

湖南大学

硕士学位论文

动态人脸表情识别技术研究

姓名：应伟

申请学位级别：硕士

专业：通信与信息系统

指导教师：邹北骥

20050508

## 摘 要

新一代人机交互界面的设计和情感智能是两项国际前沿性研究课题,在这两项课题中,一个关键的技术是如何获取人的内心情感。表情作为人的内心情感的主要表现方式,蕴涵了大量有关人的内心状态变化的信息。人脸表情自动识别技术因而受到了研究者的广泛关注。

结合心理学研究成果,本文首先建立了一个含 840 段视频序列的基本人脸表情数据库,其次针对表情的动态性特点,确定了两个表情识别平台接口,然后对目前比较成熟的表情区域定位方法、特征提取方法以及信息分类方法进行合理的选择构建了一个表情识别系统实例,最后在基本表情数据库上对其中四种基本表情进行测试,综合识别率达到了 91.7%。测试结果说明平台运行正常,可以为后续的表情识别系统的开发和改进提供良好的实验环境,并且构建的系统实例接近实际应用的要求,为将来表情识别系统应用化打下了一个基础。

表情特征提取是系统中比较核心的问题,由于应用环境的复杂性,如何得到稳定的、能够反映表情变化本质的信息是亟待解决的难题。本文对表情特征的有效归一化和合理描述两个部分进行了深入的分析,并将分析结果应用在了构建的系统实例中。

几何归一化处理对于表情特征信息的有效提取具有重要意义,常用的归一化方法由于基准特征的不稳定性容易造成误差。本文定量地分析了归一化误差对表情特征提取的影响,并提出加权优化匹配算法进行误差的矫正。算法以传统模板匹配原理为基础,根据各像素点的运动剧烈程度分配相应的匹配权值,实验证明误差得到有效地矫正,提取的表情特征信息更加真实。

利用主成分分析技术建立本征空间,在满足均方误差最小的情况下对表情信息进行重新描述,表情信息之间的相互关系能够得到更好的体现,特征数据维度也大大减少。常用的本征空间基的选择尺度是本征向量对应的贡献率,但是从表情识别的角度来讲,所选择的本征向量反映表情之间易于区分的、有代表性的特征信息才是关键。本文详细地分析了面部运动单元与本征向量的对应关系,对按照贡献率得到的本征向量进行再次筛选,使得各类基本表情之间的区分度更好,数据压缩效率也更高。

**关键词:** 表情识别; 特征提取; 归一化; 误差矫正; 主成分分析; 本征向量;  
运动单元;

## ABSTRACT

The design of the interface of human-computer interaction of new generation and affective intelligence are two international research frontiers. In these two frontiers, a key technology is how to obtain people's hidden feeling. The facial expression contains abundant information related to the change of the states of heart as the main behavior way of people's hidden feeling. Therefore the automatic recognition technology of facial expression has received the researchers' extensive concern.

Combining the psychological research results, in this thesis, we have set up one basic facial expression database which includes 840 video sequences at first, secondly, we confirm two interfaces used to structure the platform of facial expression recognition considering the dynamic characteristic, then making use of the well-known algorithms about facial regions location, feature extraction and classification, we build one instance of the system of facial expression recognition, finally, the instance has been tested using four kinds of basic facial expression in the completed database and the general recognition rate is 91.7%. The result proves the platform runs normally and can offer the good experimental environment for development and improvement of the follow-up expression recognition system, at the same time, the instance is close to demand of practical application, and lays a foundation for future application.

Expressional feature extraction is a more central question in the system, because of the complexity of the applicational environment, how to get the steady information which can reflect the facial expression essentially is a difficult problem urgently to be solved. We carry on deep analysis about the effective normalization and describing rationally of the expression feature, and has applied the analysis result to the instance structured.

Generally the geometrical normalization process is important to the extraction of facial expression feature. In commonly used normalization methods, the unstability of the fiducial features always causes normalization error. We analyze quantitatively the serious influence to feature extraction caused by normalization error, and propose the weighted optimal matching algorithm to solve it. The algorithm is founded on the basis of template matching theory, and assigns corresponding weight to the point by its motive degree when carrying on the calculation of correlative coefficient. Experimental results show that the algorithm can restrain the disturbance that normalization error produces, and the facial expression feature can be extracted more accurately.

After building the eigenspace using the principle component analysis and reprocessing

the facial feature with the least MSE, the interrelation of facial expressions can be shown better, and the data dimension also decreases. Generally the contribution ratio of eigenvectors is regarded as the rule of selection of eigenspace bases, but from the facial expression recognition's point of view, selecting the eigenvectors which can reflect the easily-differentiated and representative features is the key. We analyze the relationship between the action units (AUs) and eigenvectors and re-select the eigenspace bases. The results show that the basic facial expression recognition improves and the data compressibility increases.

**Key Words:** human facial expression recognition; feature extraction; normalization; error-correction; PCA; eigenvector; AU

## 插图索引

图 2.1 Hiroyuki KONISHI 定义的 7 种表情面部特征 <sup>[17]</sup> .....	10
图 2.2 人脸上半部分基本形变单元名称及对应的表情图像 <sup>[20]</sup> .....	12
图 2.3 表情的起因 <sup>[22]</sup> .....	13
图 2.4 人脸表情识别系统框架 <sup>[22]</sup> .....	14
图 3.1 基于混合特征提取的表情识别系统框架图 <sup>[27]</sup> .....	21
图 3.2 表情识别平台结构模型 .....	22
图 3.3 BFED 库表情序列样例帧 .....	26
图 3.4 人脸表情识别系统流程图 .....	31
图 3.5 ImgCharacter 数据结构 .....	32
图 4.1 几何归一化效果 <sup>[41]</sup> .....	37
图 4.2 存在归一化误差的表情特征区域光流计算结果 .....	38
图 4.3 存在表情形变时的图像对准效果 .....	40
图 4.4 被搜索图和模板 .....	42
图 4.5 基准特征偏移引起的归一化平移误差矫正效果 .....	43
图 4.6 加权优化匹配算法对表情信息的保留效果 .....	44
图 4.7 对同时存在误差和表情变化的图像进行矫正的效果 .....	44
图 5.1 本征向量还原图 .....	48
图 5.2 采用光流方法得到的混合 AU 特征图像 .....	50
图 5.3 基于贡献率选择的前 9 个本征向量 .....	50
图 5.4 基于本文方法得到的前 9 个本征向量 .....	51
图 5.5 改进前后识别率比较 .....	52
图 B1 与基本形变单元相关的面部肌肉图 .....	65

## 附表索引

表 2.1 Yacoob 和 Davis 采用的表情标记准则 <sup>[18]</sup> .....	11
表 3.1 BFED 库主要参数 .....	24
表 3.2 基本表情与面部器官变化的对应关系 .....	25
表 3.3 基本表情与 AU 对应关系 .....	33
表 3.4 表情识别实验参数统计 .....	34
表 3.5 表情识别实验识别结果 .....	34
表 4.1 无归一化误差时表情光流特征各分量最大值 .....	39
表 4.2 光流数据比较 .....	39
表 5.1 单个对象表情识别实验参数统计 .....	51
表 5.2 新选定的本征向量个数 .....	52
表 B1 单个形变单元 (Single Action Units) .....	62
表 B2 联合形变单元 (More Grossly Defined AUs) .....	63
表 B3 头部及眼睛方位变换描述单元 .....	64

# 湖南大学

## 学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：应伟

日期：2008年5月22日

## 学位论文授权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权湖南大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

1、保密□，在\_\_\_\_\_年解密后适用本授权书。

2、不保密□。

(请在以上相应方框内打“√”)

作者签名：应伟 日期：2008年5月22日

导师签名：廖卫平 日期：2008年5月16日

# 第 1 章 绪论

## 1.1 表情识别技术概述

### 1.1.1 研究背景

在法国的卢浮宫博物馆，有一幅有着 500 年历史的木板油彩画被誉为博物馆的三大珍品之一。这就是意大利文艺复兴时期的伟大画家达芬奇所创作的《蒙娜丽莎的微笑》。500 年过去了，蒙娜丽莎，这位佛罗伦斯的富豪夫人表露出的优雅的面容和神秘的微笑，依旧那么美，仿佛世界上一切的烦恼、可憎、可恨都可以随着她的微笑而融化。

是的，蒙娜丽莎在用她的微笑向世人传递着一种信息，世界应该是充满爱的。就此时来说，微笑的表情远远胜过了千言万语。在人们的生活中，表情，作为一种信息传递工具，无时无刻不在为人与人之间的沟通和交流起着重要的作用。心理学家 J. A. Russell 曾在他的文献中提到，在人们的交流中，只有 7% 的信息是通过语言来传递，而通过面部表情传递的信息量却达到了 55%。小女孩高兴的面部表情在向她的父母表示她很喜欢他们送的生日礼物，小男孩痛苦的表情在向体育老师表示他在足球比赛中被人踢到的地方很疼，祖母慈祥的表情则在向自己的子孙们表示：如果他们做的是善事，祖母心里会感到高兴；如果他们做错了，祖母也不会批评他们，会理解他们的行为。正因为有了面部表情这样一种信息传递途径，人与人之间的关系才更加紧密。

人脸表情识别技术随着人们对表情信息的日益重视而受到关注，成为目前一个研究的热点。如何捕捉人脸面部的运动信息、如何对脸部特征的形变特征进行分类、如何利用得到的表情信息推测人的心理状态等等都是表情识别技术需要解决的问题。作为一门多学科交叉的研究课题，人脸表情识别技术涉及到心理学、生理学、计算机科学、认知学等多个学科的知识。从目前的研究现状来看，表情识别技术进入一个研究的高潮与三个方面的推动因素有直接的关系：

1. 新一代人机交互界面的构建推动了人脸表情识别技术的发展；
2. 情感计算与情感智能的提出推动了人脸表情识别技术的发展；
3. 与人脸表情识别相关的技术的成熟推动了人脸表情识别技术的发展。

另外，计算机本身计算性能的提高也促进了人脸表情识别技术的进步。

### 新一代人机交互界面

关于人机交互界面的定义，可以这样说明：

人机交互界面是研究人、计算机以及它们相互影响的技术。更具体的说，人



机交互界面是指用户与计算机系统之间的通信，它是人与计算机之间各种符号和动作的双向信息交换<sup>[1]</sup>。这种信息交换的形式可以采用各种方式出现，如键盘上的击键、鼠标的移动、显示屏幕上的符号或图形等，也可以用声音、姿势或身体的动作等。

随着科技的发展，利用人的多种感觉通道和动作通道（如语音、手写、表情、姿势、视线等输入），以并行、非精确方式与计算机系统进行交互的多通道人机交互界面将取代目前的 WIMP 界面成为新一代人机交互界面的主流。

在 WIMP 界面中，尽管引入了图形技术，改变了原来命令语言的繁杂和单调，但是依然不能像人与人之间交流那样方便，主要的不足在于：

输入方式简陋，不能通过其它渠道譬如语音、手势等交流；

人与计算机地位不平等，计算机只能被动地接收信息，不能主动地捕捉信息；

多通道人机交互界面将彻底改变这一切，使得人与计算机之间的交互能像人与人之间的交互一样便利自然。其中，人脸表情识别技术将使得计算机能获取人脸面部信息从而推测人的心理状态，以便更人性化、更智能地为用户服务。所以说，新一代人机交互界面的构建要求更高效的人脸表情识别技术。

### 情感计算与情感智能

在日常生活、工作、交流、处理事务和决策中，人们的情感能力同正常的理性思维和逻辑推理能力一样扮演着重要的角色。神经生理学家 Damasio 研究发现，由于控制逻辑推理的大脑皮层和控制情感的边缘系统之间通道的缺损，病人尽管具有正常甚至超常的理性思维和逻辑推理能力，但严重阻碍了决策能力的运用。在决定一个人事业成功的因素中，情商即领会、运用和表达情感的能力可能比传统的智商更为重要。可以证明情感是智能的一个不可分割的部分。如果将计算机赋予类似于人类的情感能力，能够感知和表达情感，那么人与自然、人与机器将能够达成一种更加完美的和谐。

1997 年，情感计算（Affective computing）和情感智能（Affective intelligence）的概念被提了出来，并且作为一项前沿性研究受到广大研究者的关注<sup>[2]</sup>。情感计算可以定义为关于情感、情感产生以及影响情感方面的计算。在实际操作中，主要是通过各种传感器获取由人的情感所引起的表情及生理变化信号，利用合适的“情感模型”对这些信号进行识别，从而理解人的情感并做出适当的响应。情感本身并不能直接被感知，它就像思想一样，通过人的语言、姿态、行为以及其它创造性的方式如音乐、文学等来交流。

在情感的多种表现方式中，人脸表情作为最容易控制的一种而常常被用来进行情感的估计。所以，随着情感计算的深入，对人脸表情的自动识别技术的要求也会相应地得到提高。

## 相关技术的进步推动了人脸表情识别技术研究走向高潮

在过去的十几年中,人脸识别技术、人脸或局部器官定位技术、人脸跟踪技术都取得了巨大的进步。

国外研究机构和公司从事自动人脸识别的研究与实践较早,在美国、德国、日本等发达国家已经有相关产品问世。“9.11 事件”以后,美国更加大了人脸识别技术的研究力度,正在着手建立国家级的重点犯罪人口和恐怖破坏分子的人脸数据库。国内由中科院计算所—银晨网讯面像识别联合实验室自主研制的“面像检测与识别核心技术”在党的十六大会场——人民大会堂获得成功应用,鉴定专家经过详细测试认为,该系统在人脸检测、面部特征提取、人脸识别和人脸确认等技术方面均达到了国外先进水平。

在定位和跟踪方面,已有算法的鲁棒性和效率也大大增加,在充分利用多阶段由粗到精的搜索策略以及多信息有效融合的基础上,逐步与实际环境相结合,大大提高定位跟踪算法的实用性。

在人脸表情识别系统的开发过程中,人脸识别、定位、跟踪技术中有许多成熟的算法都可以应用到人脸表情识别系统中,从而大大推动了人脸表情识别技术研究走向高潮。

### 1.1.2 历史与发展

人脸表情识别技术是近几十年才逐渐发展起来的,现在成为了一个科研热点。国内外很多机构都在进行这方面的研究,尤其是美国和日本。进入 90 年代,对人脸表情识别的研究变得非常活跃,吸引了大量的研究人员和基金支持,EI 可检索到的相关文献就多达数千篇。美国、日本、英国、德国、荷兰、法国等经济发达国家和发展中国家如印度、新加坡都有专门的研究组进行这方面的研究。其中 MIT 的多媒体实验室的感知计算组、CMU、Maryland 大学的计算机视觉实验室、Stanford 大学、日本城蹊大学、东京大学、大阪大学、ATR 研究所的贡献尤为突出。最早的人脸表情研究见于文献[3], Suwa 和 Sugie 等人于 1978 年对表情识别做了一个最初的尝试,他们跟踪一段脸部视频动画,得到每帧图片上 20 个关键点的运动规律,将此运动规律与预先建立的不同表情的关键点运动模型相比较,这还只是初步的尝试。直到 1981 年才有人用仿生学方法从肌肉角度的观点为面部表情建立模型。1978 年 Terzopoulos 和 Waters 则运用了简化的 Ekman-Friesen 模型,用计算机产生人脸动画,同时也做了人脸视频序列的表情分析<sup>[4]</sup>。Mase 则使用光流来跟踪面部的运动单元,尽管只是在 8 个方向上跟踪光流特征的变化,但是所引入的表情识别模型掀起了动态表情识别研究的高潮<sup>[5]</sup>。

人脸表情识别技术的进步离不开表情库的建立,其一可以为表情识别算法的测试提供素材,另一方面也为各种算法的优劣比较提供了有效的依据。起初 Ekman

和 Friesen 系统地建立了一个有上千幅不同人脸表情的图像库<sup>[6]</sup>。近来美国卡耐基-梅隆大学、加州大学圣克鲁兹分校知觉实验室、耶鲁大学、普渡大学、马里兰大学也都各自建立了自己的人脸表情图像库。

在我国,人脸表情识别技术也逐渐开始受到重视。国内的清华大学、哈尔滨工业大学、中科院、中国科技大学、南京理工大学、北方交通大学等都有人员从事人脸表情识别的研究。2003 年 10 月,由中国科学院自动化研究所等单位牵头举办的第一届中国情感计算与智能交互学术会议将人脸表情识别与合成作为会议内容之一,吸引了国内广大研究者的极大关注<sup>[7]</sup>。2005 年 10 月,首届国际情感计算及智能交互学术会议同样在中国举行,来自世界优秀实验室的研究人员将与我国科研工作者进行密切的交流,会议还将邀请一些有突出贡献的研究者报告自己的工作内容,这将极大地推动我国在表情识别技术研究方面的进一步深入<sup>[8]</sup>。国家“863”计划、“973”项目、国家自然科学基金等都对人脸表情识别技术的研究提供了项目资助。

另外湖南大学在 2003 年 9 月承办了首届全国无处不在的计算与第二届全国可穿戴式计算学术会议<sup>[9]</sup>。会议邀请了董士海、李思昆、戴国忠、杨孝宗等著名学者做了关于人机交互、无处不在计算、可触摸的计算以及可穿戴计算方面的报告,并且提到了人脸表情识别技术在这些领域的应用前景。这次会议给湖南大学和国内其它相关研究机构在表情识别技术研究方面提供了一次良好的交流机会。

但是,和国外相比,国内的研究还只能说是刚刚起步,和美国、日本等发达国家差距还比较大,这需要我国研究者们的进一步努力。

## 1.2 表情识别技术的应用现状

### 1.2.1 应用领域

人脸表情识别技术已开始在人们的生活中得到应用,尽管只应用在某些特定的场所,功能也极为有限,但却受到了广泛的重视并带来了巨大的经济效益。目前主要的应用领域包括人机交互、安全、机器人制造、医疗、通信以及汽车领域。

#### 1. 人机交互领域

在人机交互当中,人脸表情识别技术主要应用在两个方面,其一是用户利用人脸表情来作为控制命令,指示计算机应该如何操作。通过事先进行训练,在当计算机实时监控人脸表情变化并且发现作为控制命令出现的表情信息时,计算机就会按照既定的程序作出反映,譬如用户眨一下眼睛,表示要求计算机取消当前的任务,眨两下眼睛表示要求计算机将任务保存;其二是计算机能主动地从用户的面部表情推测用户的内心情感,并自动地给出相应的人性化的服务。已有的应用例子是智能音乐播放器,随着人的心情的变化,播放器能有选择性的播放适合用例子是智能音乐播放器,随着人的心情的变化,播放器能有选择性的播放适合

用户心情的歌曲。

## 2. 安全领域

指纹、虹膜以及人脸等生物信息可以被用在身份验证等安全管理中，同样，人脸表情信息作为一种重要的生物信息也可以被用来表征身份。尽管每个人都会做出同一种类型的表情，但是表情产生的部位、表情的强度、面部器官形变的轨迹、表情持续的时间都会有较大的差异，并且这种差异相对于指纹、虹膜以及人脸来说更加不容易被伪造。

