference M&E[C]. Amsterdam, Netherlander, 2005; 317-321

[17] 吴丹, 林学闻. 人脸表情视频数据库的设计与实现[J]. 计算机工程与应用, 2004, 40(5): 177-180

[18] 薛雨丽, 毛峡, 张帆. BHU 人脸表情数据库的设计与实现[J]. 北京航空航天大学学报, 2007, 33(2): 224-228

[19] Gao Wen, Cao Bo, Shan Shi-guang, et al. The CAS-PEAL large-scale chinese face database and baseline evaluations[J]. IEEE Transactions on Systems, Man and Cybernetics, Part A, 2008, 38(1): 149-161

[20] Martinez A M, Benavente R. The AR Face Database[R]. Barcelona, Spain: The Computer Vision Center, 1998

[21] Belhumeur P, Hespanha J, Kriegman D. Eigenfaces vs. Fisherfaces; Recognition using class specific linear projection[J]. IEEE Trans. on PAMI, 1997, 19(7): 711-720

[22] Yin Li-jun, Wei Xiao-zhou, Sun Yi, et al. A 3D Facial Expression Database For Facial Behavior Research[A]//The 7th International Conference on Automatic Face and Gesture Recognition [C]. Southampton, UK, 2006; 211-216

[23] 何良华. 人脸表情识别中若干关键技术的研究[D]. 南京: 东南大学, 2005

[24] 周书仁, 梁昔明, 朱灿, 等. 基于 ICA 和 HMM 的表情识别[J]. 中国图象图形学报, 2008, 13(12): 2321-2328

[25] 周书仁. 人脸表情识别算法分析与研究[D]. 长沙: 中南大学, 2009

[26] 应自炉, 唐京海, 李景文, 等. 支持向量鉴别分析及在人脸表情识别中的应用[J]. 电子学报, 2008, 36(4): 725-730

[27] Ma L, Khorasani K. Facial Expression Recognition Using Constructive Feedforward Neural Networks[J]. IEEE Transactions on SMC SMC--Part B: Cybernetics, 2004, 34(3): 1588-1595

[28] Jiang Bin, Yang Guo-sheng, Zhang Huan-long. Comparative study of dimension reduction and recognition algorithms of DCT and 2DPCA[A]//BInternational Conference on Machine Learning and Cybernetics [C]. Kunming, China, 2008; 407-410

[29] Kyperountas M, Tefas A, Pitas I. Salient feature and reliable classifier selection for facial expression classification[J]. Pattern Recognition, 2010, 43: 972-986

[30] Zheng Wen-ming, Zhou Xiao-yan, Zou Cai-rong, et al. Facial Expression Recognition Using Kernel Canonical Correlation Analysis(KCCA)[J]. IEEE Transactions on Neural Networks, 2006, 17(1): 233-238

