

人脸表情识别研究的新进展

刘晓旻 谭华春 章毓晋

(清华大学电子工程系, 北京 100084)

摘要 人脸表情识别 (facial expression recognition, 简称 FER) 作为智能化人机交互技术中的一个重要组成部分, 近年来得到了广泛的关注, 涌现出许多新方法。本文综述了国内外近 4 年人脸表情识别 (FER) 技术的最新发展。首先, 介绍了 FER 系统的组成: 人脸检测、表情特征提取和表情分类, 并详细叙述了其中表情特征提取和表情分类的方法。然后, 对目前广泛应用的人脸表情数据库进行了介绍, 并在此基础上对当前一些 FER 系统的性能进行了比较分析。最后, 对 FER 领域的研究现状和挑战给予了评述, 对 FER 可能的发展方向进行了讨论。

关键词 人脸表情识别 表情特征提取 表情分类 人脸表情数据库

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2006)10-1359-10

New Research Advances in Facial Expression Recognition

LIU Xiao-min, TAN Hua-chun, ZHANG Yu-jin

(Department of Electronic Engineering, Tsinghua University, Beijing 100084)

Abstract As an important part of the technology for human-machine interface, facial expression recognition (FER) have drawn much attention recently and numerous methods have been proposed. In this paper, we present the up to date development of this area in recent years. First, the three steps of the FER system are introduced: face detection, expressional feature extraction and expression classification. Second, the methods of feature extraction and expression classification are detailed in different categories. Then, we also introduce the facial expression databases which are widely used at present. Based on these databases, a comparison of the performances of several FER systems is presented. At last, we demonstrate the state of the FER technique and the possible challenges, and provide some advice about the current of FER development.

Keywords facial expression recognition, expressional feature extraction, expression classification, facial expression database

1 引言

表情是人类用来表达情绪的一种基本方式, 是非语言交流中的一种有效手段。人们可通过表情准确而微妙地表达自己的思想感情, 也可通过表情辨认对方的态度和内心世界。关于表情传递信息的作用, 心理学家 Mehrabian^[1] 给出了一个公式:

感情表露 = 7% 的言词 + 38% 的声音 + 55% 的面部表情。

人脸表情识别 (FER) 所要研究的就是如何自动、可靠、高效地利用人脸表情所传达的信息。

人们对表情识别的研究可以追溯到 20 世纪 70 年代, 早期主要集中在从心理学和生物学方面进行研究和分析。Darwin^[2] 首先揭示了表情在不同性别, 不同种族的人群中的一致性。Ekman 和 Friesen^[3] 提出面部表情编码系统 (FACS), 用 44 个运动单元 (AU) 来描述人脸表情变化, 并定义了 6 种基本情感类别: 惊奇、恐惧、厌恶、愤怒、高兴、悲伤。这一系统得到了广泛的认同, 并成为后来很多表情识

基金项目: 教育部高等学校博士学科点专项科研项目 (RFDP-20020003011); 国家自然科学基金项目 (NNSF60573148)

收稿日期: 2005-09-07; 改回日期: 2005-11-01

第一作者简介: 刘晓旻 (1981 ~), 女。2006 年于清华大学电子工程系获硕士学位, 现在美国攻读博士学位。主要研究方向为图像分析。

E-mail: liuxiaomin99@mails.tsinghua.edu.cn

别研究工作的基础。人脸表情识别有广泛的应用前景,例如在多模式人机交互界面(MMHCI)中,表情与声音、视线、体态等结合起来可以获得更高效、更人性化的人机交流。另外,在如面部神经瘫痪诊断,人脸图像实时传输,人脸图像合成与动画,智能机器人,智能监控等多个领域,表情识别都有着广泛的应用。

在已有的文献中,Pantic^[4]和 Fasel^[5]的综述分别从不同的角度,系统地总结了2002年之前的FER技术和方法。本文主要针对2002年以后FER研究的进展,对诸多的研究工作和热点进行总结和归纳,为前述两篇综述提供一些新的补充,并使读者了解到FER技术目前的发展趋势与方向。

2 人脸表情识别的系统概述

一个人脸表情识别系统一般包括3个环节,即人脸检测、特征提取、表情分类,如图1所示。建立一个FER系统,第1步需要对人脸进行检测与定位,这一环节的研究实际上已成为一个独立的方向,读者可见文献[6];第2步从人脸图像或图像序列中提取能够表征输入表情本质的信息,在提取特征数据的过程中,为了避免维数危机,可能还需要特征降维、特征分解等进一步处理;第3步分析特征之间的关系,将输入的人脸表情分类到相应的类别,如AU组合或基本情感类别。

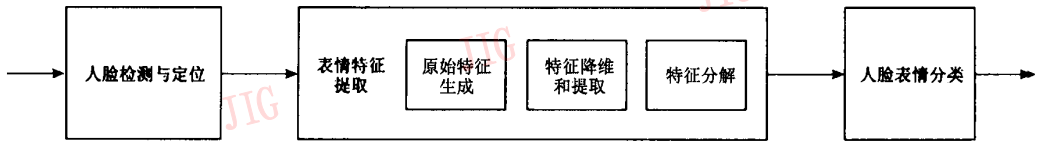


图 1 人脸表情识别系统

Fig. 1 Facial expression recognition system

3 表情特征提取的方法

表情特征提取是FER系统中最重要的部分,有效的表情特征提取工作将使识别的性能大大提高。通过对大量文献的总结,可知好的表情特征提取结果应该具备以下几个条件:

- (1)完整的表示出人脸表情的本质特征;
- (2)去除噪声、光照及其他与表情无关的干扰信息;
- (3)数据表示形式紧凑,避免过高的维数;
- (4)不同类别表情的特征之间有较好的区分性。