## 3. 机器人制造领域

在美国科罗拉多州首府丹佛举行的美国科学促进会的年会上，展出了一种有着女性脸型的新机器人，这种机器人能按照指令完整地模仿并表达人类的 28 种面部表情，而且它的面部还会随着年龄的变化出现皱纹。这个机器人名叫 K-bot<sup>[10]</sup>。在机器人的眼睛里装有两部照相机，它们能对面前人的表情进行观察并识别，随后就可以模仿出悲伤、高兴或惊讶等不同的面部动作，如它会微笑、嘲笑、皱眉甚至是扬眉。此外，它的嘴唇、脸颊和鼻子也都能移动。人脸表情识别技术的应用使得机器人不再是一个简单的、按照给定程序工作的机器，而是一个开始学习去理解人的意图，以便更好地为人类服务的“智能人”，这将逐渐改变机器人在社会中的地位。

## 4. 医疗领域

Y. Dai 等人设计了一个专门用来照看病人的电子护士，电子护士的主要设备是一台安装在病人床头的带有表情识别功能的摄像机<sup>[11]</sup>。在很多情况下，病人双手的活动往往受到一定的限制，无法便利地进行按钮呼叫，同时，对于病情严重并且需要特别护理的病人来说，任何时候守护人员的疏忽都可能导致因为对病人处理的不及时而引起的危险。电子护士通过自身带有的表情识别功能能很好地解决上述两个方面的问题，对于需要服务的病人来说，并不需要用费力去按按钮，只要做出痛苦的表情，电子护士马上就会发出呼叫信号，提醒医护人员提供必要的服务。对于需要特别护理的病人来说，病人身体状况的变化经常会反映在面部表情上，电子护士将全天候地跟踪病人的面部表情行为，并随时向医护人员通报病人的状态，以便能对紧急情况作出及时处理。

## 5. 通信领域

随着人们生活水平的提高，简单的语音通信已经不能满足人们的需要，可视电话、视频会议等应用技术的开发逐渐提上日程，但是对于目前并不富裕的带宽来说，尽管只需要显示用户的头肩部分图像，但是所传输的数据量也相当庞大。众所周知，在人与人之间交谈时，面部本身器官的结构特征是次要的，语言和面部表情才是传递信息的主要载体，设想如果在通话的双方都建立一个虚拟人头部模型，仅仅通过网络传递对方的语言和面部表情，并且利用面部表情信号驱动虚

拟人头部模型进行相应的表情变化，那么就可以实现在大大节省带宽的情况下，尽可能地传递对方要表达的信息。

## 6. 汽车领域

在对 2001 年到 2004 年京津塘高速公路北京段的交通事故进行了统计分析的过程中，数据表明造成交通事故的原因主要有疲劳驾驶、行人进入高速公路、司机措施不当、车距过近、违章并线、天气、爆胎、超速等 10 多种，其中疲劳驾驶位列第一，占有所有交通事故起因的 27%。针对此类情况，Haisong Gu 和 Qiang Ji 等人设计了一个安装在汽车上用来检测司机疲劳状态并提醒司机需要注意安全的设备<sup>[12]</sup>。这种设备会实时分析司机的精神状态并予以判断，当预测司机疲劳程度会影响安全驾驶或者发现司机有打盹现象时，设备会采取必要的措施譬如报警或者播放刺激音乐来防止可能会导致的交通事故。

### 1.2.2 面临的挑战

经过几十年的发展，表情识别技术逐渐走向成熟并且在实际生活中得到应用。可以预见在未来的十年中，表情识别技术将会使我们的生活环境变得更加和谐。目前的表情识别技术的应用还只是一个探索性的行为，距离其作为一种成熟的产品投入到人们的生活中还有很长的一段路。这其中面临的主要挑战包括如下几个方面：

#### 1. 鲁棒性有待提高：

因为表情识别系统需要实时定位跟踪人脸并提取面部表情形变信息，所以系统容易受到外界因素的干扰，其中主要的干扰是头部偏转运动以及光线的变化，有研究者提出采用多摄像头技术、旋转不变特征技术、色彩补偿技术来予以解决，尽管有一定的效果，但是当头部运动比较复杂、多种光线交叉变化时，算法识别率并不理想，所以表情识别系统的鲁棒性还有待进一步提高。

#### 2. 表情识别种类有待增加：

能识别的表情种类的多少直接反映了应用系统功能的完备程度，现在大多数表情识别应用系统都是针对人的某一类表情来进行分析，功能比较单一，系统提供的服务比较有限。

#### 3. 表情识别计算量有待降低：

在保证识别率的前提下尽可能降低系统的计算量，将能确保系统实时性的要求，同时，在表情识别系统的功耗等方面也会得到很好的改善。

#### 4. 加强多信息技术的融合：

表情识别应用的一个重要方向就是推测人的内心情感从而让计算机能提供一种人性化的服务。但是，情感并不仅仅只有面部表情这样唯一的一种表现方式，并且，在很多情况下，面部表情并没有反映人的真实的心理状态。综合多方面的

信息, 譬如语音语调、脉搏、体温等, 来更准确地推测人的内心情感, 也将是表情识别技术需要考虑的问题。

此外, 与心理学知识的有效结合、与智能处理技术的有效结合也将是表情识别技术进一步发展所面临的挑战。

## 1.3 本文所做的工作

### 1.3.1 本文的主要研究内容和特点

心理学研究表明, 在人类所有的表情类型当中, 存在六种基本表情类型, 任何一种非基本人脸表情都可以由这六种基本表情中的某几种按照一定的比例融合得到<sup>[13]</sup>。同时, 因为表情变化是一个动态的过程, 表情发生的部位、表情的强度、表情的时间性差异都直接反映了人的心态的不同, 所以, 本文首先在了解目前国内外主要的表情识别技术基础之上, 拟构建一个针对动态表情序列、识别类型为六种基本表情类型的表情识别系统, 并在此系统上对表情特征的有效提取和合理描述做相关的研究和探讨, 以便在表情识别系统的鲁棒性和效率方面做一定的改善。

其中, 主要的研究内容可归纳如下:

1. 对心理学和计算机方面有关人脸表情识别的研究内容做了详细的归纳和总结, 并对与表情识别有关的技术做了比较, 指出其在不同的应用环境下的优劣性。

2. 制定了一套用于基本表情识别实验的表情数据库采集标准, 并按照标准对 8 个实验对象的包含中性表情的七种基本表情分别采集了 15 段视频序列, 最后得到一个 840 段序列的基本人脸表情数据库。

3. 针对表情识别的动态性提出了两个接口, 并基于接口构建了一个可实际运行的表情识别系统。首先利用目前在统计特征检测方面比较有效的 Adaboost 方法进行人脸检测, 得到基于连续帧的人脸面部图像序列, 然后采用光流特征提取技术得到表情变化特征, 并采用主成分分析将特征数据压缩成特征流, 最后利用混合隐马尔可夫模型进行分类。对事先采集的人脸表情图像序列进行识别, 得到 91.7% 的综合识别率, 验证了系统的稳定性, 同时, 通过构建系统, 对人脸表情识别的各个关键技术也有了更深入的认识。

4. 结合光流特征提取方法, 定量地分析了归一化误差对表情特征提取的影响, 并提出了加权优化匹配算法进行误差的矫正。算法以传统模板匹配原理为基础, 根据各像素点的运动剧烈程度分配相应的匹配权值, 实验证明误差得到有效的矫正, 提取的表情特征信息更加真实。

5. 详细分析了人脸表情本征空间构建的原理, 利用基于贡献率的主成分选择方法选择维数较少的本征向量来表示原始表情信息, 这样的处理能使表情信息之

间的相互关系得到更好的体现，新的特征数据维度也大大减少。同时，针对基于贡献率的主成分选择方法没有考虑人脸表情信息的变化特点，只是运用基本的数学知识对数据进行统计分析的不足，提出了新的表情本征向量选择方法，实验证明，利用面部运动单元与本征向量的对应关系所选择的本征向量能更好地反映各类表情之间易于区分的、有代表性的特征信息，从而使得表情识别率也大大增加。

### 1.3.2 本文的组织结构

全文分为六个章节，内容如下：

第一章为绪论，介绍了人脸表情识别技术研究的背景、历史与发展情况，以及目前的应用现状和面临的挑战，最后对本文所做的主要工作做了一下归纳。第二章对心理学和计算机科学两个领域中人脸表情识别技术方面的研究情况做了一个详细的总结和分析，并对目前人脸表情识别系统中应用到的主要技术做了比较和分类，指出在技术层面碰到的难题。第三、四、五章是本文的研究重点。第三章运用目前比较成熟的技术，构建了一个人脸表情识别系统，并且在本文构建的基本人脸表情库上进行实验。第四章在表情特征提取的稳定性方面做了一些工作，针对图像归一化误差问题，本文提出加权优化匹配算法予以矫正，并对矫正前后的表情特征进行了比较。第五章在表情特征的描述方面做了一些改进，结合人脸表情本身的变化特性来选择新的本征向量，同样对改进前后的算法做了实验对比。最后一章对全文进行总结，并对下一步的工作做了展望。

## 第2章 动态人脸表情识别技术研究现状

### 2.1 引言

人脸表情识别技术发展到现在,最大的突破是实现了“动态”。数据源不再是事先拍摄好的表情达到最高潮的面部图片,而是分段的表情视频序列。对于人脸面部图像的定位跟踪技术也要求是动态,而且由此带来了一系列技术的进步,如表情特征信息压缩技术的改善,以及对特征流信息进行合理分类的高效率的分类器的构建等等。这种动态性的思想更合乎实际情况,人脸表情的信息千变万化,非常复杂,用简单的静态图片根本无法有效地反映出人的内心情感。尽管目前的“动态”性研究还处于实验室阶段,识别的表情类别也非常有限,但是已经是朝着实际应用的方向迈进了一大步。

其次的突破是表情分析的细化。细化主要表现在三个方面,其一是进行局部表情的分析,人的表情在不同的面部特征区域反映的表情变化信息是不同的,从局部分析入手,利用合理融合的机制对整体表情作出判断更贴近实际情况;其二是对基本形变单元的变化特征进行分析,Ekman 和 Friesen 将人脸面部各块肌肉的运动定义为基本形变单元,直接深入到对各个基本形变特征进行分析,不仅能实现对整体表情的有效把握,还能进一步理解人的真实内心情感;其三是表情量度的分析,表情量度最直观的体现是面部器官形变的程度,根据各特征部件的形变程度并结合一定的心理学先验知识可以对表情量度进行估计。

动态人脸表情识别技术的进步离不开心理学研究的进步,两者相辅相成:一方面通过对人类心理感知的研究,有针对性地对计算机表情识别系统进行改进,从而提高机器的识别能力;另一方面又可以通过机器的识别能力进一步深入探索人类的心理感知过程。并且值得一提的是,人脸表情识别技术最先是从事理学研究开始的。

本章将首先介绍心理学方面对于人脸表情识别系统构建起了重要作用的成果,其次对动态人脸表情识别系统整体构建以及相关技术进行介绍,然后对表情识别的难点进行归纳,最后是小结。

### 2.2 心理学研究成果

心理学的研究成果可以分为两大类:

其一是建立合适的表情模型。达尔文最早对人类表情的相似性以及表情的延续性进行了研究,指出不管是何种民族、何种性别以及何种年龄的人,对于同种



类型的情感，总是表现出相似的表情变化特征，并且在人类的繁衍过程中，表情行为能被继承下来。达尔文还指出，在所有表情类别中，有几种表情是比较基础的，这几种表情起源于人的生活习性并与人的生活紧密相关<sup>[14]</sup>。1971年，心理学家 Ekman 与 Friesen 的研究表明，人类有六种主要的情感，每种情感以唯一的表情来反映人的一种独特的心理活动。这六种情感被称之为基本情感，在各种种族的文化中是相通的，由愤怒、高兴、伤心、惊讶、厌恶和恐惧组成<sup>[13]</sup>。尽管已经有相关的研究，但是因为表情的复杂性，人类表情的定义和分类目前并没有统一的定论，还有待进一步探索。

其二是对人脸表情进行描述。人脸面部发生表情形变时，各个器官的形状、各个点的色彩都会因为表情形变而发生变化，如何从各种各样的特征信息中提取出与表情变化相关的特征信息以及如何将特征信息与人脸表情对应起来是众多心理学家一直在研究的课题。Ekman 与 Rosenberg 提出了 FACS AID 系统<sup>[15]</sup>，将每种表情与肌肉的运动对应起来，进行表情识别时，只需要观察判断得到肌肉的运动，就可以指出表情的类别；也有心理学家简单地将面部主要器官的角点的运动轨迹与表情类型对应起来；Irfan A. Essa 从计算机视觉技术的角度出发，运用优化的光流技术，同时结合一种三维人脸模型，直接将人脸表情全局映射为一种光流模式来对表情予以描述区分<sup>[16]</sup>；Hiroyuki KONISHI 等人根据面部器官的位置和曲率定义了 7 种表情特征来描述表情的变化，7 种特征的定义见图 2.1。

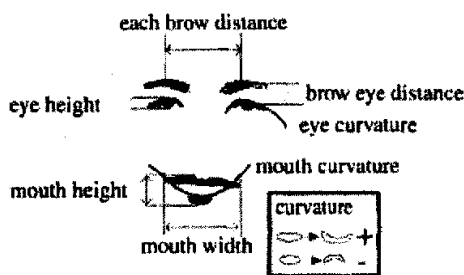


图 2.1 Hiroyuki KONISHI 定义的 7 种表情面部特征<sup>[17]</sup>

## 2.2.1 理解人脸表情

人脸表情的产生是由面部肌肉的收缩引起的，主要通过面部器官譬如眼睑、眉毛、鼻子、嘴唇以及皮肤的形变来表现，也可以通过皱纹来表现。肌肉收缩的持续时间并不长，一般在 250 毫秒和 5 秒之间。

在描述人脸表情的时候，常用的标准有三个：位置，强度以及动态性。

人脸表情的强度或者通过表情器官的几何形变来进行度量，或者通过一定区域内的皱纹密度来度量，以笑容为例，面颊上抬和嘴角上扬的幅度可以反映出笑容的强度。因为人与人之间表情变化的程度有差异，所以如果不参考待测对象的

中性表情的话,很难有办法能准确地对表情量度进行测量,尤其对于自然表情来说,可变因素太多,对表情测量的难度更大。

不仅仅只有面部器官形变信息可以反映人的心理变化,形变的时间性以及形变的演变过程也同样可以反映人的心理变化,这被称为表情的动态性。动态性可以借助于三个时间标记来进行描述,这三个标记分别为开始,高潮和结束。三个标记之间的时间间隔的不同,以及三个标记本身的转变顺序的不同都体现了表情动态性的差异。譬如笑容,如果检测到开始的标记和达到高潮的标记之间的时间间隔很短,则表示待测对象目前的心理状态可能是一种惊喜,反之,如果时间间隔很长的话,则可能是一种普通的喜悦,这两种结果是有很大的区别的。在静态图片中无法获取表情的动态性信息,因而也不能对待测对象的心理进行精细地推测。尽管动态性已经被承认有很重要的地位,但是在目前的人脸表情识别系统中对于时间标记的准确定位方法并不多,在医学上采用肌电图方法可以很准确地定位时间标记,除了这种方法之外,一般就只是通过检测事先规定的面部形变行为来进行判断了。表 2.1 是 Yacoob 和 Davis 采用的表情标记准则。

表 2.1 Yacoob 和 Davis 采用的表情标记准则<sup>[18]</sup>

表情	开始/结束	标志性准则
愤怒	开始	向内下拉眉毛,嘴巴紧缩
愤怒	结束	向外抬起眉毛,嘴巴沿展
厌恶	开始	鼻子向上抬,嘴巴沿展/张开,眉毛下降
厌恶	结束	鼻子下降,眉毛上抬
高兴	开始	嘴角上抬 或 嘴巴沿展/张开
高兴	结束	嘴角下降 或 嘴巴紧缩/关闭
吃惊	开始	眉毛上抬,下唇(颚)下降
吃惊	结束	眉毛下降,下唇(颚)上抬
伤心	开始	嘴角下降,嘴巴上抬,眉毛内端上抬
伤心	结束	嘴角上抬,嘴巴下降,眉毛内端下降
恐惧	开始	轻微嘴巴沿展和下降,眉毛内端上抬
恐惧	结束	轻微嘴巴紧缩和上抬,眉毛内端下降

### 2.2.2 FACS 系统

情绪的种类千差万别,面部特征点变化的信息也非常繁杂,但是面部表情是由面部肌肉的运动产生的,与表情变化相关的面部肌肉只有 18 块,从解剖学的知识出发,Ekman 和 Friesen 提出的一种人脸面部表情运动的描述方法—人脸运动编码系统 FACS (Facial Action Coding System),同时,根据面部肌肉的类型和

运动特征定义了基本形变单元 AU (Action Unit)<sup>[19]</sup>。基于这套理论,人脸面部的各种表情最终能分解对应到各个 AU 上来,分析表情特征信息,也就是分析面部 AU 的变化情况。

人脸运动编码系统总共定义了 44 个基本的形变单元以及一些用来描述头部和眼睛的方位变换的单元,譬如头部左右偏转等等。44 个基本形变单元又分为单个形变单元以及联合形变单元两个部分。对每个单个形变单元,系统中都提供了单元号码、单元名称以及与之相关的肌肉名称。具体形变单元的定义见附录 B.1,人脸面部肌肉的说明见附录 B.2。其中,人脸上半部分基本形变单元名称及对应的表情图像样例见图 2.2。

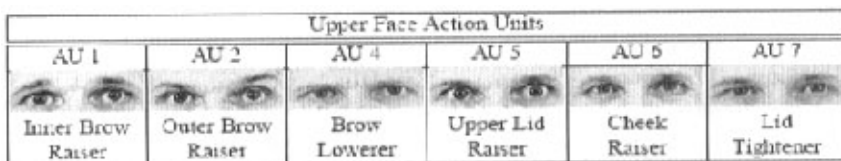


图 2.2 人脸上半部分基本形变单元名称及对应的表情图像<sup>[20]</sup>

随着研究的深入,研究者们对于 FACS 在人脸表情识别技术中的应用提出了一些新的看法。

高文和金辉认为 FACS 有两个主要弱点<sup>[21]</sup>:

1. 运动单元是纯粹的局部化的空间模板;
2. 没有时间描述信息,只是一个启发式信息。其中 FACS 不包含情绪信息,数据的分析只是单纯的描绘运动单元 AU,或者把 FACS 通过字典规则转化成情绪。

Essa 也认为 FACS 在对表情的描述方面过于简单化,人脸的表情应该是所有肌肉同时协作完成的一种行为,只是各个肌肉的影响因子存在差异而已,Essa 提出了基于视频的动态表情描述方式 FACS+, FACS+ 直接运用真实人脸样例产生表情特征模板进行表情描述,避免了 FACS 文字性的描述带来的表情失真;而且 FACS+ 充分考虑了表情的空间变化特征和时间变化特征,克服了 FACS 对动态表情特性描述的不足<sup>[16]</sup>。