- [31] 付晓峰. 基于二元模式的人脸识别与表情识别研究[D]. 杭州: 浙江大学, 2008
- [32] Xie Xu-dong, Lam Kin-man. Facial expression recognition based on shape and texture[J]. *Pattern Recognition*, 2009, 42: 1003-1011
- [33] Yacoob Y, Davis L. Recognizing Human Facial Expressions from Long Image Sequences Using Optical Flow[J]. *IEEE Transactions on PAMI*, 1996, 18(6): 636-642
- [34] Anderson K, Mcowan P. A Real-Time Automated System for the Recognition of Human Facial Expressions[J]. *IEEE Transactions on SMC—Part B: Cybernetics*, 2006, 36(1): 96-105
- [35] 杨国亮. 人工心理相关技术研究-面部表情识别与情感建模[D]. 北京: 北京科技大学, 2006
- [36] Tslakanidou F, Malassiotis S. Real-time 2D+3D facial action and expression recognition[J]. *Pattern Recognition*, 2010, 43(5): 1763-1775
- [37] Wang Jun, Yin Li-jun. Static topographic modeling for facial expression recognition and analysis[J]. *Computer Vision and Image Understanding*, 2007, 108: 19-34
- [38] Sung J, Kim D. Pose-Robust Facial Expression Recognition Using View-based 2D+3D AAM[J]. *IEEE Transactions SMC—Part A: Systems and Humans*, 2008, 38(4): 852-866
- [39] Cheon Y, Kim D. Natural facial expression recognition using differential-AAM and manifold learning[J]. *Pattern Recognition*, 2009, 42: 1340-1350
- [40] 赵浩, 吴小俊. 基于改进联合模型的人脸表情识别[J]. *计算机工程*, 2010, 36(6): 206-209
- [41] Kotsia I, Pitas I. Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines[J]. *IEEE Transactions on Image Processing*, 2007, 16(1): 172-187
- [42] Pantic M, Patras I. Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences[J]. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 2006, 36(2): 433-449
- [43] 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
- [44] 左坤隆. 人脸表情自动分析与识别系统的研究[D]. 天津: 天津大学, 2004
- [45] Zhang Yong-mian, Ji Qiang. Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences[J]. *IEEE TRANSACTIONS ON PAMI*, 2005, 27(5): 699-714
- [46] Cohen I, Sebe N, et al. Facial expression recognition from video sequences: temporal and static modeling[J]. *Computer Vision and Image Understanding*, 2003, 91(1/2): 160-187
- [47] Wang Te-hsun, Lien J. Facial expression recognition system based on rigid and non-rigid motion separation and 3D pose estimation[J]. *Pattern Recognition*, 2009, 42: 962-977
- [48] Zhan Yong-zhao, Cheng Ke-yang, Chen Ya-bi, 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
- [49] 徐琴珍, 章品正, 裴文江, 等. 基于混淆交叉支撑向量机树的自动面部表情分类方法[J]. *中国图象图形学报*, 2008, 13(7): 1329-1334
- [50] 续爽, 贾云得. 基于表情相似性的人脸表情流形[J]. *软件学报*, 2009, 20(8): 2191-2198
- [51] Wang Hai-xian, Chen Si-bao, et al. Locality-Preserved Maximum Information Projection[J]. *IEEE Transactions on Neural Networks*, 2008, 19(4): 571-585
- [52] 应自炉, 李景文, 张有为. 基于表情加权距离SLEF的人脸表情识别[J]. *模式识别与人工智能*, 2010, 23(2): 278-283
- [53] Xiao Rui, Zhao Qi-jun, Zhang David, et al. Facial expression recognition on multiple manifolds[J]. *Pattern Recognition*, 2010
- [54] Cohen I, Cozman G. Fabio, SEBE Nicu, et al. Semisupervised Learning of Classifiers: Theory, Algorithms, and Their Application to Human-Computer Interaction[J]. *IEEE Transactions on Pami*, 2004, 26(12): 1553-1567
- [55] 王上飞, 薛佳, 王煦法. 基于人脑情感注意回路的表情识别[J]. *模式识别与人工智能*, 2008, 21(5): 598-602
- [56] 森博章, 宫脇健三郎, 佐野睦夫, 等. コミュニケーションを円滑に進めるための表情変化検出方式の検討[R]. *電子情報通信学会技術研究報告*, 2008: 159-162
- [57] Tong Yan, Chen Ji-xu, Ji Qiang. A Unified Probabilistic Framework for Spontaneous Facial Action Modeling and Understanding[J]. *PAMI*, 2010, 32(2): 258-273
- [58] Kotsia I, Pitas I, Zafeiriou S, et al. Novel Multiclass Classifiers Based on the Minimization of the Within-class Variance[J]. *IEEE Transactions on Neural Networks*, 2009, 20(1): 14-34
- [59] 中野真里, 池添史隆, 田伏正佳, 等. 発声時の温度顔画像からの表情認識の効率化と性能への個人差の影響に関する検討[J]. *画像電子学会誌*, 2009, 38: 156-163
- [60] 薛雨丽, 毛峡, Caeleu C D, et al. 遮挡条件下的鲁棒表情识别[J]. *北京航空航天大学学报*, 2010, 36(4): 429-433
- [61] Lajevardi S M, Hussain Z M. Higher order orthogonal moments for invariant facial expression recognition[J]. *Digital Signal Processing*, 2010, 20(6): 1771-1779
- [62] Tan Hua-chun, Zhang Yu-jin, Chen Hao, et al. Person-independent expression recognition based on person-similarity weighted expression feature[J]. *Journal of Systems Engineering and Electronics*, 2010, 21(1): 118-126
- [63] Sorci M, Antonini G, Cruz J. Modelling human perception of static facial expressions[J]. *Image and Vision Computing*, 2010, 28(5): 790-806
- [64] Yang G, Lin Y, Bhattacharya P. A driver fatigue recognition model based on information fusion and dynamic Bayesian network[J]. *Special Issue on Intelligent Distributed Information Systems*, 2010, 180(10): 1942-1954
- [65] 卢官明, 李晓南, 李海波. 新生儿疼痛面部表情识别方法的研究[J]. *光学学报*, 2008, 28(11): 2109-2114
- [66] Park Sungsoo, Kim Daijin. Subtle facial expression recognition using motion magnification[J]. *Pattern Recognition Letters*, 2009, 30(7): 708-716