要得到满足这些条件的表情特征,特征提取的过程可能需要数个步骤来完成。首先,利用某种形式的信息来获得表情的原始特征,如特征形状与几何关系,局部纹理,光流场等,这一步骤称为原始特征获取。然而,这些原始特征一般都存在信息冗余,维数过高,区分性不够等问题。为了能够更有效地表征输入人脸表情的本质,需要对原始特征数据进行一些后处理,如特征降维和提取,特征分解等,以降低维数,去除干扰因素,得到对分类更为有利的特征数据。一些原始特征获取,特征降维和特征分解的方法如表1所示。

表 1 表情特征提取的3个步骤及方法例举

Tab. 1 The three steps of expressional feature extraction and examples of methods

| 原始特征生成 | 特征降维和提取 | 特征分解 |
|---|--|--|
| 几何特征:基准点 ^[7,8] 外貌特征:Gabor ^[13,14,48] ,HLAC ^[16] ,LBP ^[17] 混合特征:AAM ^[24-26] 序列特征:光流 ^[34,35] ,运动单元 ^[46] | PCA ^[9] ,LDA ^[19] ,ICA ^[12] ,CDA ^[39] 排序 PCA + LDA ^[40] MPPCA ^[25] | 高阶奇异值分解 ^[24] 双线性分解 ^[26] |

3.1 原始特征生成的方法

3.1.1 针对静态图像

- (1) 基于几何特征的方法

Pantic^[7]在以前工作^[8]的基础上改进并完善了一种自动提取脸部器官和轮廓基准点的方法。采用多检测器的方法,从正面人脸提取19个特征点,从

侧面人脸提取 10 个特征点。通过一致性检验为每个提取的点赋予一个确定性因子 CF (certainty factor), 比较 CF 来从多个检测器的信息中选择最精确的检测结果。

几何特征的识别效果对基准点提取的准确性要求较高, 在图像质量低和背景复杂的情况下难以实现。同时几何特征的提取忽略了脸部其他部分的信息如皮肤的纹理变化等, 因此在识别细微变化的表情时区分性不够高。

(2) 基于外貌特征的方法

外貌 (appearance) 特征泛指使用全部人脸图像像素的特征, 反映了人脸图像底层的信息。基于灰度图像的方法最简单, 由于灰度像素受到光照和噪声影响较大, 因此需要较好的预处理方法, 同时结合 PCA 和 DCT 变换等方法进行进一步提取^[9, 10]。

外貌特征提取最主要的方法是基于局部特征的方法, 利用一组滤波器对图像进行滤波, 结果可以反映局部像素之间的关系 (梯度、相关性、纹理等)。近来, 基于 Gabor 小波的方法被广泛应用于人脸表情的特征提取^[11-14], 它能够检测多尺度、多方向的纹理变化, 同时受光照变化影响较小, Donato^[12] 曾比较了几种方法识别脸部 AU 的性能, 结果表明 Gabor、ICA 优于其他的方法。Wen^[14] 在一系列手工标定的局部小区域提取平均 Gabor 小波系数作为纹理特征, 同时引入了一种在人脸合成中使用的基于比例图的方法来对纹理提取区域进行预处理, 以降低不同人脸差异和光照变化引起的人脸反照度不均的影响。另外, 一种类似 Harr 小波的矩形特征也被应用于表情特征的提取^[15], 它可以对整幅图像进行快捷的计算, 从而获得较好的实时效果。局部特征提取方法还有基于边缘, 高阶局部自相关 (HLAC), 局部二元模式等。Shinohara^[16] 对人脸图像提取共 35 种主要的高阶局部自相关特征, 同时与 Fisher 加权映射的方法相结合, 从而增强一些重要特征区域, 削弱不重要区域。Feng^[17] 使用局部二元模式 (LBP) 来提取脸部图像的纹理特征, 对脸部区域分块计算 LBP 直方图, 并连结起来作为表情识别的特征。Buciu^[18] 采用非负矩阵分解和局部非负矩阵分解来提取图像特征, 并得出局部非负矩阵分解的识别结果优于 PCA 方法的结论。

其他可归为外貌特征的方法还有: 基于脸部不对称性的方法^[19], 基于不变矩的方法^[20, 21]。Mitra^[19] 对左右两边人脸图像提取密度的差异和边

缘方向的相似性组成所谓的不对称脸 (asymmetry face), 并认为这一脸部的不对称性在表情识别中也具有很强的区分能力。Zhu^[20] 定义了包含脸部重要特征的 7 个区域, 然后对每个区域计算 4 种改进的不变矩, 并定义一个尺度因子 α 来调节和平衡各个区域的特征值强度。

(3) 基于混合特征的方法

几何特征能够简洁地表示出人脸宏观的结构变化, 而另一方面外貌特征侧重于提取局部的细微变化, 一些研究者将两者结合起来, 用混合的特征进行表情识别。如文献 [13]、[14] 中将几何特征与局部点或局部区域的 Gabor 小波系数结合起来作为表情特征。Zhang^[22] 提出用多种特征融合的方法进行表情识别, 所使用的特征提取方法包括: 基于特征点的几何特征提取, 法令纹 (nasolabial fold) 检测, 前额区域边缘检测, 后两者均为脸部瞬时变化出现的特征, 如图 2 所示。

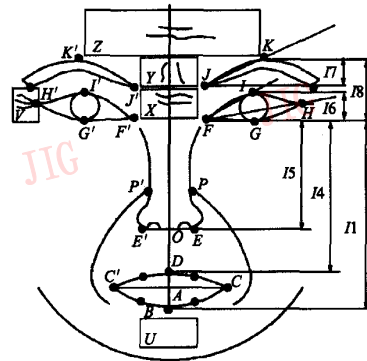


图 2 特征点的几何关系 (方形区域表示皱纹)

Fig. 2 The geometrical relationship of feature points (where the rectangles represent the furrows and wrinkle)