### 2.2.3 表情与情感

因为目前表情识别的目的主要是用于理解人的几种基本情感,所以,很多表情识别系统研究者往往会将表情与情感的概念、表情与人的心理状态的概念混淆,这样的理解存在局限,会使表情识别技术的研究进入一种进退两难的地步。对此,心理学家进行了详细地区分。

表情的产生主要由四个方面的因素引起,分别是精神活动、说话、示意以及生理活动。情感只是属于精神活动里面的一个部分。它们之间的关系见图 2.3。

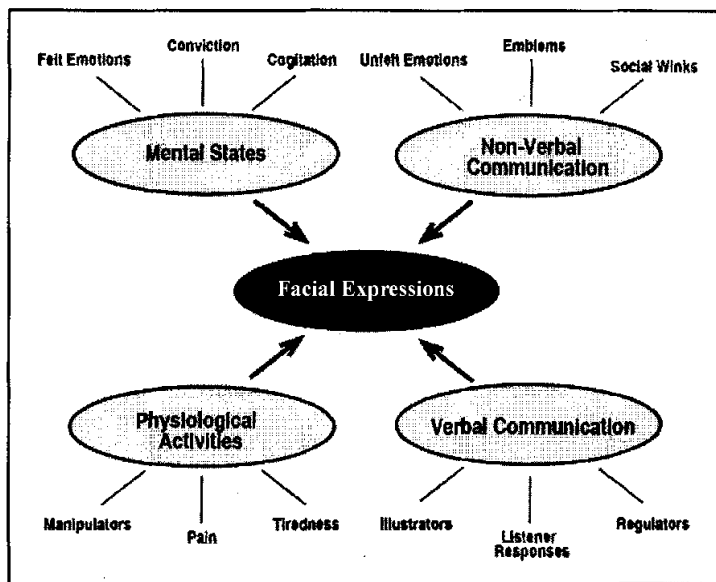


图 2.3 表情的起因<sup>[22]</sup>

同时，因为生活习惯不同，或者环境受到限制，当人的情感发生变化时，也有可能不会出现表情变化，或者表情变化反映的不是真实的情感。心理学研究表明，反映情感的方式还可能是：音调，姿态，手势，凝视的方向，甚至脉搏，体温等等。

所以，对表情和情感的关系应该描述如下：

情感是表情发生的起因之一，而表情也只是反映情感的方式之一。

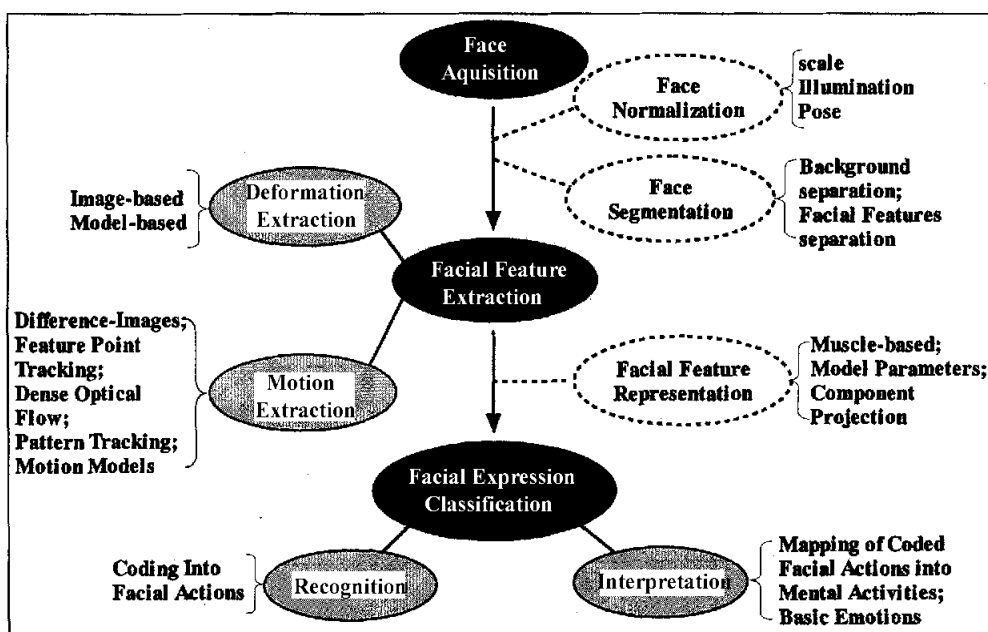
将表情与情感进行明确区分，有利于利用多个方面的信息来理解人的真实情感，譬如目前提出的多功能状态机等，就是将多个方面的特征信息进行适当的融合来分析人的心理状态的变化。同样，将表情与情感区分开来，能够对表情的起因有更深入的认识。

## 2.3 计算机领域的研究

### 2.3.1 人脸表情自动识别系统框架

人脸表情自动识别系统主要分为三个部分，分别是人脸/器官定位部分、表情特征提取部分以及表情分类部分。系统框架图见图 2.4。

其中，在表情特征提取之前，通常会进行面部图像的归一化处理 and 人脸图像与背景、面部器官图像与人脸图像的分离，这样可以便于表情特征更为稳定、有效地提取；在将表情特征信息送入分类器之前，也往往需要将信息进行一定的组合。这几个步骤在框架图中是用虚线椭圆框表示的，在表情识别系统中并不一定都会用到，可以按照表情特征提取方式的不同以及分类器的不同进行合理选择。

图 2.4 人脸表情识别系统框架<sup>[22]</sup>

表情分类部分可以有两种分类方式，其一是识别，将提取的表情信息转化成人脸运动编码系统中的形变单元，也即通常所说的 AU 识别。将表情信息分解为 AU 单元的组，可以进行情感的理解，也可以在人脸表情仿真等方面应用；其二是解释，直接将表情信息分类为情感类别，可以借助于 FACS AID 系统进行情感解释，也可以按照情感类别的不同进行直接映射。

框架图中大括号所包含的内容是各个部分常用到的技术。

### 2.3.2 表情区域定位方法比较

按照表情特征提取区域的不同，定位的对象有两类，一类为整个人脸区域的定位，另外一类为面部器官区域的定位，下面分别予以介绍。

#### 人脸检测

人脸检测问题最初作为自动人脸识别系统的定位环节被提出，近年来由于其 在人脸表情识别、安全访问控制、视觉监测、基于内容的检索和新一代人机界面 等领域的应用价值，开始作为一个独立的课题受到研究者的普遍重视。

人脸检测(face detection)是指在输入图像中确定所有人脸(如果存在)的位 置、大小、位姿。人脸的检测过程实际上就是对人脸模式特征的综合判断过程， 包括多种方法：

##### 1. 肤色区域分割与人脸验证方法

对于彩色图像，在确定肤色模型之后，首先可以进行肤色检测。在检测出肤

色像素后, 需要根据它们在色度上的相似性和空间上的相关性分割出可能的人脸区域, 同时利用区域的几何特征或灰度特征进行是否是人脸的验证, 以排除其它色彩类似肤色的物体。

## 2. 基于启发式模型的方法

基于启发式模型的方法首先抽取几何形状、灰度、纹理等特征, 然后检验它们是否符合人脸的先验知识。

## 3. 基于统计模型的方法

基于统计模型的方法主要有以下几类: 基于特征空间的方法、基于人工神经网络的方法、基于概率模型的方法以及基于支持向量机的方法。基于统计模型的方法被认为是解决复杂的人脸检测问题的最为有效的方法。它具有如下优点:

不依赖于人脸的先验知识和参数模型, 可以避免不精确或不完整的知识造成的错误;

采用了实例学习的方法获取模型的参数, 统计意义上更为可靠;

通过增加学习的实例可以扩充检测模式的范围、提高检测系统的鲁棒性;

基于统计模型的方法大多适用于复杂背景图像中的人脸检测;

对于人脸检测研究的趋势, 应该是利用多种线索(头发、肤色、器官、轮廓、模板等), 综合多种分类方法(混合高斯模型、概率模型、神经网络与支持向量机等), 将启发式信息与统计学习方法相结合来进行人脸的检测。

## 面部器官定位

对于面部器官的定位, 根据定位所依据的基本信息的类型, 可以将现有的面部器官定位的方法分为 5 大类, 分别是基于先验规则定位、基于几何形状信息定位、基于色彩信息定位、基于外观信息定位和基于关联信息定位<sup>[23]</sup>。

基于先验规则的特征区域定位方法根据脸部特征的一般特点总结出一些经验规则, 搜索前, 先对输入图象作变换, 使目标特征得到强化, 而后根据上述规则从图中筛选出候选点或区域。基于几何形状信息是根据脸部特征的形状特点构造一个带可变参数的几何模型, 并设定一个评价函数量度被检测区域与模型的匹配度, 搜索时不断调整参数使评价函数最小化, 从而使模型收敛于待定位的脸部特征。在基于色彩信息的方法中, 首先使用统计方法建立起脸部特征的色彩模型, 搜索时遍历候选区域, 根据被测点的色彩与模型的匹配度筛选出候选点。基于外观信息是将脸部特征附近一定区域(窗口)内的子图象作为一个整体, 映射为高维空间中的一个点, 这样, 同类脸部特征就可以用高维空间中的点集来描述, 并可以使用统计方法得到其分布模型, 在搜索中, 通过计算待测区域与模型的匹配度即可判定其是否包含目标脸部特征。基于关联信息是在局部信息的基础上, 引入脸部特征之间的相对位置信息, 以缩小候选点范围。

面部器官相对于人脸来说，特征信息模糊度更大，在面部器官检测时，也可以首先检测到人脸，再在此基础上对器官进行定位，如果单独处理的话，可以考虑将多种方法结合起来进行，这样准确率会大大提高。

### 2.3.3 表情特征提取方法分类

表情特征提取方法按照特征提取的部位、特征产生的方式、是否利用三维模型等可以进行如下几种分类：

#### 1. 全局方法和局部方法

全局方法：将整个面部当成一个整体对待，将面部整体的变化模式与表情类型关联起来；

局部方法：集中分析由表情变化引起的面部特征或者区域变化。

#### 2. 变形方法和行为方法

行为方法：直接依赖由表情产生的面部变化；

变形方法：依赖于中性表情图像或者人脸模型来提取面部表情特征，可以应用到基于静态图片和基于序列的表情识别，在以视频序列为数据源时，也只是单独地处理每一帧。

两者相比行为方法能提取比较底层的特征流信息，对细微表情特征的提取能力更强一些。

#### 3. 图片方法和模型方法

图片方法：直接从图片中获取信息，无需对目标有过多地了解，操作起来比较简单，计算量也比较小，不足之处在于鲁棒性比较差，当环境变得复杂时，特征提取效果会变差；

模型方法：通过二维或者三维人脸模型来描述面部结构，前者基于面部表观行为，后者试图恢复立体几何结构。三维模型有两种，肌肉模型和行为模型。三维模型需要繁杂的、计算量大的适应过程，而且，构造精确的头部和面部模型必须手工进行，任务量繁重。

#### 4. 表观方法和肌肉模型方法

表观方法：从图象上测得的变化并不完全反映实际的运动情况，称之为表观运动，利用表观运动图像和二维模型方法集中分析面部肌肉运动产生的视觉效果；

肌肉模型方法：通过获取的视觉信息（譬如光流），来改变肌肉模型中肌肉的运动，肌肉模型所反映出来的行为可以用来推测表情的变化。

### 2.3.4 表情识别方法分类

人脸表情自动识别系统的最后一个处理步骤是表情特征的识别，即分类。分类有两种方式，一种是基于时空的方式，另外一种是基于空间的方式。

基于时空的方式：

隐马尔可夫模型在语音识别系统中得到了广泛的应用，因为其对动态性序列有很好的描述能力，既可描述瞬态的随机过程又可描述动态的随机过程的转移特性，所以在表情识别中也常被采用。尤其在应用基于行为的表情特征提取方法的系统中，隐马尔可夫模型常被用来对特征流信息进行分类<sup>[24]</sup>。递归型神经网络 RNN (Recurrent neural networks) 在人脸表情识别系统中也应用比较多，相对隐马尔可夫模型来说描述能力更强<sup>[25]</sup>。在 RNN 中，参考点根据不同的情况自动地改变，在构建 RNN 进行动态表情序列识别的时候，有三个功能需要考虑：

能感知人脸表情特征信息中细微的运动；

能感知面部器官的整体运动矢量；

在进行当前表情状态识别的时候，需要结合上一步得到的表情状态信息。

另外一种对表情在时间上的演变性具有描述能力的模型被称为时空行为能量模板 (spatio-temporal motion-energy templates)，人脸表情行为被映射为二维行为场，根据行为模板之间的欧几里德距离来对表情类型进行区分<sup>[16]</sup>。

基于空间的方法：

神经网络方法被用在人脸表情识别方法中，模型由简单的、并行处理的互连处理单元构成，每个神经元与其它神经元相连并从与其相连的神经元接受输入，输入加权后输出送给其它的神经元，各神经元之间的连接强度通过连接权值来描述。在表情识别中，输入数据或者是直接的人脸图像信息，或者是经过一些特征转化方法处理后的信息，常用的特征转化方法包括主成分分析方法 (PCA)、独立成分分析方法 (ICA) 或 Gabor 小波方法。

但是神经网络方法的训练过程的工作量非常大。以往主要是对基本表情类型进行分类，类别比较少，训练的工作量相对较少，但是如果是自然表情类型，那么需要解决的第一个问题就是：在 FACS 框架中，用来描述自然表情类型的 AU 组合达到了 7000 多种，对各种类别的 AU 组合的训练问题需要合理的解决。目前已有关于如何减少神经网络方法计算量的解决方案被提了出来。

## 2.4 动态人脸表情识别技术难点分析

用计算机来自动分析、识别面部表情是一个非常复杂的问题，它关键在于建立一个合理的人类情绪 (human emotions) 模型并将它与人脸面部特征及表情的变化特征联系起来。但是人脸是个柔性体，不是刚体，很难用模型来精确地描绘。而且，表情的识别还依赖于以下几个方面的因素：

### 1. 对人脸的熟悉程度

人脸作为表情的载体，对表情的识别有至关重要的作用。在不同的人之间以及同一个人的不同年龄阶段之间，面部呈现的特征并不一致，这与人的生理结构



有关,这种差异势必会影响表情的分析。在表情识别系统中,有两个重要部分依赖于人脸结构,其一是提取的特征的幅度,其二是建立的人脸三维模型。对人脸进行充分地认识,对可能存在的变化予以适当地处理,能使对表情的理解变得更加深入和精细。

## 2. 对各种表情的体验

表情的表现有缓和的和激动的、细微的和强烈的、轻松的和紧张的等诸多形式,它的生理因素也是细微多变的,所以非常复杂。从统计的角度来看,表情的多样性主要表现在表情强度的差异以及表情类型的差异两个方面,同样的高兴表情,不同的人有不同的表现,有些只是嘴角稍微向两侧上抬;有些不仅仅上抬嘴角,而且眼睛部位也出现上下眼睑的相对运动;有些在高兴时并不习惯上抬嘴角,而是整个嘴先轻微张开,然后向左右两边运动等等之类。对表情的体验还需要结合心理学方面的知识,所以,对表情的充分认识一直是表情识别的难点所在。

## 3. 对脸部的注意程度

在实际应用中,待测目标的运动是无法受到限制的,从非正面的人脸面部得到的表情信息是不完整的,这种不完整性会导致识别结果的可信度下降,甚至导致识别错误。实时获得人脸面部变化的完整信息,或者采用合理的补偿方法进行补偿也是表情识别系统中需要处理的难题。

## 4. 非视觉的因素也给表情识别带来一定的困难

人脸面部的头发遮挡、眼睛等附属器件的干扰以及面部化妆带来的对表情识别的影响,都是表情分析中必须面对的问题。这种干扰的随机性比较大,影响程度也无法事先确定,目前的解决方法是在训练的时候尽可能地将带有以上干扰的样本纳入训练集,或者直接给予限制。

## 5. 数据来源方面

目前人脸库的建立已经比较完善,但是对于人脸表情来说,得到大家广泛认可的、并且兼顾了表情动态性的自然表情库并不存在。已有的较大规模的表情库通常也只是对几种基本表情数据进行了采集,并且还带有较为突出的人为性质。采集自然表情的难度有三个方面,其一是自然表情的种类繁多;其二心理学方面对于表情的定义目前还不统一,无法提供有力的理论依据;其三是对采集手段要求比较高,既要进行采集,又不能对待测对象有事先的准备。

## 6. 光照等外界环境因素对图像的影响大

光照等外界环境的变化会对人脸检测带来一定的干扰,尤其是对基于肤色的面部区域检测方法来说干扰更为严重。在特征的提取过程中,外界环境的变化会使提取的特征中存在较多的干扰杂质。在分类阶段,训练时往往无法对光照情况进行事先的预测,也就无法对可能出现的外界环境变化进行覆盖性训练,外界环境的变化会直接导致识别出现错误。

## 2.5 小结

新一代人机交互界面的构建、情感智能以及情感计算概念的提出使得对人脸表情的研究日益受到重视，作为一种重要的生物特征信息，人脸表情反映了用户的心理特征，特别是情感方面的特征，这种信息的准确获取能使计算机更加人性化、更加智能地为用户服务。在过去的几十年中，无论是心理学方面的研究，还是计算机方面的研究都取得了重要的成就，心理学方面为表情识别提供了坚实的理论基础，相对于以往来说能更深入地对表情变化的本质作出说明，并指出了表情变化与人的心理状态之间的联系和区别，这对计算机工作者的研究起到了指导作用。表情并不等同于情感，情感也并不是只有表情这样唯一的一种表现方式，单纯对人脸表情进行分析是不可能充分地理解人的情感的，必须借助于多方面的特征，譬如上下文关系、语气等等来对情感进行综合估计。计算机方面的工作主要是对人的六种基本表情进行识别，从过去基于静态图像发展到了目前基于视频序列来进行表情的分析，对表情的认识更加深入，识别效果也得到了改善。

但是，目前的表情识别系统还仅仅限于对背景比较单一、表情类型比较简单的情况作出处理，对自然表情分析、表情程度的估计、多信息有效融合等方面的研究还存在严重不足，本章 2.4 小节分析了表情识别的难点所在，对这些难点提出适当的解决方法将有利于表情识别技术的进一步发展。

## 第3章 表情识别平台设计及系统实例构建

### 3.1 引言

从前面几章的介绍可以看到,随着人脸表情识别系统的应用要求越来越高,表情识别系统的功能也越来越完善。从发展趋势来看,系统的改进途径有两条主线,其一是将系统的各个部分的研究独立出来,分别利用相关领域的技术来予以完善,譬如人脸定位跟踪可以单独作为一个课题,借鉴在人脸识别、人体跟踪等领域产生的新思想来提高自身的效率,表情分类也可以借鉴语音、汉语词性等分类方法中的思路来降低时间开销;其二是通过分析各部分之间的联系来对系统进行改进,譬如结合提取的表情特征的特点来选择相对合适的分类算法等等。考虑到实验室刚刚涉足表情识别研究领域,对其研究还主要停留在对文献的阅读上,所以本次研究首先是对目前较为成熟的表情识别系统进行分析,从其发展的趋势以及目前实际应用的需要出发,确定了针对动态性分析的表情识别平台接口,大大提高了表情识别系统功能的模块化程度。在这个平台上,实验者可以快速利用不同的技术构建出新的表情识别系统,同时也可以方便地看到局部算法的改进对整体识别率的改善效果。其次,结合表情识别系统框架结构,通过对各个部分的算法进行合理选择,构建了一个表情识别系统实例,系统实例满足提出的平台接口要求。最后需要说明的是,为了对表情识别技术有更为深入的认识和后续实验的需要,作者利用实验室资源,构建了一个含840段视频序列的基本人脸表情数据库,并结合建立的表情识别系统对数据库中的表情序列进行识别,综合识别率达到了91.7%,比较理想。实验结果一方面证明了本章设计的表情识别平台的合理性,另外一方面也为表情识别系统走向实际应用奠定了一定的基础。

### 3.2 平台设计

#### 3.2.1 表情识别系统的分类

通过对2.3.1节表情自动识别系统框架的认识,可以知道系统包含三个主要部分,各个部分有多种解决策略,并且依据实验环境的不同和选择的特征的差异,还存在一些可以自由选择的处理部分,譬如图像归一化处理等等。

根据选择的解决策略的差异以及与其它方法的结合程度可以将目前已有的表情识别系统分为五大类:

基于变形(deformation)特征提取的表情识别系统;

基于行为(motion)特征提取的表情识别系统;

基于混合特征提取的表情识别系统；

联合表情识别系统；

多模式表情识别系统；

其中第一类和第二类的区别主要是特征提取的方式不同，前者有代表性的系统是 Lyons 等人提出的基于 Gabor 小波进行特征提取的表情识别系统<sup>[26]</sup>；后者有代表性的系统是 Essa 等人提出的包含自动人脸检测的复杂视觉系统<sup>[16]</sup>。其中面部行为采用全局致密光流和人脸三维模型进行提取，三维模型的引入能够使得在了解人脸纹理结构的同时，还能对肌肉的激励作用引起的表情动作进行描述，最后对 52 个正面人脸表情序列进行分析，识别率为 98%。基于混合特征提取的表情识别系统结合前两类系统各自的优势进一步提高了识别效率。Bartlett 提出的混合系统识别模式框图见图 3.1。

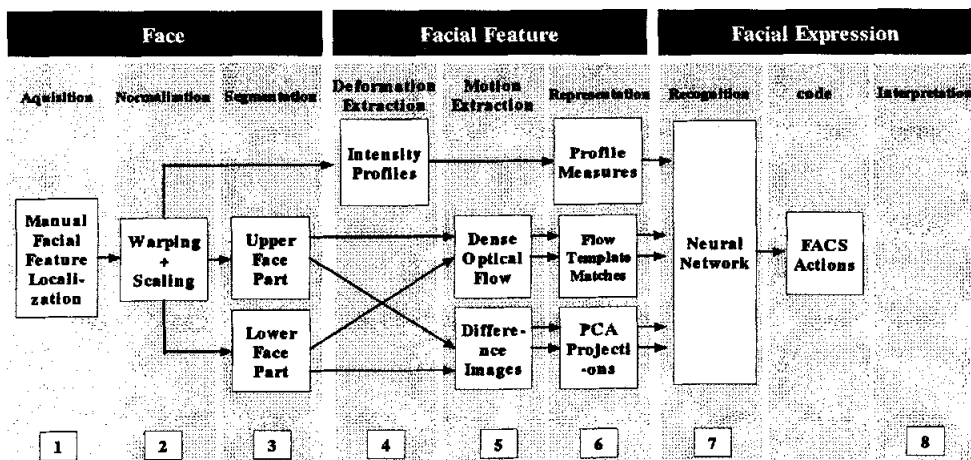


图 3.1 基于混合特征提取的表情识别系统框架图<sup>[27]</sup>

人脸表情识别性能的提高不仅仅可以通过优化利用面部表情动作，还可以借助于面部器官特征所反映出来的性别、年龄以及种族等信息。联合表情识别系统就是根据这种思想提出来的。其中有代表性的研究是 Lanitis 等人提出的系统<sup>[28]</sup>，他们利用二维活动形状模型(ASMs)来匹配人脸得到面部表观参数，从参数中分析出人的头部的姿态、性别以及表情变化，参照这些信息来进行情感的推测。

大多数的人脸表情识别系统都是单一模式的，它们主要是研究面部特征运动对人的表情的反映。从日常生活中可以发现，当人的心理情感发生变化时，外在表现不仅仅只有面部特征运动的变化，而且很多时候还表现在语调、身体运动方面，譬如高兴的时候，身体会扭动，语调也会变得高昂轻快等等。所以多模式通道信息的获取能提高表情推测的正确率，对于实际应用来说鲁棒性更强。目前常见是结合语音的双模式表情识别系统，Cohn 等人集中分析语音的基本频带部分，因为这个部分的特征能更好地反映人的情感的变化，并且易于提取<sup>[29]</sup>。

### 3.2.2 平台设计策略

本次研究设计的表情识别平台和一般的表情识别系统并不完全一致，区别在于一般的表情识别系统是从整体的性能出发来进行构建的，往往是先确立需要进行识别的表情类别，然后进行特征的选择，根据特征的特性选择合适的分类方法，最后构成一个完整的系统，而表情识别平台的重点是接口。无论识别的情感种类有多少种，也不需要考虑各个部分采取的方法是何种类型，只要满足两种主要接口，都可以纳入到这个表情识别平台中来进行算法推演和系统构建。

平台两个主要接口是：

基于连续帧的人脸图像接口

基于特征流的分类信息接口

人脸面部表情是由面部肌肉的运动产生的，这是个动态的过程，肌肉运动的速度、激烈程度等因素反映了人的不同心理历程，因此在设计表情识别系统的时候，针对的应该是包含表情变化的视频序列，而不是静止的图像。基于连续帧的人脸图像接口是针对动态表情分析要求而设定的。

同样，对于分类来说，也不再是采用几何距离就能对表情类型作出判断，表情的特征也不仅仅是几十个或者几百个数据，特征流的概念继承了表情的动态性，同时也对分类机制提出了更高的要求，通过对大量并且连续的表情样本的学习建立合理的表情模型，运用统计的思想，对表情的类别作出基于概率的推测。

平台结构模型见图 3.2。

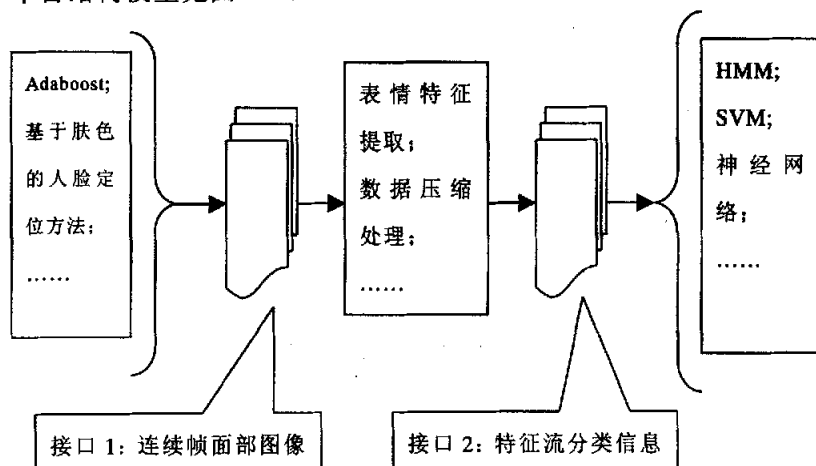


图 3.2 表情识别平台结构模型

从平台的结构模型可以看出，对于多模式的表情识别系统以及联合表情识别系统暂时无法在平台上实现。

### 3.2.3 基于平台的系统实例

本章后续内容将围绕构建一个可以实际运行的系统来展开, 首先获得一组表情视频序列作为数据源, 然后通过 Adaboost 方法来定位跟踪人脸, 采用第四章将要介绍的仿射变换方法对人脸图像进行几何归一化处理并进行误差矫正, 得到连续帧的面部图像。利用光流特征提取方法进行表情特征的提取, 并利用第五章中介绍的本征向量选择方法建立合适的表情本征空间, 对表情信息进行重新度量, 最后将每一帧的新的表情度量信息串行连接构成用于分类的特征流信息送入混合隐马尔可夫模型进行识别。

## 3.3 基本人脸表情库的建立

### 3.3.1 建立标准

近年来, 国外建立了不少用于人脸表情分析实验的数据库。CMU 的 PIE 数据库在 CMU 的 3D Room 中录制完成。该数据库包含 13 个姿态, 43 种光照条件, 收集了人脸常见的 4 种表情: 中性表情、微笑、眨眼、说话, 规模为 68 人<sup>[30]</sup>。CMU 的 Cohn-Kanade 人脸表情数据库包含 210 人的灰度图像动作序列, 是一个基于形变单元编码的人脸表情数据库, 它根据人脸运动编码系统, 把脸部动作分为 46 种 AU, 采集了人脸的大部分基本 AU 的动作序列<sup>[31]</sup>。MIT 人脸库有 16 人, 每人 27 幅图片, 有光照、尺度和头部偏转变化<sup>[32]</sup>。UMIST 人脸库有 20 人 564 幅图像, 包含从正面到侧面的各个角度<sup>[33]</sup>。Kodak 人脸库为彩色图像, 有尺寸、姿态和光照变化<sup>[34]</sup>。

国内清华大学的吴丹和林学闯建立了一个适合我国人脸表情研究的、公用的、检验人脸识别、表情识别与合成算法的标准数据库<sup>[35]</sup>。该数据库采集时采用正面充分光照, 从 3 个不同的视角记录图像数据, 内容包括了 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, 提供了一个中文语音动态视位系统, 并对表情序列采用了 FACS 进行评价。

但是, 已有的数据库很多只提供了表情图片, 并不是作为视频序列数据提供, 不适合进行动态表情分析; 包含了视频序列的表情库很多也不是免费可以获得, 因为本次研究主要是作为一个试探性研究, 并不需要和其它算法进行识别率的比较, 同时, 表情数据的采集也是人脸表情识别系统中一个非常重要的部分, 和最终的特征提取和表情分类都有直接的关系, 所以, 在研究比较成熟的表情数据库建立技术的基础上, 建立自己的数据库, 可以在以后的处理中更加灵活地解决可能碰到的问题, 在人脸表情识别系统的开发方面也将会积累宝贵的经验。

本次研究采集的人脸表情数据库是针对基本表情而建立, 被定名为基本人脸表情数据库 BFED(Basic Facial Expression Database)。主要参数见表 3.1。

表 3.1 BFED 库主要参数

被拍摄对象及环境	
对象数目	8
男	50%
女	50%
人种	黄种人
年龄	23 岁-25 岁
装饰	没有附带配件, 没有头发干扰
拍摄环境	背景单一, 光照均匀
图像序列	
单频图像	320*240(pixel), 24 位彩色
视频速率	7 帧/秒
视频长度	3 秒
姿态	正面
情感类型	7 类
单一情感序列个数	15 段
总序列个数	840 段
AU 典型组合	
高兴	6
伤心	1+4
厌恶	4+6+7
吃惊	1+2+5
愤怒	4+5
恐惧	1+4+5+7

需要说明的是, 心理学研究表明, 表情一般持续时间为 250 毫秒到 5 秒之间, 同时, 拍摄表情序列时, 要求实验对象的表情变化模式为: 准备→启动→高潮→放松→结束, 所以最终选定的表情变化时间长度为 3 秒。实践证明, 3 秒的时间长度让实验对象感觉比较自然, 并且能够按照既定的模式完成整个表情变化过程。

另外, 目前在心理学研究中对于人脸表情的强度的定义并不明确, FACS 系统中只是简单地分了几个等级。在表情数据采集时, 我们以实验对象自我感觉比较自然为标准来限定表情强度。

对于基本表情与面部器官变化的对应关系我们参考了 FACS AID 系统<sup>[13]</sup>以及金辉和高文提出的表情结构特点表<sup>[21]</sup>, 最终对应关系见表 3.2。

表 3.2 基本表情与面部器官变化的对应关系

表情	额头、眉毛	眼睛	脸的下半部
惊奇	眉毛抬起，变高变弯；眉毛下的皮肤被拉伸；皱纹可能横跨额头	眼睛睁大，上眼皮抬高，下眼皮下落；	下颌下落，嘴张开，唇和齿分开，但嘴部不紧张，也不拉伸
恐惧	眉毛抬起并皱在一起；额头的皱纹只集中在中部，而不横跨整个额头	上眼睑抬起，下眼皮拉紧	嘴张，嘴唇轻微紧张，向后拉；
厌恶	眉毛压低，并压低上眼睑	在下眼皮下部出现横纹，脸颊推动其向上，但并不紧张	上唇抬起；下唇与上唇紧闭，推动上唇向上，嘴角下拉，唇轻微凸起；鼻子皱起；脸颊抬起
愤怒	眉毛皱在一起，压低；在眉宇间出现竖直皱纹	上眼皮拉紧，抬起或不抬起；上眼皮拉紧，眉毛压低；眼睛瞪大，可能皱起；	唇紧闭，唇角拉直或向下；鼻孔可能张大
高兴	眉毛稍微下弯	下眼睑下边可能有皱纹，可能鼓起，但并不紧张；鱼尾纹从外眼角向外扩张	唇角向后拉并抬高；一道皱纹从鼻子一直沿伸到嘴角外部；脸颊被抬起；
悲伤	眉毛内角皱在一起，抬高，带动眉毛下的皮肤	眼内角的上眼皮抬高	嘴角下拉；嘴角可能颤抖

### 3.3.2 表情序列的后期处理及样例

表情序列的后期处理主要是对不合格样本予以排除。不合格样本按照差错成因可以分为两类：

1. 由于光源变化、摄像机参数变化、头部运动等因素造成样本不合格；
2. 由于对表情理解不到位，或者在表情变化过程中受到干扰而导致表情没有按照既定模式进行变化造成样本不合格。

为了保证基本表情的标准性，对于各个实验对象都要求按照统一的模式进行表情变化，但是因为人的面部肌肉的差异，以及个人平常的生活习性，在采集数



据时，在表情的自然性和既定模式之间会出现一些矛盾，所以最后还会参照标准对表情序列进行再次筛选。部分合格表情序列样例帧见图 3.3。



(a) 伤心表情序列



(b) 高兴表情序列



(c) 恐惧表情序列



(d) 愤怒表情序列



(e) 厌恶表情序列



(f) 吃惊表情序列

图 3.3 BFED 库表情序列样例帧

### 3.4 表情识别系统实例采用的关键技术

依赖于平台构建一个可以实际运行的系统。对其中人脸定位部分、特征提取部分和表情分类部分采用的技术进行介绍，满足平台要求得到的接口数据，实现对最终输入表情序列的分类。

在人脸定位阶段，考虑到表情特征需要依据前后两帧图像的变化来提取，所以对前后帧中表情区域定位的准确性要求较高。基于 Adaboost 方法的面部区域定位技术是目前被公认为准确率比较高、代表基于统计方法未来趋势的一种技术，本节将对这种技术进行介绍并应用到构建的系统中来。

基于行为的特征提取方法直接分析表情变化时面部的行为，往往是基于前后帧进行特征计算，提取的特征包括面部器官几何位置的变化、形状的变化以及对特征区域的纹理形变。正因为基于行为的特征提取方法较基于变形的特征提取方法能更加充分的对表情变化进行描述，同时，对表情变化的时间性也有较好的说明，所以，在目前的基本表情识别技术研究以及日渐兴起的自然表情、混合表情识别技术研究中，基于行为的特征提取方法往往被采用。光流特征提取方法是基于行为的特征提取方法中比较有代表性的一种。

隐马尔可夫模型能够解决时变的非平稳信号或过程的模型化问题，用概率或统计范畴的理论成功地解决了怎样辨识具有不同参数的短时平稳的信号段的问题。它既可描述瞬态的随机过程又可描述动态的随机过程的转移特性。人的面部表情图像序列可以看成几个模型之间的转换过程，可以用隐马尔可夫理论进行分类。

#### 3.4.1 基于 Adaboost 方法的面部区域定位技术

Adaboost 算法是一种分类器算法，是由 Yoav Freund 和 Robert E. Schapire 在 1997 年提出来的<sup>[36]</sup>。其基本思想是利用大量的分类能力一般的简单分类器 (weaker classifier) 通过一定方法叠加起来，构成一个分类能力很强的强分类器，因为我们知道，发现大量而粗略的经验规则要比找到一条高度准确的预测规则容易得多。理论证明，只要每个简单分类器分类能力比随机猜测要好，当简单分类器个数趋向于无穷时，强分类器的错误率将趋于零。

Adaboost 用于人脸检测时，从人脸中抽取大量的一维简单特征。这些简单特征都有一定的人脸和非人脸区分性。最终系统使用数千个一维简单分类器，组合起来达到很好的分类效果。

Adaboost 算法描述如下：

已知  $n$  个训练样本  $\{x_1, y_1\}, \dots, \{x_n, y_n\}$ ，其中  $y_i = \{0, 1\}$  对应样本的假和真；在训练样本中共有  $m$  个假样本， $l$  个真样本；待分类物体有  $k$  个简单特征，表示为  $f_j(\bullet)$ ，

其中  $1 \leq j \leq k$ ，对于第  $i$  个样本  $x_i$ ，它的特征为  $\{f_1(x_i), f_2(x_i), \dots, f_j(x_i), \dots, f_k(x_i)\}$ ，对每一个输入特征  $f_j$  有一个简单二值分类器。

定义 第  $j$  个特征的简单分类器为

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

分类器由阈值  $\theta_j$ ，一个偏置  $p_j$  决定，其中偏置  $p_j$  决定不等式方向，只有  $\pm 1$  两种情况。

训练目标 找出最优的  $T$  个简单分类器，优化组成一个强分类器。

训练方法

1. 设  $\omega_{t,i}$  为第  $t$  次循环中第  $i$  个样本的误差权重，对训练样本中的误差权重按如下公式初始化：对于  $y_i = 0$  的样本， $\omega_{t,i} = \frac{1}{2m}$ ，对于  $y_i = 1$  的样本， $\omega_{t,i} = \frac{1}{2l}$ 。