基于混合特征的一类重要方法是基于模型的方法。这类方法用模型的方式来描述人脸的结构, 几何模型 (点模型) 也是一类简单的模型方法。活跃外貌模型 (AAM)^[23] 是目前广泛应用的一种人脸特征提取的方法^[24-27]。AAM 方法结合形状和纹理信息建立对人脸的参数化描述, 然后再用 PCA 进行降维。Chang^[28] 使用 Hu^[29] 提出的活跃小波网络 (AWN) 进行人脸的校准。与 AAM 中的纹理模型不同, AWN 使用 Gabor 小波网络的表示方法来对纹理的变化进行建模, 这一方法对部分阻挡和一些光照的变化有较好的鲁棒性。相对于基于外貌特征的方法, 基于模型的方法可以得到更为可靠的人脸特征

参数,但同时它也有计算较为复杂,而且初始点获取困难等缺点。

3.1.2 针对图像序列

基于图像序列的表情特征提取有两类方法,一类仅利用帧间信息进行特征的跟踪和定位,可称为基于帧内特征的方法,另一类不仅需要跟踪和定位,而且还利用帧间信息来作为表情识别的特征,可称为基于序列特征的方法。

(1) 基于帧内特征的方法

在帧内提取特征的方法与静态图像使用的方法大致相同,因此主要讨论特征跟踪的方法。

① 基于特征点跟踪的方法

基于特征点的跟踪通常只将特征点选在灰度变化较大的区域,例如眼角点,嘴角点等。这些点便于跟踪,从而可以进一步得到脸部特征的位移或形变信息。Zhang^[22]分别结合 Kalman 滤波和红外(IR)设备定位瞳孔两种方法来决定预测的特征位置,然后在预测位置附近匹配两点之间 Gabor 核的相似度以检测特征的位置。由于 IR 设备的引入,该方法对大的头部运动也能得到鲁棒的跟踪结果。Bourel^[30]手工选择 12 个特征点并进行跟踪,然后用基于状态的特征选择方法来提取表情特征。Pardàs^[31]提出一种基于活跃轮廓的跟踪方法,可以提取人脸的特征并转换为 MPEG 编码方案中的人脸动画参数(FAPs)。

② 基于模型跟踪的方法

不同于特征点的跟踪,基于模型的跟踪针对整个人脸。这里模型可以是 2 维或 3 维的,但大多数模型都需要较复杂的计算。Braathen^[32]用 3 维模型跟踪自然场景中头部的转动,然后将人脸图像用 3 维变形的的方法转换到规范的姿势,从而更方便地进行表情识别。

Huang^[33]使用多分辨率的 3 维可变形人脸模型来对人脸建模,并采用了多层次的跟踪方法。全局形变的跟踪由底层的数千个节点来完成,对于局部的纹理变化,用非刚性形状配准方法来进行获取。该方法适用于 3 维人脸深度扫描数据,可以对细微的表情细节进行提取和跟踪。

(2) 基于序列特征的方法

① 基于光流的方法

光流反映了帧间运动的重要信息,在表情的动态分析中有较多的应用^[34-36]。基于光流方法的缺点在于,光照不均和脸部非刚性运动等因素将会影

响特征提取的结果。金辉等人^[34]在人脸面部物理-几何结构模型的基础上,提取面部表情特征区域,通过光流估计计算其运动场,进而计算特征流向量,然后把一组图像序列的运动向量组成运动特征序列,对表情的运动进行分析。

② 基于模型的方法

这类方法采用模型来跟踪人脸,并根据跟踪得到的模型参数和一些前后帧的信息来表示人脸的运动。Tao 和 Huang^[37]提出一种基于 Bezier Volume 的跟踪系统,在第 1 帧用交互式的方法确定眼角和嘴角等标志性特征的位置,然后调整网状的 3 维模型来匹配选定的特征位置。在跟踪之后,用一些定义好的脸部运动单元(motion unit)来表示运动特征。这些运动单元不仅表示了人脸区域的运动有无,还表示了运动的方向和强度。应用这一模型及运动单元来提取特征的方法有文献[14]、[46]、[54]。Gokturk^[38]使用 3 维的可变形人脸模型来进行人脸跟踪,然后用得出的形状向量及其在帧间的差异来作为表情识别的特征。

3.2 特征降维和提取的方法

由于特征表示方法尤其是外貌特征的空间维数通常非常巨大,因此需要通过一些映射或变换将它们转换到低维的子空间来表示。这样不仅可以使特征的维数明显降低,同时这些低维空间特征的有效性也将得到提高。常用的特征降维及提取的方法有主元分析(PCA),线性判别分析(LDA),ICA 等。这些方法在进行特征提取时各有优势,如 PCA 提取了最有代表性的特征,可以有效地消除冗余,降低维数,但它没有考虑不同类别数据之间的区分性。而 LDA 则通过最大化数据的类间离散度和最小化类内离散度来选择合适的投影方向,侧重于寻找具有最大分辨力的方向。近来,关于特征降维和提取的方法,也有一些新的研究。

Chen^[39]在 LDA 的基础上提出一种基于聚类的判别分析,针对类内存在多种子类的情况,首先使用模糊 C 均值的方法进行聚类,然后以分离不同类的子类为目标,寻找最佳的投影方向。该方法对同一表情由于光照等变化产生类内多种聚类的情况,有较好的识别效果。

Dubousson^[40]首先对特征用 PCA 方法进行分析,然后按照主元对某一识别任务的重要性将它们进行排列,采用前向逐步选择的方法选择和保留最有区分性的 k 个主元方向,最后使用 LDA 对排列后

的子空间进行计算,产生 $C-1$ 维(C 是分类的类别个数)的判别子空间进行分类。该方法称为排序 PCA + LDA 方法,可以对不同的识别任务建立最优的子空间。

Zalewski^[25] 引入了混合概率 PCA 的方法,与传统的 PCA 方法不同,概率 PCA 为 PCA 定义了一个概率模型,对于主元方向以外的信息并不是直接丢弃,而是作为高斯噪声进行估计,通过最大似然解方法来估计模型参数而得到最佳概率模型。该方法可以解决较大姿态变化带来的特征空间中非线性的分布,并可以结合贝叶斯判别方法进行表情识别。

3.3 特征分解的方法

人脸图像包含了丰富的信息,对不同的识别任务来说,所利用的信息也各不相同。人脸检测寻找的是人脸图像共有的一致性,人脸识别需要利用表示人脸个体差异的信息,而表情识别则需要表示各种表情之间差异的信息,对一种识别任务有利的信息有可能反而对其他识别任务造成干扰。近来,一种新的解决思路是把人脸不同的因素如表情因素和个体因素分离开来,使得识别能够在相应的子空间中进行,避免其他因素的干扰。

Aboud^[26] 提出用双线性模型的方法来进行人脸的外貌分解,首先使用 AAM 的方法获取人脸的特征参数,然后通过双线性分解模型的方法将人脸特征参数分解为与个体无关的表情分量以及与个体相关的线性映射的乘积。实验结果表明,该方法在训练样本数较少的情况下识别结果优于 LDA 方法。

Wang^[24] 将高阶奇异值分解(HOSVD)的方法用于脸部表情的分解。与双线性方法不同,HOSVD 是一种多因素的分析方法,在人脸识别和运动分析中都有应用。该方法将不同人、不同表情的图像用一个 3 阶的张量来表示,3 阶分别表示个体、表情和特征(用 AAM 提取特征)。对张量用 HOSVD 方法进行分解,得到个体子空间、表情子空间和特征子空间,如图 3 所示。训练时从包含 7 种表情的图像集中得到人的个体子空间和表情子空间模型。这类方法由于分解了人脸各种不同的因素,不仅可以较好地应用于表情识别,也可以方便地进行人脸表情的合成,甚至增加新的因素来满足不同光照和不同角度的合成要求。

基于特征分解的方法在分类过程中需要对已知表情类别的样本库进行遍历搜索。首先假设测试样本的个体因素来自第 1 个训练的个体,分解得到相

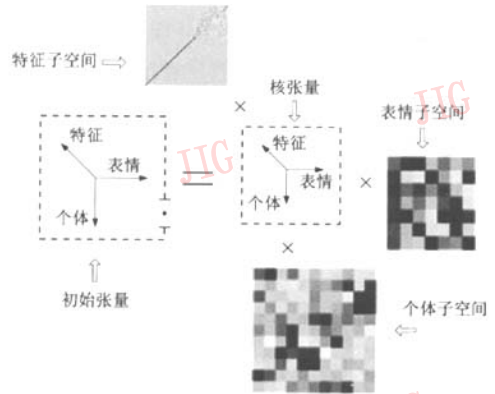


图 3 脸部表情分解到人脸子空间、表情子空间和特征子空间
Fig. 3 Facial expression decomposition to person subspace, expression subspace, feature subspace

应的表情分量,计算测试样本表情分量与该训练样本表情分量的相似性(如余弦距离或欧式距离),对所有训练个体重复这一过程,最终将测试人脸分类到最相近的表情类别。

4 表情分类方法

表情分类指定义一组类别,并设计相应的分类机制对表情进行识别,归入相应类别。如按照脸部动作分类(FACS),将脸部的动作分类到 44 个 AUs (action units);或按照感情分类,将表情分类到 Ekman 和 Friesen 定义的 6 种基本感情。根据是否利用时间信息来进行分析,人脸表情分类的方法可分为基于空间分析的方法和空时结合的方法两类。

4.1 空间分析方法

4.1.1 基于专家规则(Expert rules)的方法

这类方法从人的观察角度出发来制定一些规则,对特征提取得出的参数,判断它符合哪一类规则,将其归入相应的类别。Pantic^[7,8] 利用 AU 编码来描述输入的表情,然后与每一种表情类别的 AU 编码描述进行比较,将其分类到最相近的一类基本表情中。使用这类方法的还有文献[41]、[42]。基于专家规则的方法可以更精确地描述人脸表情,从而更好地识别混合表情,同时由于它对表情进行了编码,因此便于进行人脸表情的合成。

4.1.2 基于机器学习的方法

(1) 人工神经网络

人工神经网络在静态图像的人脸表情识别中有

很多运用^[9,13,43,44]。Gueorguieva^[43]使用多层感知的神经网络来进行表情识别,训练并测试了 4 种网络,得出 S 形函数和径向基函数的神经单元混合能较好地适合于前馈神经网络的结论。Ma^[9]在全脸图像中用 2 维离散余弦变换进行特征检测,然后用一个结构性单隐层的前馈神经网络作为表情分类器,效果比传统的神经网络好。神经网络方法的缺点在于,当识别很多无限制的混合表情时,对分类器的训练将会比较困难。

(2) 贝叶斯分类

Sebe^[45]首先使用朴素(Naive)贝叶斯分类器进行表情的识别,并证明将特征分布的假设由高斯分布改为柯西分布可以提高分类的性能。朴素贝叶斯分类器的缺点在于,它假设特征之间是相互独立的,而事实上人脸的表情发生过程中,各特征的变化并不是完全独立的,而是存在着相互制约的关系。Cohen^[46]在柯西分布的贝叶斯分类器基础上,使用高斯树状分类器对特征之间的从属性进行建模。并且讨论了如何使用无标签的样本来训练贝叶斯网络分类器,从而在有标签样本数较少的情况下提高分类器的性能^[47]。Zhang^[22]采用一个 3 层的贝叶斯网络对脸部表情和脸部 AU 之间的因果关系进行建模,底层为可观察数据层,包括一些基本的特征数据,如眼、鼻、嘴、皱纹等等,中间层(即隐含层)为 AU 层,上层为表情层。同时结合 HMM,从时间和空间关系上同时对表情进行建模。