2. For  $t = 1, \dots, T$ . a. 权重归一化，使得  $\omega_i$  为  $\omega_{t,i} \leftarrow \frac{\omega_{t,i}}{\sum_{j=1}^n \omega_{t,j}}$ 。b. 对于每个特征  $j$ ，训练出其简单分类器  $h_j$ ，也就是确定阈值  $\theta_j$  和偏置  $p_j$ ，使其目标函数  $\varepsilon_j = \sum_i \omega_{t,i} |h_j(x_i) - y_i|$  达到最小。c. 从 b 中确定的简单分类器中，找出一个具有最小的错误  $\varepsilon_t$  的简单分类器  $h_t$ 。d. 对所有样本的权重进行更新  $\omega_{t+1,i} = \omega_{t,i} \beta_i^{1-\varepsilon_i}$ 。e. 确定方法：如果刚才  $x_i$  被  $h_t$  正确分类， $\varepsilon_i = 0$ ；反之  $\varepsilon_i = 1$ 。 $\beta_i = \frac{\varepsilon_i}{1 - \varepsilon_i}$ 。

3. 最后的强分类器是：

$$h(x) = \begin{cases} 1 & \sum_{i=1}^T a_i h_i(x) \geq 0.5 \sum_{i=1}^T a_i \\ 0 & \text{otherwise} \end{cases} \quad \text{其中 } a_i = \log \frac{1}{\beta_i} \quad (3.2)$$

以上训练的物理意义可以表述为：当已经提取的分类器对于某些样本分类正确，那么减小这些样本的权重。当分类错误，增加这些样本的权重。这样，后面训练提取的简单分类器就会更加强化对这些分类错误样本的训练。

AdaBoost 算法有如下优点：

1. 充分利用弱算法产生的预测规则的准确性，理论上可以达到任意精确的程度；
2. 容易处理弱算法提供的实数值形式的分类预测；
3. 克服了早期 Boosting 算法的许多实际应用的困难。

### 3.4.2 光流技术应用用于表情特征提取

#### 运动场与光流场

物体与摄像机之间的任何相对运动都将导致视平面上空间物体对应点的变化,严格的说,空间运动在视平面上引起的对应运动称为运动场。

从图像中可测的是图像辐照度的变化,这种变化并不总是与物体和摄像机之间的相对运动对应,如单色圆球绕其直径自转;另外,图像中光源的运动而不是视频目标的运动会导致灰度的变化。因此这种从图象上测得的变化并不完全反映实际的运动情况,称之为表观运动(apparent motion)。所谓光流(optical flow)是指亮度模式引起的表观运动,光流场是一个二维矢量场,它包含的信息即是各像点的瞬时运动速度矢量信息。就大多数现实世界的投影运动而言,光流分析是有效的运动场估算方法。而且,运动估算研究的是“视在运动”而非真实运动,可以不用考虑任何与图像亮度变化无关的运动,在这一点上来说,可以将运动场与光流场对应起来。

#### 光流算法与表情

由 Horn 和 Schunk 在 80 年代早期建立的“光流分析法”,是数字视频处理学科中二维运动估算的重要方法<sup>[37]</sup>。此类算法研究的对象是二维运动(或称投影运动),即真实世界中的三维运动以透视或正交的投影方式在图像平面上形成的“视在运动”,这一运动对应着图像中不同目标的相对位置的变化,也就对应着图像前后帧相应位置灰度的变化, Horn 和 Schunk 将图像平面特定坐标点上的灰度瞬时变化率定义为“光流矢量”。本章利用 Horn 和 Schunk 提出的基于时空梯度的运动分析方法来分析视频序列中人脸的表情变化信息。

以 Horn 与 Schunk 为代表提出的基于梯度的光流计算方法,比较适合于皮肤的变形计算,而且计算量比较简单,只是逐点地估计位置的瞬时速度场,算法主要公式见式(3.3)和式(3.4)。

$$\frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial t} = 0, \quad \text{令 } u = \frac{dy}{dt}, \quad v = \frac{dx}{dt} \quad (3.3)$$

$$E_s = \iint (u_x^2 + u_y^2 + v_x^2 + v_y^2) \mu_x d_y \quad (3.4)$$

其中  $V=(u, v)^T$  即为光流矢量,  $f$  是图像中对应点的灰度值,在基于梯度的方法中,时空梯度之间的关系是及其重要的,它构成了对光流计算的一个重要的约束。设在时间  $t$  和  $t+dt$  时有两幅连续的投影,在投影图上有灰度为  $f(x,y)$  的象素点,这里  $x,y$  为该点的坐标,该点在  $t+dt$  上移至  $f(x+dx,y+dy)$  时,满足  $f(x,y,t) = f(x+dx,y+dy,t+dt)$ 。当变化是连续的,而且两帧间的时间间隔  $dt \rightarrow 0$

时,由泰勒级数展开可以得到式(3.3),式(3.3)常称为梯度基本约束方程;依据同一运动物体引起的光流场应该是连续的、平滑的,即同一物体上相邻点的速度是相似的,那么其投影到图象上的光流变化也应该是平滑的这一特点,提出了一种加在光流场上的附加约束,式(3.4)就是为了解决不适定求解问题而引入的平滑性约束项,由式(3.3)和式(3.4)可以联立求解光流矢量。

### 3.4.3 MHMM 应用于表情分类

隐马尔可夫模型 HMM (Hidden Markov Model)是一个统计模型,具有十分丰富健壮的数学结构,适用于动态过程时间序列建模并具有强大的时序模式分类能力,理论上可处理任意长度的时序,它为 HMM 提供了一个非常广泛的应用范围。目前,HMM 已经成功地用于语音识别,而人脸识别和人脸表情识别是 HMM 的又一个新的应用领域。

一个 HMM 模型由两个随机过程组成,一个是马尔可夫链,由初始状态分布概率  $\pi = \{\pi_i\}$  和转移概率矩阵  $A = \{a_{ij}\}$  描述,输出为状态序列,但是这个输出序列隐藏在模型中;另一个随机过程由  $B = \{b_j(O_i)\}$  描述,输出为观察序列但不一定是马尔可夫过程,这也是 HMM 名称的由来。

设有观察序列  $O = O_1 O_2 \cdots O_T$  和状态集  $S = \{S_1 S_2 \cdots S_T\}$ , 定义 HMM 为三元组  $\lambda = (A, B, \pi)$ , 其中  $A = \{a_{ij}\}$  为状态转移概率矩阵,  $a_{ij} = P\{S_j \text{ at } t+1 | S_i \text{ at } t\}$ ,  $1 \leq i, j \leq N$ ;  $B = \{b_j(O_i)\}$  为观察符号概率分布,如果  $B$  是离散的并有  $M$  个观察值  $\{v_1 v_2 \cdots v_M\}$ , 则称此时的隐马尔可夫模型为离散隐马尔可夫模型 (DHMM), 并且

$b_j(O_i) = P\{O_i = v_k | S_j \text{ at } t\}$ ,  $1 \leq j \leq N, 1 \leq k \leq M$ ; 如果  $B$  是连续的并用 Gaussian 混合概率分布来描述,则称此时的隐马尔可夫模型为高斯隐马尔可夫模型 (GHMM), 并且有  $b_j(O_i) = \sum_{k=1}^M c_{jk} N(O_i | \mu_{jk}, \Sigma_{jk})$ , 其中  $c_{jk}$  为第  $j$  次混合系数,  $\mu_{jk}$

和  $\Sigma_{jk}$  分别为观察向量的均值和协方差矩阵;  $\pi$  为初始状态分布概率, 记为  $\pi = \{\pi_i\}$ , 其中  $\pi_i = P\{S_i \text{ at } t=1\}$ ,  $1 \leq i \leq N$ 。在模式识别中, 高斯分布是最容易接近原始分布的一种, 对于比较复杂的情况, 往往利用多个高斯混合就可以基本代表原始数据的分布了, 这种利用多个高斯模型来表示观测值概率的模型被称为混合高斯隐马尔可夫模型 (MHMM)。

在人脸表情识别系统中, 为每一种表情建立一个混合高斯隐马尔可夫模型, 如果是将面部分不同区域譬如眼睛区域、嘴唇区域分开进行表情度量的话, 则还需要为每个区域每种表情分别建立混合高斯隐马尔可夫模型。状态数目是依据表情变化的模型以及预先的实验数据确定, 观测序列  $O = O_1 O_2 \cdots O_T$  由通过光流提取的表情特征向量组成, 其概率分布  $B$  用 Gaussian 混合概率分布描述。

对应于各个基本表情的 MHMM 构成一个人脸表情分类器, 对于每个 MHMM 求  $P(O|\lambda_c), 1 \leq c \leq 6$ , 如果取  $c' = \arg \max_{1 \leq c \leq 6} [P(O|\lambda_c)]$ , 则  $c'$  就是系统识别出的人脸表情编号。

### 3.5 表情识别系统实例构建

根据平台模块化设计的思想, 以及选定的各部分采用方法的要求的不同, 本文采用 VC++6.0 和 Matlab 相结合的方式系统的开发。其中, 人脸定位跟踪、图像归一化处理、光流表情特征提取部分利用 VC++6.0 开发, 这样可以利用到 VC++6.0 所具有的文件读写速度快、视频数据处理函数丰富的特点。表情本征空间建立、数据重新描述等过程牵涉到图像像素的分析以及大量矩阵的运算, 选择 Matlab 工具进行设计可以利用其强大的矩阵计算能力及其直观的描述语言大大加快系统开发的进程。最后的表情特征分类部分因为是在 Kevin Murphy 提供的 HMMBNT 基础工具包<sup>[38]</sup>上设计的, 所以这部分也放置在了 Matlab 环境中进行开发。

#### 3.5.1 系统流程图

根据平台规范以及前面所述的方法, 得到系统流程图见图 3.4。

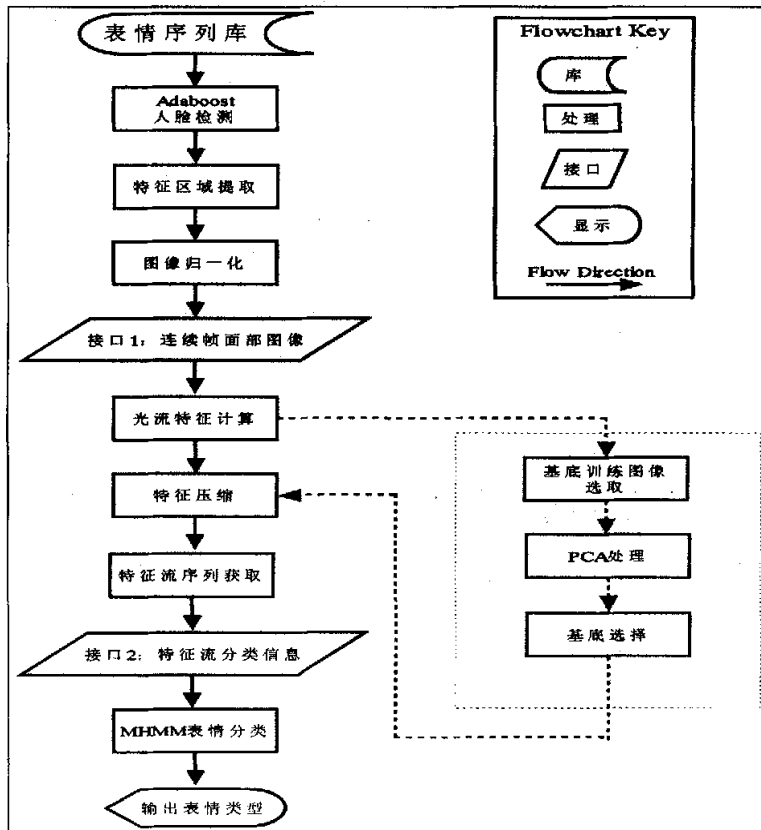


图 3.4 人脸表情识别系统流程图

### 3.5.2 数据结构

#### 1. 原始视频文件采用 Avi 格式保存

AVI(Audio/Video Interleaved)文件是 MS—Windows 的视频文件，其文件扩展名为“ . AVI”。它一般包含三部分：文件头、数据块和索引块。其中数据块包含实际数据流，即图像和声音序列数据。这是文件的主体，也是决定文件容量的主要部分。索引块包括数据块列表和它们在文件中的位置，以提供文件内数据随机存取能力。文件头包括文件的通用信息、定义数据格式、所用的压缩算法等参数。

#### 2. 图片文件在 VC++中以 OpenCV 库<sup>[39]</sup>提供的 IplImage 格式保存

#### 3. VC++与 Matlab 交换数据的格式为 CvMat 类型

#### 4. ImgCharacter 保存进行分类的表情特征流信息

ImgCharacter 采用 Matlab 环境中的矩阵格式，结构见图 3.5。每一列对应某一种表情一个序列的所有帧的特征向量的集合。因为在实验中我们选择了 4 种表情进行识别，每种表情采集了 15 组视频序列，每个表情序列为 21 帧（7 帧/秒，总共 3 秒），每帧对应的光流特征含  $N$  个  $U$  分量， $N$  个  $V$  分量，所以总的大小就应该为：

矩阵宽：4（种表情） $\times$ 15（组）=60

矩阵高：1（个视频序列） $\times$ 21（帧/序列） $\times N \times 2$ （光流值）=42 $N$

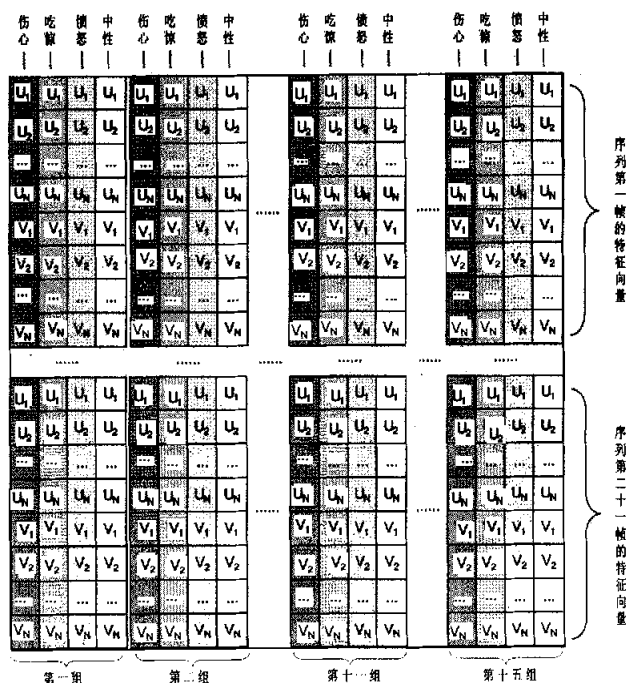


图 3.5 `ImgCharacter` 数据结构

另外，表情特征流信息的构成过程是：将每一帧表情图像得到的光流特征的两个分量串联起来，各个分量的大小由 PCA 压缩后取一定比例确定，设为  $N$ ，串联起来后构成一个特征向量，向量大小为  $2N$ ，将一个序列所有帧对应的特征向量再串联起来，就得到了大小为序列帧数的  $2N$  倍的特征流信息。

### 3.5.3 分类特征选取

在本次实验中，因为最终目的并不是强调表情识别率，而主要是对本章所构建的平台进行测试，构建一个识别系统实例；同时，考虑到心理学方面对表情定义还没有统一的定论，存在一定的模糊性，以整个人脸为基础进行表情的识别相对来说比较复杂。所以，仅仅选择眼睛眉毛区域作为表情特征提取区域，待识别表情类型也只包含了伤心、吃惊、愤怒以及中性表情四种表情，具体选择原因如下：

人的面部总共有 44 个单独的 AU 单元，通过对人脸表情库样本的分析并结合心理学研究的成果<sup>[40]</sup>，得到如表 3.3 所示的非中性基本表情与各个 AU 的对应关系。

表 3.3 基本表情与 AU 对应关系

情感类别	主要的 AU 单元					次要的 AU 单元				
	AU	AU	AU	AU	AU	AU	AU	AU	AU	AU
高兴	6	12				25	26	16		
伤心	1	15	17			4	7	25	26	
厌恶	9	10				17	25	26		
吃惊	5	26	27	1+2						
愤怒	2	4	7	23	24	17	25	26	16	
恐惧	20	1+5	5+7			4	5	7	25	26

从图 2.2 得知，人脸上半部分的 AU 单元只有 AU1、AU2、AU4、AU5、AU6 以及 AU7，各自形变特征比较明显。从表 3.3 中可以看出，在反映高兴、伤心、吃惊以及愤怒 4 种基本表情时，人脸上半部分的 AU 单元相互之间没有包含且区分度比较好。因为引起高兴表情的 AU6 的主要形变区域并不在眼睛和眉毛区域，而是分布在眼睛下方靠近面颊的部分，与其它 AU 形变位置差别较大，所以高兴表情不纳入待识别表情之列，最终确定的待识别表情类别为伤心、吃惊、愤怒以及中性表情，从人脸上半部分的对称性考虑，表情特征区域选取的时候只取了左边部分的眼睛和眉毛区域。

### 3.5.4 实验与分析

我们在 BFED 表情库上进行实验。对表情库中所有对象共 8 人的四种表情分别



选取 15 组序列进行识别, 总识别序列数目为 480 个。采样频率为每秒钟 7 帧, 每组内表情图像的帧数为 21 帧, 所取表情特征区域为人脸上半部的左边的眼睛、眉毛区域, 归一化后的表情特征区域大小为  $70 \times 70$  (pixel); 然后对每组序列的每帧图像进行时间-空间上的平滑, 用 Horn 和 Schunk 提出的基于梯度的光流方法提取特征部件区域的运动信息, 得到垂直方向 U 分量和水平方向 V 分量的运动图像; 在对运动特征进行降维时, 选择其中有代表性的运动场图像作为训练基底, 按照第 5 章介绍的本征向量选择方法建立新的表情描述空间, 并分别对特征区域的 U 和 V 分量进行降维, 特征向量长度按照首次筛选累计贡献率和二次筛选累计贡献率皆取 90% 得到大小为 156 (U、V 方向各取 78); 在对 MHMM 进行训练时, 各种表情类型分别取各个对象的 9 组对应表情序列进行 MHMM 的训练。

实验各种参数统计见表 3.4。

表 3.4 表情识别实验参数统计

参数名称	参数值	说明
识别对象	8人	取自BFED表情库
表情类型	4种	伤心, 吃惊, 愤怒, 中性
单一表情序列数	120个	8个对象, 每个对象15组序列
总表情序列个数	480个	—
单帧特征向量长度	156	U、V方向各取78
特征流长度	3276	单帧特征向量为156, 每个序列总共有21帧
分类器训练集	288	8个对象每种表情分别取9组序列
单一表情训练序列数	72	288/4
分类器测试集	192	8个对象每种表情分别取6组序列
单一表情测试序列数	48	192/4

得到的识别结果见表 3.5。

表 3.5 表情识别实验识别结果

表情类型	识别率	误识结果
伤心	79.2%	愤怒 (3), 中性 (7)
吃惊	95.8%	中性 (2)
愤怒	91.7%	伤心 (1), 中性 (3)
中性	100%	—