(3) 其他

支持向量机是在统计学习理论的基础上发展起来的一种分类方法,在解决小样本、非线性和高维问题上有很多优势。目前支持向量机是机器学习中的应用最多的分类器之一,近年来也被应用于表情识别中^[10,48]。Adaboost 方法将多个弱分类器结合起来训练形成强分类器,不同分类器针对不同的特征,通过训练可以达到特征选择的目的,在模式识别领域如图像检索和人脸检测中都有成功的应用。Bartlett^[48]使用 Adaboost 选择特征与 SVM 分类相结合的 AdaSVM 方法进行分类,取得较好的识别结果。Wang^[15]用 Adaboost 方法训练查找表型的弱分类器进行表情识别,在自己收集的数据库上的测试结果略优于 SVM。

Guo^[49]提出特征选择线性规划的方法,将模式分类的问题转化为最小化平均分类错误的 1-范数的问题,从而可以用线性规划的方法来进行求解。

而特征选择则通过限制分类面法向量的分量个数来完成。在实验中该方法可以用很少的特征数目达到接近于 SVM 的识别正确率。

4.2 空时分析的方法

在以前的文献中,空时结合的分析方法还包括基于规则的推理以及循环神经网络等,但近几年的文献中,只有隐马尔科夫模型(Hidden Markov Model)被普遍采用^[20,32,35,46,50-52]。

Muller^[51]运用伪 3 维的隐马尔科夫模型(P3DHMMs)进行动态人脸表情的分析。对序列中的每一幅图像用一个伪 2 维的隐马尔科夫模型(P2DHMM)进行建模,再用 1D-HMM 对时间状态进行建模,最后得到表示完整表情动作的 P3DHMM。Yeasin^[52]提出两步分析法,首先将一系列线性分类器用于 frame level,每次观察结果结合起来产生时间信号,然后用这一时间信号作为训练数据训练离散 HMMs。

用 HMM 方法进行识别需要确定其初始和结束的状态,因此一般用于单独的表情序列或分割好的表情序列。Cohen^[46]提出多层次的 HMM,第 1 层为分别针对 6 种表情的 6 个 HMM 模型;第 2 层为表示 6 种表情之间状态变化的 Markov 模型。将单个表情 HMM 的状态输出联合起来作为高层 Markov 模型的输入,并通过训练得到 6 种表情之间的转移概率,从而可以自动将视频分割为不同的表情段进行分析。

5 人脸表情数据库及算法比较

目前在人脸表情识别研究中使用最广泛的是 CMU 的 Cohn-Kanade 数据库^[53],在很多研究中被列为算法比较的标准。Cohn-Kanade 数据库是基于 AU 编码的数据库,含有 210 个对象的大约 2 000 个图像序列,每个人有一系列的脸部活动,包括单个 AU 或者 AU 组合。其次,日本 ATR 的女性表情数据库(JAFFE)^[11]也得到较多使用。JAFFE 是以 7 种基本表情为基础的数据库,包括 10 位日本女性,每种表情有大约 3、4 幅图像,总共 213 幅图像。此外,还有一些数据库也可应用于人脸表情识别,如 Ekman 和 Friesen 的人脸表情数据库, Yale 人脸数据库, Purdue 的 AR 人脸数据库, CMU 的 PIE 数据库等。

随着人脸表情识别研究的深入,出现了一些比较各种人脸表情识别算法的文献^[54,55]。例如,

Sebe^[54]建立了一种在自然环境下的真实人脸表情数据库,然后测试并比较了一系列基于机器学习理论的分类器,如贝叶斯分类器,决策树,SVM,KNN等。结果表明 KNN 的识别效果最好,但同时具有计算速度较慢并且空间复杂度较高的缺点。Littlewort^[55]系统比较了几种表情识别中的机器学习的方法,包括 SVM, Adaboost, LDA, 获得最佳识别效果的方法为基于 Gabor 小波特征的 Adaboost 与 SVM 结合的方法。

由于目前很多的人脸表情识别系统所报告的结果都是在不同的数据库、不同的实验设置上进行的,

因此,对哪个系统更为优越,很难给出一个统一的评价。表 2 给出了使用 Cohn-Kanade 数据库和 JAFFE 数据库的一些算法及其结果,对每一种算法列出了相应的识别任务(AU 或基本表情等),数据类型(静态、序列或序列中抽取若干帧),数据量,以及训练/测试比等实验的客观条件。测试结果中①表示该实验的测试集中的人不在训练集中出现,②则表示该实验测试集中的人可能在训练集中出现。为了对近年来识别方法的进展提供一个纵向的比较,表中也列出了一些经典的方法以及他们的测试结果参见文献[35]、[11]。