表 3.5 中的识别率是测试集中各种类型的表情被正确识别的序列数目与测试集中各种表情序列总数的比值。误识结果是当前表情类型被误识为其它表情类型的名称及数目。

最终的综合识别率为 91.7%，尽管此处识别率是在一定的限制条件下而且基于本地表情库得到，但从系统运行的角度来看，识别结果能够说明本章构建的平台运行良好，可以为后续表情识别系统的研究和改善提供较好的实验环境；同时，也说明了光流表情特征流信息能够很好地反映表情变化的实质，验证了 MHMM 模型对于动态短序列分类的高效性。

从误识结果可以看出，总的误识序列个数为 16 个，其中有 12 个都被误识为中性表情，主要原因初步估计包含两个方面，其一是表情特征区域较小，特征不明显；其二是因为特征提取时由于各种不稳定因素导致出现较多的干扰信息。另外，各个对象由于对表情有不同的理解以及个体面部肌肉的区别，对于伤心和愤怒两种表情的表现并不到位，所以出现了这两种表情的互相误识。

### 3.6 小结

本章首先介绍了表情识别平台的定义，之所以称之为平台，是因为它结合目前表情识别技术的发展现状，定义了几个关键的接口，使得各个主要技术都成为一个独立的模块，便于后续地改进以及新系统的开发。其次构建了一个实际的识别系统，对主要技术进行了简要介绍，最后的识别效果证明了平台的可靠性。

BFED 是作者本人采集的一个基本人脸表情数据库，在采集时结合了目前表情数据库普遍适用的标准，对 8 个对象总共采集了 840 段表情序列，满足基本表情识别的实验需要。下一步对数据库的完善主要是两个方面：

1. 增加不同年龄段的实验对象，分析年龄对于表情识别的影响因子；
2. 对每种表情按程度不同进行类型的扩展，譬如高兴的表情将扩展到“非常高兴，兴高采烈”、“一般高兴”、“轻度高兴，微笑”等表情类别。

对数据库进行第二个方面的完善，将满足表情识别系统对量度分析的需要。

## 第4章 人脸表情图像归一化误差处理技术

### 4.1 引言

在人脸表情识别技术研究中,环境的复杂性以及人体自身的运动会对表情信息的有效提取产生干扰,预先对人脸图像进行几何归一化处理是有实际意义的。一般来说,定位跟踪只是确定人脸图像的大致位置,并没有对位置的准确性以及人脸的朝向作出说明,同时由于人与摄像机之间距离的改变,定位跟踪到的面部图像会出现大小不一的现象,几何归一化处理就是根据定位和跟踪的结果对图像进行一定的调整,将人脸面部图像变换到固定的位置、统一的朝向以及相同的大小。在基于图像序列的表情识别系统中,为了获得帧与帧之间表情变化的真实信息,几何归一化处理显得尤为重要。通常意义上所说的归一化处理包括几何归一化处理和色彩归一化处理,为了减少篇幅,本章简称几何归一化处理为归一化处理。

然而在实际情况中,归一化处理后的效果并不都很理想,通常存在误差,其主要原因有以下几点:

1. 虹膜等基准特征的自身运动导致误差;

2. 表情器官发生形变时,由于面部肌肉的相互牵引,虹膜等基准特征会被迫地产生位移,这种现象也会导致误差;

3. 视频拍摄时,图像质量有时候较差,定位方法本身精确度不高,这些因素引起的虹膜等基准特征定位的不准确也会导致误差。

本章首先对基于视频序列的图像归一化问题做了一个简单的介绍,并就目前比较有影响力的归一化方法做了归纳和评价;然后对归一化误差中最常见的平移误差进行了定量地分析,指出归一化误差会对表情特征提取产生严重干扰。最后提出加权优化匹配算法对误差进行矫正。

### 4.2 基于视频序列的图像归一化技术

#### 4.2.1 归一化问题描述

归一化处理又称为标准化处理,在人脸定位跟踪系统、人脸识别系统、表情识别系统中并非关键的一个步骤,但是却是一个系统实用化进程中必须要面对的问题。在实验环境中,为了便于处理以及为了对其它步骤的算法效果进行评价,往往会要求实验对象尽可能保持身体处于静止状态,摄像机与实验对象的距离总是保持相对的固定,即使必须考虑运动情况的时候,也是尽可能的在一定的速度

限制之内，以事先约定的方式运动，同时，对光照条件、背景等外部环境也加入了诸多限制。但是在实际应用中，这种限制是不合理的，这也是目前很多系统不能马上投入到实际应用中去的主要原因。归一化处理就是针对这类问题而提出的。因为很多实验往往是对正面人脸图像进行分析和处理，所以需要将存在非平面角度偏转、平面内旋转以及缩放的人脸图像统一变换到正面并且相对固定的大小上来。图 4.1 中第一行图像为定位到的人脸，但是存在一定的偏移，这种偏移势必会影响到后续的分析，对其进行几何归一化处理后，处理效果如第二行图像所示。



图 4.1 几何归一化效果<sup>[41]</sup>

#### 4.2.2 归一化问题处理方法

归一化处理方法通常基于三个层次：仿射变换，主特征对准以及基于像素级的致密特征对准。不同的层次对应不同的特征。

目前广泛采用的归一化方法是仿射变换(affine transformation)方法<sup>[42][43]</sup>。首先利用虹膜等基准特征确定眼睛的大致位置，然后结合器官的对称性对双眼的位置进行标定，依据双眼的位置及相互之间的距离对人脸图像进行仿射变换，使面部器官变换到相应固定的位置上去，实现图像的归一化处理，有时也利用眼睛与嘴唇之间的相对位置关系来达到归一化处理的效果。仿射变换主要的优点是处理方式直观、计算量小，当面部基准特征比较稳定时，仿射变换方法能对存在平面内旋转、缩放、相对位移的面部表情器官图像进行有效地归一化处理。

活动形状模型 ASMs(Active Shape Models)<sup>[44]</sup>和活动表观模型 AAMs(Active Appearance Models)<sup>[45]</sup>是基于主特征的层次中比较影响力的归一化处理方法。活动形状模型方法将全局形状子空间限制融合到局部纹理匹配中来。通过不断修改局部搜索策略与全局限制，使得最终的迭代在人脸面部轮廓上得到收敛。活动表观模型方法建立了一个包含纹理和形状信息的模型，通过不断优化模型参数，使得模型能准确地匹配到人脸的各个轮廓曲线。这类方法精确度比较高，能获取到局部器官特征的边缘曲线，但是在建模和优化的过程中，其繁琐的人脸特征点标定工作以及开销过于庞大的迭代计算让其还只能停留于非实时处理阶段。

基于光流技术，Beymer 等人<sup>[46]</sup>提出了一种名为矢量器(vectorizer)的方法，

在这种方法中,前后帧图像的二维形状变化被表示为一种光流场。通过分析光流场的特性,能够得出图像的几何变化,对这种变化进行逆方向的补偿,能得到较好的归一化效果。矢量器方法的性能依赖于光流计算的精度,其不稳定性的合理解决是目前一个研究的热点。

### 4.3 人脸表情图像归一化误差分析

因为人脸在三维空间活动时具有 3 个自由度( $X, Y, Z$ ),所以当人脸面部平面和  $XY$  平面保持一致时,归一化误差可以分为以下 6 种类型: $X$  方向平移误差, $Y$  方向平移误差, $Z$  方向缩放误差,绕  $X$  方向旋转误差,绕  $Y$  方向旋转误差,绕  $Z$  方向旋转误差。本节主要针对基于仿射变换的归一化处理中最为常见的  $X$  方向和  $Y$  方向的平移误差进行分析。

一般来说,出现归一化平移误差时,表情特征区域相对位移的范围只有几个到几十个像素。但是,对于表情特征提取来说,归一化平移误差导致的干扰却是比较大的。结合基于光流的表情特征提取方法,定量地分析归一化平移误差对表情特征提取的干扰情况。

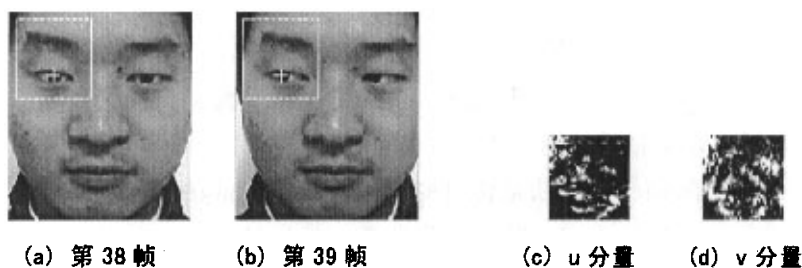


图 4.2 存在归一化误差的表情特征区域光流计算结果

图 4.2(a)和图 4.2(b)是某一段表情图像序列中第 38、39 帧归一化后的效果图,为强调归一化后表情特征区域的平移误差,在采集时限制了头部的转动和前后伸缩运动,仅允许头部在平面内移动。人脸面部区域的初始定位和跟踪采用 Adaboost 方法[36],因为 Adaboost 方法是一种基于统计分类的方法,所以尽管每帧中人脸图像大小近乎一致,但是最终得到的面部图像定位框却在一定范围内漂移。归一化时设定标准框(在图 4.2 中即为人脸图像边框)大小统一为  $145 \times 165$  (pixel),十字叉表示定位到的瞳孔中心位置,在将归一化后的面部图像放置到标准框中时,要求十字叉中心点处在相对标准框左上角点水平偏右 35 像素,垂直偏下 50 像素处。标准框中的白色矩形框是参照瞳孔中心位置并且依据人脸器官的结构分布特性确定的包含眼睛和眉毛的表情特征区域,矩形框大小为  $60 \times 63$  (pixel)。

先由图 4.2(a)和图 4.2(b)中对应表情特征区域图像计算得到光流矢量值,

然后将光流矢量的  $u$  分量值和  $v$  分量值按照绝对值大小映射为图像灰度值，显示结果如图 4.2 (c)和图 4.2 (d)所示。

因为图 4.2 (a)和图 4.2 (b)中表情器官并没有发生形变，不存在表情变化，只是因为归一化时依据的基准特征发生了平移，从而使得表情特征区域无法对准，所以图 4.2 (c)和图 4.2 (d)中的光流值主要是由于归一化误差所导致。

记  $U_{\max}$  为特征区域图像中各点光流  $u$  分量的绝对值的最大值， $V_{\max}$  为图像中各点光流  $v$  分量的绝对值的最大值， $M$  为图像宽度， $N$  为图像高度。

$$U_{\max} = \max(|u_{i,j}|), \quad V_{\max} = \max(|v_{i,j}|) \quad i \in (0, M-1), j \in (0, N-1) \quad (4.1)$$

由公式 (4.1) 可以得到：

$$U_{\max} = 5.31, \quad V_{\max} = 5.46$$

采用和上一段表情序列一致的表情变化速率和表情序列长度，另外采集了 6 个基本表情图像序列。同时，人的头部运动被严格限制，双眼被要求注视一个固定物体，保证瞳孔位置不发生变化，这样归一化后效果比较理想，计算得到的光流矢量基本反映了实际的表情变化情况。

取包含眼睛和眉毛的表情特征区域进行分析，记  $U'_{\max}$  为序列各帧中特征区域各点光流  $u$  分量的绝对值的最大值， $V'_{\max}$  为各点光流  $v$  分量的绝对值的最大值， $M$  为图像宽度， $N$  为图像高度， $F$  为图像序列总的帧数。

$$U'_{\max} = \max(\max(|u_{i,j}|))_k, \quad V'_{\max} = \max(\max(|v_{i,j}|))_k \\ i \in (0, M-1), j \in (0, N-1), k \in (1, F) \quad (4.2)$$

由公式 (4.2) 可以得到各个序列表情变化时的  $U'_{\max}$  和  $V'_{\max}$  值，见表 4.1。

表 4.1 无归一化误差时表情光流特征各分量最大值

表情类型	愤怒	厌恶	伤心	吃惊	高兴	恐惧
$U'_{\max}$	6.46	5.96	5.82	5.37	5.51	6.04
$V'_{\max}$	9.07	9.41	5.07	7.09	8.84	8.57

进而可以得到比较数据，见表 4.2。

表 4.2 光流数据比较

归一化平移误差产生的光流		表情变化产生的光流		比例
$U_{\max}$	5.31	$U'_{\max}$	6.46	82.20%
$V_{\max}$	5.46	$V'_{\max}$	9.41	58.02%

由表 4.2 的比例得知，归一化平移误差产生的光流分量值占到了表情变化能达到的最大光流分量值的 82.20%和 58.02%，而图 4.2 (a)和图 4.2 (b)中存在的归一化平移误差只有 5~6 个像素，由此可以证明这种误差对于表情特征提取的干扰是比较严重的。

## 4.4 归一化误差解决方案

### 4.4.1 加权优化匹配算法

在基于图像序列的表情识别系统中,表情特征的提取往往是在序列的前后帧中进行,考虑到归一化平移误差主要表现在表情特征区域发生了位置的偏移,并没有图像大小的变化,所以误差的矫正主要是将相邻的两帧依次对准。当两幅图像不存在表情特征变化时,对准的概念比较简单,本文采用模板匹配的原理从全局的角度对图像进行匹配,充分利用非基准特征的边缘和灰度信息,找到最佳匹配点,达到对基准特征偏移引起的误差进行矫正的目的。但是,当两幅图像既存在归一化平移误差,又存在相对表情形变的时候,对准时必须尽可能地保留表情特征变化的信息,让真正发生了表情形变的特征处于不对准状态。



图 4.3 存在表情形变时的图像对准效果

图 4.3 (c) 所示是图 4.3 (a) 和图 4.3 (b) 经过简单模板匹配处理后的效果图,其中,对准后的第  $i+1$  帧的边缘图像覆盖在了第  $i$  帧上面。可以明显地看出,因为眉毛部分的面积在图像中占据比例较大,其它部分相似性并不是很强,所以对准时在眉毛部分造成了局部最优,反映出的结果是眉毛静止不动,眼睛向下运动并伴有张开趋势,而实际的表情应该是眉毛和上眼睑上扬、下眼睑下垂的吃惊表情。所以,对于存在非刚性形变的表情特征区域来说,对准时要充分考虑表情的变化特性。

本文提出利用加权优化匹配算法进行归一化平移误差的矫正。加权优化匹配算法以传统模板匹配原理为基础,在进行相关系数的计算时,根据各像素点的运动剧烈程度分配相应的匹配权值,运动剧烈程度大的像素点分配较小的匹配权值,运动剧烈程度小的像素点分配较大的匹配权值。加权优化匹配算法有如下几个特点:

1. 因为视频图像的分辨率往往较低,并且面部表情器官存在非刚性形变,所以,直接利用表情特征区域中的特征曲线或者特征点的几何信息来矫正误差的方法并不可行。基于模板匹配原理进行算法的设计有利于从全局来考虑特征的对应关系,不会因为局部信息的误差造成矫正的失败;

2. 基准特征的不稳定是导致归一化平移误差的主要原因,而这种不稳定往往会表现出较大的运动剧烈程度,为基准特征分配较小的匹配权值能降低其对误差

矫正的干扰；

3. 给反映实际表情变化的特征象素点分配较小的权值，能有效地保留表情信息；

4. 对于只存在归一化误差而不存在相对表情形变的情况而言，计算得到的各点的运动剧烈程度近乎一致，因此分配的权值也基本相等，这时，加权优化匹配算法就相当于一般的模板匹配算法，同样也能对归一化误差进行有效矫正。

#### 4.4.2 算法实现

设进行归一化平移误差矫正的前一帧表情特征区域图像为 F1，后一帧表情特征区域图像为 F2，两幅图像宽皆为 M 个象素，高皆为 N 个象素。加权优化匹配算法主要步骤如下：

第一步：估算前后帧表情特征区域中各点的运动剧烈程度。光流算法可以逐点地进行前后帧象素点位移矢量的计算，所以在此被用来进行运动剧烈程度的估计。F1 对应的权值图记为 W，F2 对应的权值图记为 R，权值图大小与特征区域图像大小一致。权值图各点的值由公式 (4.3) 和 (4.4) 计算得到。

$$W(i, j) = 1 - \frac{|V_{(i,j)}|}{v_{\max}} \quad i \in (0, M-1), j \in (0, N-1) \quad (4.3)$$

$$R(i, j) = 1 - \frac{|V'_{(i,j)}|}{v'_{\max}} \quad i \in (0, M-1), j \in (0, N-1) \quad (4.4)$$

式 (4.3) 中  $V_{(i,j)}$  为以 F1 为当前帧，以 F2 为下一帧计算得到的 F1 中各点的光流矢量值， $v_{\max}$  是 F1 中各点的光流矢量的模的最大值；式 (4.4) 中  $V'_{(i,j)}$  为以 F2 为当前帧，以 F1 为下一帧计算得到的 F2 中各点的光流矢量值， $v'_{\max}$  是 F2 中各点的光流矢量的模的最大值。

第二步：将 F1 居中放置在一个大的背景图中，构成被搜索图 S，被搜索图宽为 B 个象素，高为 H 个象素，F2 作为模板 T，模板宽为 M 个象素，高为 N 个象素。其中被搜索图和模板大小关系为：B = M + 2\*X\_sur，H = N + 2\*Y\_sur。X\_sur 和 Y\_sur 为依据表情器官形变特性统计得到的平移误差上限值，包含眼睛和眉毛器官的表情特征区域一般设误差上限值为区域对应边长的 15%。被搜索图和模板见图 4.4。

同时，得到被搜索图对应的权值图，记为 W'，W' 和被搜索图大小一致，计算公式见式 (4.5)。

$$W'(i, j) = \begin{cases} W(i - X\_sur, j - Y\_sur) & i \in (X\_sur, X\_sur + M - 1) \\ & j \in (Y\_sur, Y\_sur + N - 1) \\ 0 & \text{else} \end{cases} \quad (4.5)$$

由式 (4.5) 可以看出，在被搜索图中，除 F1 部分外的其它各点对应的权值



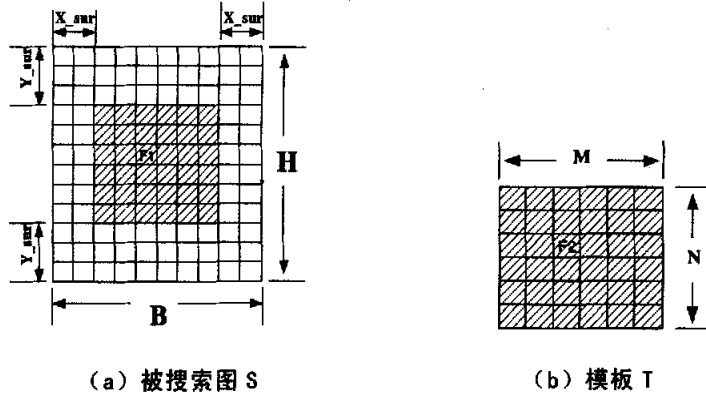


图 4.4 被搜索图和模板

为 0，所以在模板匹配时，实际纳入计算的象素点属于模板 T 与被搜索图中 F1 部分重叠的区域。

第三步：进行相关系数的计算，找出最佳匹配点，实现平移误差的估计。匹配时，模板覆盖被搜索图的那块区域叫子图  $C_{ij}$ ， $(i, j)$  为子图左上角在被搜索图 S 上的坐标（令图像左上角为坐标原点，向右为 X 轴增长方向，向下为 Y 轴增长方向），搜索范围是：

$$\begin{cases} 0 \leq i \leq 2 * X\_sur \\ 0 \leq j \leq 2 * Y\_sur \end{cases}$$

进而可以得到每一次匹配时模板 T 与被搜索图中 F1 部分重叠的区域在搜索图中的坐标位置及大小：

重叠区域起点  $(SB_x, SB_y)$ ：

$$SB_x = \max(i, X\_sur)$$

$$SB_y = \max(j, Y\_sur)$$

重叠区域终点  $(SE_x, SE_y)$ ：

$$SE_x = \min(i+M-1, X\_sur+M-1)$$

$$SE_y = \min(j+N-1, Y\_sur+N-1)$$

重叠区域宽度： $K = SE_x - SB_x + 1$

重叠区域高度： $L = SE_y - SB_y + 1$

令模板 T 中坐标系与被搜索图 S 中的坐标系一致，可以得到重叠区域在模板 T 中的坐标位置，记为  $(TB_x, TB_y)$ ： $TB_x = SB_x - i$ ， $TB_y = SB_y - j$

加权优化匹配算法中对应子图  $C_{ij}$  的相关系数记为  $R(i, j)$ ，计算公式见式 (4.6)。

$$\begin{aligned}
 R(i, j) = & \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} [W'(m + SB_x, n + SB_y) S(m + SB_x, n + SB_y) \\
 & \times R(m + TB_x, n + TB_y) T(m + TB_x, n + TB_y)] \\
 & \times \frac{1}{\sqrt{\sum_{m=0}^{K-1} \sum_{n=0}^{L-1} [W'(m + SB_x, n + SB_y) S(m + SB_x, n + SB_y)]^2}} \\
 & \times \frac{1}{\sqrt{\sum_{m=0}^{K-1} \sum_{n=0}^{L-1} [R(m + TB_x, n + TB_y) T(m + TB_x, n + TB_y)]^2}}
 \end{aligned} \quad (4.6)$$

在搜索范围内进行相关系数的遍历计算，找出相关系数最大值对应的子图，这时子图相对于被搜索图中 F1 部分偏离的距离就是对归一化平移误差的估计值。

## 4.5 实验与分析

图 4.5(a) 中前两幅图像分别取自图像序列中连续的两帧，由于基准特征偏移，归一化后的两幅图像存在平移误差，尽管没有明显的表情变化，但是计算得到的光流表情特征矢量的 U、V 分量图却反映出较大的值，如图 4.5(a) 后两幅图像所示。这种光流值对于表情特征提取来说，是一种干扰。

通过加权优化匹配算法调整后的图像及相应的光流分量图如图 4.5(b) 所示。通过寻找最大相关系数，得到两幅图像中表情器官的相对位移，利用求得的位移值对归一化平移误差进行矫正，从光流分量图可以看出，矫正后误差干扰大大减少。

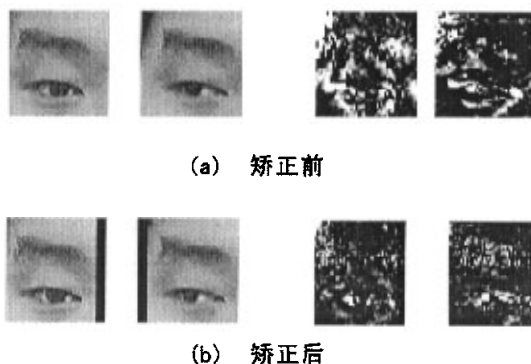


图 4.5 基准特征偏移引起的归一化平移误差矫正效果

图 4.6 中第  $i$  帧和第  $i+1$  帧归一化效果比较理想，几乎不存在归一化误差，但是存在明显的表情变化。直接运用简单模板匹配会造成表情信息丢失（见图 4.3），加权优化匹配算法为表情变化特征点分配较小的权值，对准时能比较完整地保留表情信息，结果如图 4.6(c)、图 4.6(d) 所示，光流图反映的结果是眉毛和

上眼睑上扬、下眼睑下垂的吃惊表情，与实际表情变化情况相符。

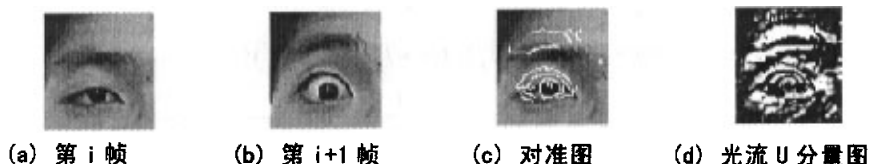


图 4.6 加权优化匹配算法对表情信息的保留效果

图 4.7(a) 中的两幅图像既存在由于基准特征偏移引起的归一化平移误差，又存在明显的表情变化，实际的表情是眉毛下拉、右侧眉头轻微抬起的厌恶表情。图 4.7(b) 显示误差得到了有效地矫正。图 4.7(a) 和图 4.7(b) 最后一幅图像反映竖直方向上的表情变化情况，也即光流  $U$  分量图，可以对比看出，矫正后眉毛下拉的表情变化特征得到了完整的保留，并且由于去掉了干扰，特征更加明显。

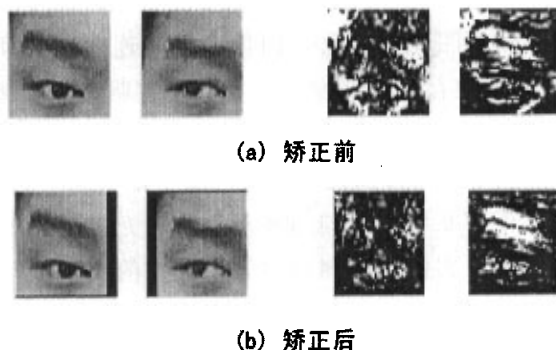


图 4.7 对同时存在误差和表情变化的图像进行矫正的效果

## 4.6 小结

在基本表情识别技术研究以及日渐兴起的自然表情、混合表情识别技术研究中，如何有效地从图像序列中得到真实的表情变化信息是一个备受关注的研究热点。在实际应用环境中，无法对人本身的运动作出如实验室环境下相似的限制，必须考虑当存在头部偏转、光照变化等因素下的干扰情况。为此，图像归一化处理作为重要的步骤纳入到表情识别系统的前期处理中。本章首先对归一化问题做了简要地描述，并就目前比较有影响的归一化解决方法做了总结。总的来说，目前的归一化解决方法的鲁棒性还值得进一步提高，对实验环境的要求相对来说还是比较高，譬如图像要求有较好的清晰度等等。活动表观模型方法的效率改进目前已有一些成果<sup>[47]</sup>，但其训练图像的标定的工作量依然比较高，寻找自动的标定方法将是其继续改进的方向。接着本章主要对表情特征区域的归一化误差问题进行了详细的分析，研究数据表明，因为图像归一化时依据的基准特征存在不稳定性，归一化后的结果会存在一定的误差，这样得到的表情变化信息无法反映出实

际的表情变化情况，最终将导致表情识别率降低。针对其中比较普遍的归一化平移误差，提出了加权优化匹配算法进行误差的矫正。加权优化匹配算法从全局的角度出发，利用统计匹配的思想对误差进行估计，同时根据像素点运动剧烈程度引入相应的匹配权值。实验结果表明，算法不仅能有效地解决归一化误差问题，而且对真实的表情变化信息予以保留。对归一化中的旋转和缩放误差同样可以采用本文提出的方法予以矫正，不过在匹配搜索时应该限定搜索的方式，如在矫正旋转归一化误差时，搜索方式应该设置为绕一定的角度进行。

下一步的工作主要是分析自遮挡的图像在进行归一化时产生的误差，自遮挡的图像在进行归一化时还牵涉到像素的合理插值技术，在进行误差分析时也将考虑插值带来的表情信息失真问题。

## 第5章 表情本征空间建立技术及改进

### 5.1 引言

“本征空间”的概念最早被 Sirovich 和 Kirby<sup>[48]</sup>用在人脸识别系统中。即通过线性变换,从采集的人脸图像中得到一组互不相关的本征向量,利用这组向量构成一个新的特征表示空间,在新空间中对人脸图像进行再次投影,得到的新特征能更好地从统计的角度对人脸图像进行描述,并且数据量大大减少。这种方法被称为主成分分析,这组向量本身还原的图像被称为本征脸 (Eigenface),表示了一种能代表人脸某一方面主要特征的面部形状和纹理,构成的新空间即是本征空间 (Eigenspace)。在表情识别系统中,待识别的是表情变化图像,因为人脸表情变化有其固有的规律,同时各个人、各种表情变化信息之间因为人脸的普遍相似性也存在很强的关联性,并且相对于人脸识别来说还避免了因为肤色差异带来的干扰,所以同样可以利用主成分分析的技术来对表情信息进行再次处理。

在所有的本征向量中,有一些对应表情信息的低频部分,反映了表情信息的主要特征;另外一些则对应高频部分,更多的是由于干扰以及次要面部形变特征引起。滤除掉对应高频部分的本征向量,用剩余的本征向量作为空间基底,能够减少新空间中表情信息的维度,并且已有研究表明<sup>[49]</sup>,只要本征向量选择适当,最终得到的识别率不仅不会因为空间基底减少而降低,还会得到更好地改善。

在解决本征向量取舍问题的时候,许多研究者采用了以贡献率作为评价准则的主成分选择方法<sup>[50][51]</sup>。通过协方差矩阵分解,对于每一个本征向量,会相应地得到一个特征值,按照特征值大小与特征值总和的比值确定各本征向量的贡献率,选择按大小排序后的前  $L$  个本征向量作为新空间的基,这  $L$  个本征向量的贡献率之和要大于一个事先设定的百分值,一般取 90% 以上。这种方法计算比较简单,尽管滤除了部分向量,但相对于原特征只是丢弃了一些干扰信息,主要表情变化特征能得到很好的保留。这种方法的不足在于它是基于纯数据统计的角度来对本征向量进行选择的,没有很好地和表情变化的特点相结合,其目标只是能最大程度地还原原始表情信息中聚类性较好的数据。从表情识别的角度来讲,所选择的本征向量反映各类表情之间易于区分的、有代表性的特征信息才是关键。

基于这种考虑,本章首先简要地介绍主成分分析技术与基于贡献率的主成分选择方法的实现过程。然后结合 FACS 系统,详细地分析了与表情变化紧密相关的面部 AU 单元和各个本征向量之间的对应关系,利用其相关程度在按照贡献率得到的特征集中进行再次筛选,最后利用实验对算法效率进行验证。

## 5.2 表情本征空间的建立

### 5.2.1 主成分分析技术

在基于视频的人脸表情识别系统中，当直接利用从视频帧中提取的表情特征作为分类依据时，得到的识别率往往比较低，主要原因是提取的表情特征只是对应图像帧中各像素点的位置及属性变化，没有体现各像素点之间的差异及其整体的分布规律。为此，通常需要对表情描述空间进行合理转换，利用新的空间基对表情信息进行重新描述。目前常用的空间转换方法是主成分分析方法<sup>[52]</sup>，在数学上又称为 K-L 变换方法。主成分分析方法本身是一种基于统计的多元正交线性变换。它利用统计理论从表情特征数据中找到一组互不相关的本征向量，以这组本征向量作为基底建立新的表情描述空间，将原始表情信息在新空间中进行投影得到新的表情特征值。因为本征向量体现的是表情数据的聚类性，所以以新的表情特征作为分类依据往往能得到好的分类结果。

在主成分分析方法中，首先将样本图像逐行首尾相连表示成一个列向量，然后将所有样本图像放置到一个大的训练集矩阵  $X'$  中， $X'$  大小为  $S \times N$ ， $N$  对应样本数目， $S$  对应列向量的维度。对训练图像进行标准化调整：

$$X = (X' - \overline{X'}) \quad (5.1)$$

式 (5.1) 中， $X$  是调整后的训练样本图像集， $\overline{X'}$  为训练样本图像集的平均影像。标准化调整使数据之间的偏移关系更加明显。

训练样本集的主成分可以通过式 (5.2) 计算得到：

$$L * V' = \Lambda * V' \quad (5.2)$$

式中， $L$  为  $X^T$  的自协方差矩阵， $\Lambda$  为特征值组成的对角阵， $V'$  为特征值相对应的本征向量组成的正交阵。

单个特征值记为  $\lambda_i$  ( $i = 1, 2, \dots, N$ )，单个本征向量记为  $V'_i$  ( $i = 1, 2, \dots, N$ )，向量维数为  $S$ ，具体形式见式 (5.3)。

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix}, \quad V' = [V'_1, V'_2, \dots, V'_{N-1}, V'_N] \quad (5.3)$$

本征向量的长度并不影响本征空间的构建，所以常把本征向量的长度缩放为单位长度，得到标准本征向量矩阵  $V$ ，转换公式见式 (5.4)。

$$V_{(j,i)} = V'_{(j,i)} / \sqrt{\sum_{m=1}^S (V'_{(m,i)})^2} \quad j \in [1, S], \quad i \in [1, N] \quad (5.4)$$

### 5.2.2 基于贡献率的本征向量选择算法

主成分分析方法计算得到的各个本征向量反映了表情数据不同的统计特征，在利用其构建的新空间中，各种表情信息之间的关系得到了很好的体现。特征值作为本征向量的一种对应权值，反应了本征向量在描述数据分布特性方面的重要程度，特征值越大，则在对应的本征向量表示的方向上数据的分布位置的方差越小。基于贡献率的主成分选择方法就是一种以特征值的大小作为评价标准来对本征向量做出选择的方法。

由式 (5.3)、式 (5.4) 得知，对应本征向量  $V_i$  的特征值记为  $\lambda_i$ 。贡献率  $r$  可以根据各向量对应的特征值计算得到：

$$r = \lambda_i / \sum_{j=1}^N \lambda_j \quad i \in [1, N] \quad (5.5)$$

设定一个判决阈值，即可以选择出合适的本征向量。

目前更为常用的是基于累计贡献率  $R$  的判决方法，累计贡献率能更加直观地说明选择的本征向量对原始信息的整体表现能力。

$$R = \sum_{m=1}^L \lambda_m / \sum_{j=1}^N \lambda_j \quad L \in [1, N] \quad (5.6)$$

选择  $L$  使  $R \in [90\%, 95\%]$ ，得到的  $L$  个本征向量取代原  $N$  个本征向量构成了新空间的基， $L$  小于  $N$ 。

对人脸上半区域左边部分进行表情特征的提取，然后利用主成分分析技术并结合基于贡献率的选择方法得到构成新空间的本征向量。图 5.1 所示的是  $R$  取 96% 时得到的 16 个本征向量还原的图像。可以看出，第一个本征向量反映的是表情变化主要的区域，其余的向量分别反映的是不同部分的表情变化统计特征。

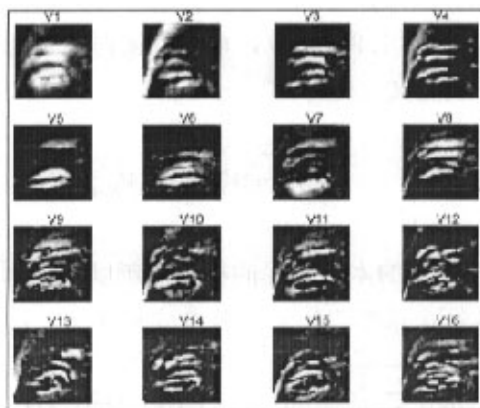


图 5.1 本征向量还原图

## 5.3 改进后的本征向量选择算法

### 5.3.1 算法描述

基于贡献率的主成分选择方法强调的是能最大程度地还原原始表情信息中聚类性较好的数据,在进行主成分计算的时候,并没有考虑人脸表情信息的变化特点,只是运用基本的数学知识对数据进行统计分析。从人脸表情识别的角度来说,当经过处理后的信息作为输入信号送到分类器中进行识别的时候,为了得到理想的分类结果,不仅对数据量有要求,而且对于每种表情来说,要求对应的输入信息能反映最有代表性的并且是容易和其它表情进行区分的特征。所以,可以结合表情变化本身的先验知识,对基于贡献率得到的本征向量进行再次筛选。

对于表情变化本质的认识,不同的心理学专家提出了各自不同的看法。其中被表情自动识别系统研究者们广泛接受的一种描述机制是 Ekman 和 Friesen 提出的人脸运动编码系统 (FACS) 系统<sup>[16]</sup>。从 2.2.2 小节的分析可以得知,人脸面部的各种表情最终能分解对应到各个 AU 上来,而且 AU 形变特征也最能反映不同表情类型之间的差别。

通过对实验图像进行观测得知,运用主成分方法得到的本征向量和 AU 基本单元并没有必然的一一对应关系,从主成分的公式推导也可以看出,在主成分的计算式中并没有引入 AU 变化信息作为附属约束,所以,直接利用各个 AU 选择对应的本征向量作为空间基并不可行。但是根据本征空间的定义,同时依据 AU 变化特征可以作为表情形变特征的基本组成单元的先验知识,可以将各个 AU 投影到本征空间,得到 AU 与每一个本征空间基的相关权值,权值大小反应了空间基在构成表情形变单元时的贡献率。如果对各个 AU 投影得到的权值进行综合考虑,可以给出一种新的评价标准,利用新的标准对本征向量进行再次筛选,能使原始表情信息在新空间中的投影值更适合作为表情分类的依据。

### 5.3.2 算法实现

首先利用表情特征提取方法提取出各个 AU 形变所产生的特征图像,经过标准化调整后各行首尾相连构成图像向量,记为  $A_i$ , 以各个  $A_i$  作为列向量构成 AU 特征图像矩阵  $A$ 。矩阵  $A$  的宽度即为选择的 AU 数目,大小记为  $K$ , 高度为表情特征图像的高与宽的乘积,记为  $S$ 。

设通过基于贡献率的主成分选择方法确定的空间基矩阵为  $V$ :

$$V = [V_1, V_2, \dots, V_{L-1}, V_L]$$

其中训练集中的每一幅图像也是采用相同的表情特征提取方法得到的表情变化特征图像,累计贡献率取  $R$ ,  $L$  为按照贡献率  $R$  选择的主成分数目,主成分向量维数为  $S$ 。



由式 (5.7) 计算得到投影权值矩阵  $P$ ， $P$  的大小为  $L \times K$ ， $P$  的每个列向量是对应的 AU 特征图像与各个空间基的相关权值向量。

$$P = V^T * A \quad (5.7)$$

对  $P$  中各个列向量对应位置的值求和得到综合权值向量  $Q$ ，综合权值向量  $Q$  的维数为  $L$ 。 $Q$  的计算公式见式 (5.8)。

$$Q_i = \sum_{j=1}^K P_{(i,j)} \quad i \in [1, L] \quad (5.8)$$

对向量  $Q$  的各分量值按照从大到小的顺序排序后再次进行贡献率累计，累计贡献率取  $R'$ 。最终选定的主成分个数记为  $Z$ ， $Z$  小于等于  $L$ 。