表 2 一些 FER 系统及其识别结果列举

Tab.2 Several examples of FER systems and their recognition results

| 数据库 | 特征提取方法 | 分类方法 | 识别任务 | 数据类型 | 测试结果(%) | | 数据量 | 训练/测试比 |
|---------------------------------|--|----------------------------|-----------------|--------------------|--------------------|--------------|--------------|----------------|
| | | | | | ① | ② | | |
| Cohn-Kanade | 光流、特征点跟踪、边缘检测 3 种方法 ^[35] | HMM | 3 种上部 AU | 序列 | 85 ~ 92 | — | 100 ~ 270 | 5:8 ~ 4:5 |
| | | | 7 种下部 AU | 序列 | 81 ~ 92 | — | | |
| | 红外照相机检测虹膜,恢复眼眉区域形状参数 ^[10] | SVM 分类器 | 5 种上部 AU 和中性 | 不详 | 81.22 | — | 25 人 | 24:1 |
| | Boosting 方法定位特征,提取 Gabor 幅值 ^[55] | SVM 分类器 | 7 种上部 AU | 每序列抽取二帧:开始,峰值 | 92.9 | — | 90 人, 625 帧 | — |
| | | | 7 种表情 | 同上 | 93.3 | — | | |
| | PBVD 跟踪人脸,提取运动单元(Motion Units) ^[46] | 贝叶斯网络 | 7 种表情 | 静态逐帧 | 73.22 | — | 53 人 | 1:4 重复 |
| | PBVD 跟踪人脸;提取几何特征 + 局部纹理特征 ^[14] | 混合高斯模型(GMM),在线的基于 EM 的适应算法 | 4 种表情 | 静态逐帧 | 75.4 | 87.6 | 47 人, 2981 帧 | ① 6:4 ② 1:1 |
| | 局部脸部不对称性的表示 - 不对称脸 ^[19] | 有序前向特征选择算法与线性判别分析(LDA) | 3 种表情(高兴、愤怒、沮丧) | 每序列抽取 3 帧:开始,中间,峰值 | FNR 3.6 FPR 1.8 | — | 55 人 | 6:5 重复 |
| ACM 跟踪,转换为 FAPs ^[31] | 半连续 HMM | 6 种表情 | 序列 | 84 | — | 90 人 | 89:1 重复 | |
| | | 6 种表情 + “说话” | 已分割序列 | 81 | — | | | |
| 连续序列 | 64 | | — | | | | | |
| 图像序列的光流 ^[52] | 两步法:kNN + HMMs | 6 种表情 | 序列 | 90.9 | — | 97 人, 488 序列 | — | |
| JAFFE | 手工标定 34 个点 + 18 个 Gabor 滤波器 ^[11] | LDA | 6 种基本表情 + 中性 | 静态显著表情 | 75 | 92 | 9 个人, 193 帧 | ① 9:1 ② 8:1 |
| | HOLA 特征 + Fisher 加权映射 ^[16] | Fisher 线性判别 | 6 种基本表情 + 中性 | 静态显著表情 | 69.4 | — | 9 个人 | 8:1 重复 |
| | 局部二元模式(LBP) ^[17] | 改进的模版匹配法, 两步分类 | 6 种基本表情 + 中性 | 静态显著表情 | 77 | — | 9 个人 | 8:1 重复 |
| | 手工标定的 34 个特征点 + 18 个 Gabor 滤波器 ^[49] | 特征选择线性规划法 | 6 种基本表情 + 中性 | 静态显著表情 | — | 91.0 | 10 个人 | 9:1 重复 |

6 讨论与总结

通过前面几节的介绍,可以看出人脸表情识别方面的研究取得了很大进展,但从识别系统的性能来看,与实用化还有一定的距离。目前看来,FER系统的性能主要受到以下几个方面的限制:

(1) 特征提取的方法存在一些缺陷,很多模型的方法需要手工介入来标定一些点或区域,外貌特征的方法不需要手工帮助,但提取出的信息往往不够可靠而且易受干扰;

(2) 某些表情易于识别,而某些表情不易识别,心理学方面的研究认为,最容易辨认的表情是快乐、痛苦,较难辨认的是恐惧、悲哀,更难辨认的是怀疑、怜悯,在 FER 系统中也面临着对某些类别的表情识别率偏低的情况;

(3) 目前的研究大多数都针对特定条件下采集的图像或者图像序列。但要做到对任意采集的图像或图像序列都适应,还需要人脸检测等其他领域的发展。

虽然受到一些限制因素的影响,国内外研究者仍提出了很多思路,力图在这些方面有所突破。对本文介绍的一些方法进行总结,可以得出近来 FER 研究的几个热点,也是未来可能的发展方向:

(1) 由于人脸表情变化表现在图像上会产生多种不同的变化,如五官位置、形状,以及细微的皮肤纹理、阴影等,因此将多种特征混合起来,可以最大限度地利用表情变化产生的信息,更完整地表示表情,混合法将成为特征获取中的一个重点;

(2) 由于模型的方法往往需要手工的帮助,因此外貌特征就被较多地应用于自动的 FER 系统。由此带来的问题是特征维数可能很高,而且特征的有效性不高。因此特征降维、提取和选择等方法的研究也成为重点;

(3) 近年随着机器学习领域的发展,越来越多的机器学习方法被用于表情的识别和分类,并取得了较好的结果,这也应是未来识别方法的主流。但同时人的表情是有一定的心理学规律可循的,是否可以将心理学和生物学知识与机器学习的方法结合起来提高分类的性能,这是有待研究的问题;

(4) 在目前的人脸表情识别系统中,往往是对比较夸张的表情进行分析和识别。但在实际的生活,很多表情可能是混和的表情,或者并不是很夸张

的表情。对混和表情、表情的强度方面,目前的研究还非常的缺乏。已经有研究针对表情的强度方面做了一些尝试,但研究仍然处于初级阶段;

(5) 人脸表情会受到多种因素的影响,如光照、姿态变化,脸部阻挡等。针对脸部出现阻挡的情况,有一些初步的研究^[22, 56],但与实际情况还有一定距离。在目前的研究中,针对光照和姿势的变化,一般都采用 3 维人脸建模的方法来解决。但目前 3 维人脸建模的技术尚不能达到自动建模的要求。同时,精度也受到一定的限制。另外,特征分解是解决这类问题的一种新方法,但这方面的研究刚刚出现,有待进一步发展。

参考文献 (References)

- Mehrabian A. Communication without words[J]. *Psychology Today*, 1968, 2(4): 53 ~ 56.
- Darwin C. *The Expression of the Emotions in Man and Animals*[M]. London: J. Murray, 1872.
- Ekman P, Friesen W V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*[M]. Palo Alto: Consulting Psychologists Press, 1978.
- Pantic M, Rothkrantz L. Automatic analysis of facial expressions: the state of the art[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(12): 1424 ~ 1445.
- Fasel B, Luetttin J. Automatic facial expression analysis: A survey [J]. *Pattern Recognition*, 2003, 36(1): 259 ~ 275.
- Yang M, Kriegman D J, Ahuja N. Detecting faces in images: A survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(1): 34 ~ 58.
- Pantic M, Rothkrantz L. Facial action recognition for facial expression analysis from static face images[J]. *IEEE Transactions on Systems, Man and Cybernetics-Part B*, 2004, 34(3): 1449 ~ 1461.
- Pantic M, Rothkrantz L. Expert system for automatic analysis of facial expression[J]. *Image Vision Computing*, 2000, 18(11): 881 ~ 905.
- Ma L, Khorasani K. Facial expression recognition using constructive feedforward neural networks[J]. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 2004, 34(3): 1588 ~ 1595.
- Kapoor A, Qi Y, Picard R W. Fully automatic upper facial action recognition[A]. In: *Proceedings IEEE International Workshop on Analysis and Modeling of Faces and Gestures [C]*, Nice, France, 2003: 195 ~ 202.
- Lyons M, Budynek J, Akamastu S. Automatic classification of single facial images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(12): 1357 ~ 1362.
- Donato G, Bartlett S, Hager C, et al. Classifying facial actions[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(10): 974 ~ 989.

- 13 Tian Y, Kanade T, Cohn J. Evaluation of Gabor wavelet-based facial action unit recognition in image sequences of increasing complexity [A]. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition [C], Washington, DC, USA, 2002: 26 ~ 30.
- 14 Wen Z, Huang T. Capturing subtle facial motions in 3d face tracking [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Nice, France, 2003, 2: 1343 ~ 1350.
- 15 Wang Y, Ai H, Wu B, *et al.* Real time facial expression recognition with adaboost [A]. In: Proceedings of International Conference on Pattern Recognition [C], Cambridge, UK, 2004, 3: 926 ~ 929.
- 16 Shinohara Y, Otsu N. Facial expression recognition using fisher weight maps [A]. In: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition [C], Seoul, Korea, 2004: 499 ~ 504.
- 17 Feng X. Facial expression recognition based on local binary patterns and coarse-to-fine classification [A]. In: Proceedings of International Conference on Computer and Information Technology [C], Wuhan, China, 2004: 178 ~ 183.
- 18 Buciu I, Pitas I. Application of non-negative and local nonnegative matrix factorization to facial expression recognition [A]. In: Proceedings of International Conference on Pattern Recognition [C], Cambridge, UK, 2004, 1: 288 ~ 291.
- 19 Mitra S, Liu Y. Local facial asymmetry for expression classification [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], Washington, DC, USA, 2004, 2: 889 ~ 894.
- 20 Zhu Y, DeSilva L C, Ko C C. Using moment invariants and HMM in facial expression recognition [J]. Pattern Recognition Letters, 2002, 23(1-3): 83 ~ 91.
- 21 Jiang Lu, Zhang Pin-zheng, Shu Hua-zhong. Moment application to human facial expression recognition [J]. Journal of Southeast University (Natural Science Edition), 2004, 34(4): 557 ~ 560. [姜璐, 章品正, 舒华忠. 矩在面部表情识别中的应用 [J]. 东南大学学报, 2004, 34(4): 557 ~ 560.]
- 22 Zhang Y, Ji Q. Active and dynamic information fusion for facial expression understanding from image sequences [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 699 ~ 714.
- 23 Cootes T F, Edwards G J, Taylor C J. Active appearance models [A]. In: Proceedings of the 5th European Conference on Computer Vision [C], Freiburg, Germany, 1998, 2: 484 ~ 498.
- 24 Wang H, Ahuja N. Facial expression decomposition [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Nice, France, 2003, 2: 958 ~ 965.
- 25 Zalewski L, Gong S. Synthesis and recognition of facial expressions in virtual 3D views [A]. In: Proceedings of IEEE 6th International Conference on Automatic Face and Gesture Recognition [C], Seoul, Korea, 2004: 493 ~ 498.
- 26 Abboud B, Davoine F. Appearance factorization based facial expression recognition and synthesis [A]. In: Proceedings of International Conference on Pattern Recognition [C], Cambridge, UK, 2004, 4: 163 ~ 166.
- 27 Zuo Kun-long, Liu Wen-yao. Facial expression recognition using active appearance models [J]. Journal of Optoelectronics Laser, 2004, 15(7): 853 ~ 857. [左坤隆, 刘文耀. 基于活动外观模型的人脸表情分析与识别 [J]. 光电子激光, 2004, 15(7): 853 ~ 857.]
- 28 Chang Y, Hu C, Turk M. Probabilistic expression analysis on manifolds [A]. In: Proceedings of International Conference on Computer Vision and Pattern Recognition [C], Washington DC, USA, 2004, 2: 520 ~ 527.
- 29 Hu C, Feris R, Turk M. Real-time view-based face alignment using active wavelet networks [A]. In: Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures [C], Nice, France, 2003: 215 ~ 221.
- 30 Bourel F, Chibelushi C C, Low A A. Robust facial expression recognition using a state-based model of spatially-localized facial dynamics [A]. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition [C], Washington, DC, USA, 2002: 106 ~ 111.
- 31 Pardas M, Bonafonte A, Landabaso J L. Emotion recognition based on MPEG4 facial animation parameters [A]. In: Proceedings of IEEE Acoustics, Speech, and Signal Processing [C], Orlando, FL, USA, 2002, 4: 3624 ~ 3627.
- 32 Braathen B, Bartlett M S, Littlewort G, *et al.* An approach to automatic recognition of spontaneous facial actions [A]. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition [C], Washington, DC, USA, 2002: 231 ~ 235.
- 33 Huang X, Zhang S, Wang Y, *et al.* A hierarchical framework for high resolution facial expression tracking [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop [C], Washington, DC, USA, 2004: 22 ~ 22.
- 34 Jin Hui, Gao Wen. Analysis and application of the facial expression motions based on eigen-flow [J]. Journal of Software, 2003, 14(12): 2098 ~ 2105. [金辉, 高文. 基于特征流的面部表情运动分析及应用 [J]. 软件学报, 2003, 14(12): 2098 ~ 2105.]
- 35 Lien J. Automatic Recognition of Facial Expression Using Hidden Markov Models and Estimation of Expression Intensity [D]. Pittsburgh: The Robotics Institute, CMU, 1998.
- 36 Minamitani H, Hoshino Y, Hashimoto H, *et al.* Computerized diagnosis of facial nerve palsy based on optical flow analysis of facial expressions [A]. In: Proceedings of the IEEE International Conference of Engineering in Medicine and Biology Society [C], Cancun, Mexico, 2003: 663 ~ 666.
- 37 Tao H, Huang T. Explanation-based facial motion tracking using a piecewise Bezier volume deformation model [A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], Collins, CO, USA, 1999: 23 ~ 25.
- 38 Gokturk S B, Bouguet J Y, Tomasi C, *et al.* Model-based face tracking for view-independent facial expression recognition [A]. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition [C], Washington, DC, USA, 2002: 272 ~ 278.

- 39 Chen X, Huang T. Facial expression recognition: a clustering based approach [J]. *Pattern Recognition Letters*, 2003, 24 (9-10): 1295 ~ 1302.
- 40 Dubuisson S, Devoine F, Masson M. A solution for facial expression representation and recognition [J]. *Signal Processing: Image Communication*, 2002, 17(9): 657 ~ 673.
- 41 Matsugu M, Mori K, Mitari Y, *et al.* Subject independent facial expression recognition with robust face detection using a convolutional neural network[J]. *Neural Networks*, 2003, 16(5-6): 555 ~ 559.
- 42 Haneda K, Muraguchi T, Nakamura O. The recognition of facial expressions using expert system [A]. In: *Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering* [C], Montreal, Canada, 2003, 2: 1195 ~ 1198.
- 43 Gueorguieva N, Georgiev G, Valova I. Facial expression recognition using feedforward neural networks [A]. In: *Proceedings of the International Conference on Artificial Intelligence* [C], Las Vegas, NV, USA, 2003: 285 ~ 291.
- 44 Seyedarabi H, Aghagolzadeh A, Khanmohammadi S, *et al.* Recognition of six basic facial expressions by feature-points tracking using RBF neural network and fuzzy inference system [A]. In: *Proceedings of IEEE International Conference on Multimedia and Expo* [C], Taipei, Taiwan, China, 2004: 1219 ~ 1222.
- 45 Sebe N, Cohen I, Garg A, *et al.* Emotion recognition using a Cauchy naive Bayes Classifier [A]. In: *Proceedings of International Conference on Pattern Recognition* [C], Québec City, Canada, 2002, 1: 17 ~ 20.
- 46 Cohen I, Sebe N, Garg A, *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 Cohen I, Sebe N, Cozman F G, *et al.* Learning bayesian network classifiers for facial expression. recognition with both labeled and unlabeled data [A]. In: *Proceedings of International Conference on Computer Vision and Pattern Recognition* [C], Madison, Wisconsin, USA, 2003, 1: 595 ~ 604.
- 48 Bartlett M S, Littlewort G, Frank M, *et al.* Recognizing facial expression: machine learning and application to spontaneous behavior [A]. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* [C], San Diego, CA, USA, 2005, 2: 568 ~ 573.
- 49 Guo G D, Dyer C R. Learning from examples in the small sample case: face expression recognition [J]. *IEEE Transactions on System, Man and Cybernetics-Part B, Special Issue on Learning in Computer Vision and Pattern Recognition*, 2005, 35(3): 477 ~ 488.
- 50 Jin Hui, Cao Wen. Analysis and recognition of facial expression image sequences based on HMM [J]. *Acta Automatica Sinica*, 2002, 28(4): 646 ~ 650. [金辉, 高文. 基于 HMM 的面部表情图像序列的分析与识别 [J]. *自动化学报*, 2002, 28(4): 646 ~ 650.]
- 51 Muller S, Wallhoff F, Hulsken F, *et al.* Facial expression recognition using pseudo 3-D hidden Markov models [A]. In: *Proceedings of International Conference on Pattern Recognition* [C], Québec City, Canada, 2002, 2: 32 ~ 35.
- 52 Yeasin M, Bullot B, Sharma R. From facial expression to level of interest: a spatio-temporal approach [A]. In: *Proceedings of International Conference on Computer Vision and Pattern Recognition* [C], Washington, DC, USA, 2004, 2: 922 ~ 927.
- 53 Kanade T, Cohn J F, Tian Y. Comprehensive database for facial expression analysis [A]. In: *Proceedings of the Fourth International Conference of Face and Gesture Recognition* [C], Grenoble, France, 2000: 46 ~ 53.
- 54 Sebe N, Lew M, Cohen I, *et al.* Authentic facial expression analysis [A]. In: *Proceedings of International Conference on Automatic Face and Gesture Recognition* [C], Seoul, Korea, 2004: 517 ~ 522.
- 55 Littlewort G, Bartlett M, Fasel I, *et al.* Dynamics of facial expression extracted automatically from video [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Face Processing in Video* [C], Washington DC, USA, 2004: 80 ~ 80.
- 56 Buciu L, Kotsia I, Pitas I. Facial expression analysis under partial occlusion [A]. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* [C], Philadelphia, PA, USA, 2005, 5: 453 ~ 456.