## 5.4 实验比较

实际操作中，获取单个 AU 形变引起的表情变化信息是比较困难的，主要原因是缺乏心理学知识支持，实验者对于各个 AU 形变的认识并不到位。考虑到本文方法会对各个 AU 投影得到的权值进行综合考虑，并不是对单个 AU 产生的权值进行分析，所以并不需要获得单个 AU 形变引起的表情变化信息。本次实验选择几种混合 AU 形变信息进行权值的获取，最后在综合评价时按照各个 AU 的混合数目进行比例的调整。图 5.2 是采用光流方法得到的几种混合 AU 对应的特征图像。

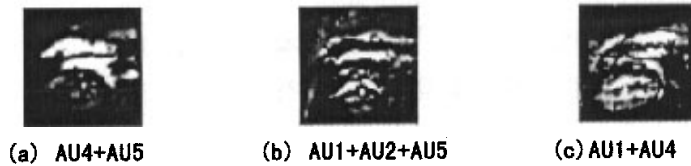


图 5.2 采用光流方法得到的混合 AU 特征图像

图 5.3 给出了基于贡献率的主成分选择方法得到的前九个本征向量以及相应的贡献率，图 5.4 是利用本文方法再次筛选后得到的前九个本征向量，从左至右、从上到下分别按综合权值评价结果从大到小进行排列，并且各个本征向量上面标注了未经筛选前的序号及相应的贡献率。

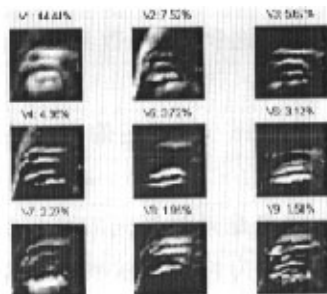


图 5.3 基于贡献率选择的前 9 个本征向量

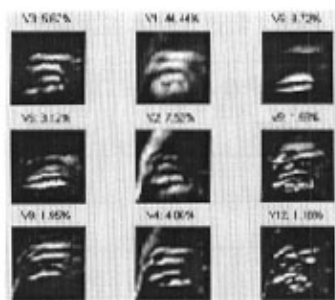


图 5.4 基于本文方法得到的前 9 个本征向量

从图 5.3 和图 5.4 可以看出, 本文方法确定的本征向量和基于贡献率的主成分选择方法确定的本征向量并不完全一致。

基于第三章构建的表情识别系统, 本章还对两种方法获得的本征向量进行了对最终识别率的影响因子的比较。为了避免因为对象个体之间的区别造成的干扰, 本次实验仅选择了一个对象的表情序列进行分析。

实验的各种参数统计见表 5.1。

表 5.1 单个对象表情识别实验参数统计

参数名称	参数值	说明
识别对象	1人	取自BFED表情库
表情类型	4种	伤心, 吃惊, 愤怒, 中性
单一表情序列数	15个	1个对象, 每种表情15组序列
总表情序列个数	60个	—
分类器训练集	36	1个对象每种表情分别取9组序列
单一表情训练序列数	9	36/4
分类器测试集	24	1个对象每种表情分别取6组序列
单一表情测试序列数	6	24/4

主成分分析时, 因为伤心、吃惊、愤怒基本表情序列包含了中性表情帧, 所以仅以这三种序列共 945 帧作为训练样本。取累计贡献率为 90%, 按照基于贡献率的特征选择方法得到的本征向量个数为 62。参照本文方法, 利用混合 AU 特征图像对新空间进行投影, 对得到的向量权值进行综合评价, 按照不同的二次累计贡献率进行重新筛选, 选定的本征向量个数见表 5.2。

识别率结果如图 5.5 所示, 原点对应的纵坐标表示采用基于贡献率的主成分选择方法得到的识别率, 星号点对应的纵坐标表示采用本文方法得到的表情识别率。横轴只代表本征向量的个数, 对应于两种方法的本征向量并不相同。数据表明, 经过本文方法重新筛选后的本征向量能更好地还原原始表情特征中易于区分

表 5.2 新选定的本征向量个数

二次累计 贡献率	重新筛选后的 向量个数	二次累计 贡献率	重新筛选后的 向量个数
50%	16	80%	36
60%	22	90%	46
70%	28	100%	62

的、有代表性的那一部分信息,当选择的本征向量个数不大于 46 个时,采用本文方法得到的识别率一直比采用原方法得到的识别率要高,当本征向量个数等于 46 的时候,本文方法就已经实现了采用原方法时需要选择至少 62 个向量才能达到的最高识别率。

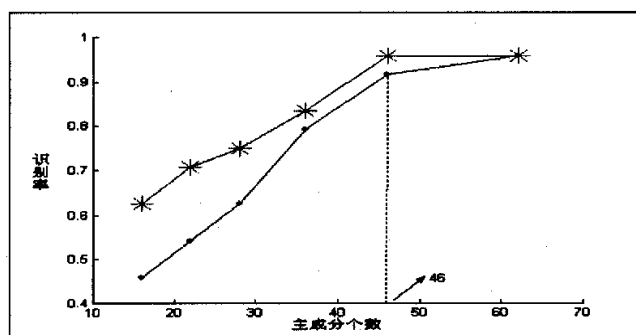


图 5.5 改进前后识别率比较

## 5.5 小结

主成分分析方法具有良好的统计分类能力以及简单直观的计算过程,因而在人脸识别、特征选取、图像归一化等领域得到了广泛的应用。其在人脸表情识别系统中的应用主要是对表情特征数据进行统计分析并构建新的特征描述空间。利用基于贡献率的主成分选择方法能够对新空间的基底进行一定的筛选,滤除次要信息,同时也到达降维的效果。但是这种方法并没有和表情变化固有的规律以及表情识别的类型有效地结合起来。本章在详细地分析面部运动单元与本征向量之间对应关系的基础上,对按照贡献率得到的本征向量进行再次筛选,保留其中对于易于区分的、有代表性的特征信息描述能力较好的本征向量。实验结果表明,新提出的方法不仅更大程度地提高了数据降维的效率,而且还能得到更好的表情识别结果。

将主成分选择方法与基于待识别目标的先验知识相结合的思路还可以在其他领域应用,这对于主成分分析技术的进一步推广有一定的积极作用。

## 结论与展望

### 结论

让计算机能自动分析人的内心状态变化,更智能、更人性化地为人类服务是目前计算机研究工作的一个重点内容。心理学研究表明,人的内心状态有多种表现途径,包括人脸表情、语调、体态、脉搏、体温等等。通过对外在表现特征的分析,并结合对人的内心状态的先验认识,计算机能较好地实现对人的意图的估计。在所有的表现途径中,人脸表情具有最容易控制、特征明显、提取方便等优势,人脸表情识别技术也因而成为了一个研究的热点。最初对人脸表情进行识别研究是针对静态图片来进行的,随着对表情认识的深入,基于静态图片的表情识别方式因为识别的表情类型单一、数据源获取过程过于人为化、识别结果与真实情况误差较大等原因,已逐渐被基于视频序列的表情识别方法所取代。近几年,围绕基于视频序列的动态表情识别方法,研究者们做了大量的工作。

本文在紧密跟踪目前国内外主要研究机构的工作基础之上,对与动态表情识别技术关系最为密切的心理学和计算机科学领域的成果做了详细地总结和分析。从整体来看,动态人脸表情识别技术研究还处于一个起步的阶段,由于人脸表情变化的复杂性,心理学研究对于表情的定义、表情量度的定义、表情与人的内心情感之间的关系等的认识都还无法达成统一。计算机方面的工作主要是针对几种基本表情类型进行识别,对系统运行环境的要求比较高。可以说动态表情识别技术距离实际应用还有很长一段路,需要研究者们进一步努力,也需要各个学科知识的互相促进。

为了对表情识别各个主要技术点有深入地认识以及为了后续研究工作的需要。本文首先建立了一个包含 840 段视频序列的基本人脸表情数据库,然后结合目前比较成熟的表情区域定位方法、特征提取方法以及信息分类方法构建了一个动态人脸表情识别系统实例。基本人脸表情库中的表情序列长度为 3 秒,表情变化过程严格按照“准备→启动→高潮→放松→结束”的模式进行,对各种表情类型的定义参考了 FACS 系统规则,背景单一。基本人脸表情数据库适合于进行比较基础的基本表情识别实验。识别系统实例满足两个事先针对表情动态性建立的接口要求,通过对表情库中伤心、吃惊、愤怒以及中性表情进行测试,得到综合识别率为 91.7%。由于基于的数据库并不一致,此时得到的识别率和其它研究机构的结果不具备可比性,但是可以证明系统运行正常,这为以后新的表情识别系统的开发和改善提供了良好的实验平台。

因为实验的数据源是视频序列,所以待处理的数据量非常庞大,而且表情特征信息比较容易受到干扰。图像归一化处理技术能根据定位和跟踪的结果对图像进行一定的调整,将人脸面部图像变换到固定的位置、统一的朝向以及相同的大小,从而保证了表情信息的有效提取。主成分分析技术在表情识别中的应用是利用其多元正交线性变换特性对原始表情信息进行重新描述,并结合基于贡献率的方法对线性变换基底进行筛选,从而使得新的表情信息区分度更好,维数更少。但是,在实验中两种方法的效果并不是很理想。针对归一化处理技术,本文结合光流特征提取方法定量地分析了归一化误差对表情特征提取的影响,并提出了加权优化匹配算法进行误差的矫正,这样的处理使得表情系统的抗外界干扰能力得到进一步增强,当对象的头部运动比较剧烈的时候,表情信息也能有效地被提取。针对表情信息的合理描述问题,本文结合表情变化本身的特性,利用面部运动单元与本征向量的对应关系对基于贡献率得到的本征向量进行再次筛选,实验结果证明了改进后的方法在保证识别率不降低的情况下进一步提高了数据压缩比例,这为系统的实时性要求提供了更为可靠的保障,在将来的实际产品开发中,也将极大地降低系统对硬件设备的要求。

## 展望

本次研究主要是对基本表情进行分析和识别,下一步的工作主要在两个方面展开。

其一,对表情识别系统中各个阶段的方法进行改良。这其中包括作为关键步骤的人脸/面部区域定位方法、特征提取方法、表情信息分类方法,还包括表情数据源采集方法、表情图像预处理方法、特征数据压缩和重新描述方法等辅助方法。改良的途径分为两类,一类是关注其它领域譬如人脸识别、视频数据压缩的最新进展,利用其它领域的新思路、新方法来改善表情识别系统,另一类就是积极分析表情识别系统各个部分之间的关联性,对各部分方法进行合理的筛选和融合,以达到整体识别效率最高的结果。

其二,积极追求在自然表情识别研究、表情程度分析研究等方面的突破。人的内心状态的变化是无常的,并且可能是细微的。仅仅对基本表情或者单一程度的表情进行分析是远远不够的。目前已有相关的研究报道,但是由于技术的限制以及心理学依据的缺乏,自然表情识别以及表情程度分析都只是在基本表情识别的基础上进行简单的线性扩展,没有本质的飞跃。这需要我们的进一步努力。

另外,积极跟踪心理学研究成果,并利用其指导表情识别系统的开发也是下一步工作中要继续关注的重要内容。

## 参考文献

- [1] Maja Pantic, Rothkpantz, L.J.M. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 2003, 91(9): 1370-1390
- [2] R. W. Picard. *Affective Computing*. Cambridge, MA: MIT Press, 1997
- [3] M.Suwa, N.Sugie, K.Fujimora. A Preliminary Note on Pattern Recognition of Human Emotional Expression. In: *Proceedings of the 4th International Joint Conference on Pattern Recognition*. Kyoto, Japan: 1978, 408-410
- [4] Terzopoulos D, Waters K. Analysis and Synthesis of Facial Images Using Physical and Anatomical Models. In: *Proceedings of the International Conference on Computer Vision*. Vancouver, Canada: 1990, 727-732
- [5] Mase K. Recognition of facial expressions from optical flow. *IEICE Transactions, Special Issue on Computer Vision and its Applications*, 1991, E74(10): 3474-3483
- [6] P.Ekman, W.Friesen. *Pictures of Facial Affect*. Palo Alto, CA, USA: Consulting Psychologists Press, 1976
- [7] 陶建华. 第一届中国情感计算及智能交互学术会议. [http://www.cad.zju.edu.cn/chinagraph/chinese/academically/Academic\\_ACH03.htm](http://www.cad.zju.edu.cn/chinagraph/chinese/academically/Academic_ACH03.htm), Dec 8-9, 2003
- [8] Jianhua Tao. The First International Conference on Affective Computing & Intelligent Interaction. <http://www.affectivecomputing.org/2005/index.htm>, Oct 22-24, 2005
- [9] 杨磊. 首届全国无处不在的计算与第二届全国可穿戴式计算学术会议. [http://www.chinagraph.org/chinese/academically/Academic\\_Anywhere.htm](http://www.chinagraph.org/chinese/academically/Academic_Anywhere.htm), Sep 25-26, 2003
- [10] Dan Ferber. The Man Who Mistook His Girlfriend for a Robot. <http://www.popsoci.com/popsoci/science/article/0,12543,473054-1,00.html>, Aug 11, 2003

- [11] Y.Dai, Y.Shibata, K.Hashimoto et al. A New Method of Facial Action Extraction and Expression Recognition of Person without Language. In: Seventh International Conference on Parallel and Distributed Systems: Workshops. Iwate, Japan: 2000, 129-136
- [12] Haisong Gu, Qiang Ji. An Automated Face Reader for Fatigue Detection. In: The 6th International Conference on Automatic Face and Gesture Recognition. Seoul, Korea: 2004, 111-116
- [13] P.Ekman, W.V.Friesen. Constants Across Cultures in the Face and Emotion. Journal of Personality and Social Psychology, 1971, 17(2): 124-129
- [14] C.Darwin. The Expression of the Emotions in Man and Animals. J.Murray, London: Gavin DeBeer Collection QP401 DAR, 1872
- [15] P.Ekman, E.Rosenberg, J.C.Hager. Facial Action Coding System Affect Interpretation Database (FACSAID). <http://nirc.com/Expression/FACSAID/facsaid.html>, Jul 1998
- [16] I.Essa, A.Pentland. Coding, Analysis, Interpretation and Recognition of Facial Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 757-763
- [17] Hiroyuki Konishi, Mitsuteru Kokubun, Kirokazu Iguchi. System for analysis of involuntary facial expressions. In: IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition. Nara, Japan: IEEE Computer Society, 1998, 454-459
- [18] M. J. Black, Y. Yacoob. Recognizing Facial Expressions in Image Sequences using Local Parameterized Models of Image Motion. International Journal of Computer Vision, 1997, 25(1): 23-48
- [19] P.Ekman, W.V.Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, Calif: Consulting Psychologists Press, 1978
- [20] P.Ekman. An argument for basic emotions. Cognition and Emotion, 1992, 6(3):169-200

- [21] 高文,金辉. 面部表情的分析与识别. 计算机学报, 1997, 20(9): 782-789
- [22] Beat Fasel, Juergen Luetttin. Automatic Facial Expression Analysis: A Survey. Pattern Recognition, 2003, 36(1): 259-275
- [23] 林维训, 潘纲, 吴朝晖等. 脸部特征定位方法. 中国图象图形学报, 2003, 8(8): 849-859
- [24] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 1989, 77(2): 257-286
- [25] H.Kobayashi, F.Hara. Dynamic recognition of basic facial expressions by discrete-time recurrent neural network. In: Proceedings of the International Joint Conference on Neural Network. NAGOYA, JAPAN: 1993, 155-158
- [26] M.J.Lyons, J.Budynek, S.Akamatsu. Automatic Classification of Single Facial Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(12): 1357-1362
- [27] G.L.Donato, M.S.Bartlett, J.C.Hager et al. Classifying Facial Actions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(10): 974-989.
- [28] A.Lanitis, C.J.Taylor, T.F.Cootes. Automatic Interpretation and Coding of Face Images using Flexible Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 743-756
- [29] J.Cohn, G.S.Katz. Bimodal Expression of Emotion by Face and Voice. In: Workshop on Face/Gesture Recognition and Their Applications, the Sixth ACM International Multimedia Conference. Bristol, UK: 1998, 41-44
- [30] Sim T, Baker S, Bsat M. The CMU Pose, Illumination and Expression(PIE) database. In: Proceedings of Automatic Face and Gesture Recognition. Washington, DC, USA: IEEE Computer Society, 2002, 53-58
- [31] Kanade.T, Cohn.J, Tian.Y. Comprehensive database for facial expression analysis. In: Proceedings of International Conference on Face and Gesture Recognition. Grenoble, France: IEEE Computer Society, 2000, 46-53
- [32] Turk M, Pentland A. Eigen faces for recognition. Journal of Cognitive



- Neuroscience, 1991, 3(1): 71-86
- [33] Graham D B, Allinson N M. Characterizing virtual eigensignatures for general purpose face recognition. *Face Recognition: From Theory to Applications*, 1998, 163: 446-456
- [34] Loui A C, Judice C N, Liu S. An image database for benchmarking of automatic face detection and recognition algorithms. In: *Proceedings of International Conference of Image Processing*, Chicago, Illinois, USA: IEEE Computer Society, 1998, 146-150
- [35] 吴丹, 林学闾. 人脸表情视频数据库的设计与实现. *计算机工程与应用*, 2004, 40(5): 177-180
- [36] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997, 55(1): 119-139
- [37] Horn B, Schunck B. Determining optical flow. *Artificial Intelligence*, 1981, 17:185-203
- [38] Kevin Murphy. Hidden Markov Model (HMM) Toolbox for Matlab. <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html#hmm>, May 4, 2003
- [39] Intel Corporation. Open Source Computer Vision Library. <http://www.intel.com/research/mrl/research/opencv/>, Jun 14, 2000
- [40] Yongmian Zhang, Qiang Ji. Facial Expression Understanding in Image Sequences Using Dynamic and Active Visual Information Fusion. In: *Ninth IEEE International Conference on Computer Vision*. Nice, France: IEEE, 2003, 1297-1304
- [41] 赵慧琳, 王林泉, 葛元. 人脸图像定位与标准化算法. *计算机工程*, 2003, 29(22): 158-160
- [42] 朱长仁, 王润生. 基于单视图的多姿态人脸识别方法. *计算机学报*, 2003, 26(1): 104-109
- [43] Xiujuan Chai, Shiguang Shan, Wen Gao et al. Pose Normalization Using Generic 3D Face Model as a Prior for Pose-Insensitive Face Recognition. In: *SINO BIO-*

- METRICS'04. Guangzhou, China: Springer-Verlag in Lecture Notes in Computer Science, 2004, 144-152
- [44] T.F.Cootes, C.J.Taylor, D.Cooper et al. Active Shape Models-Their Training and Application. *Computer vision and image understanding*, 1995, 61(1): 38-59.
- [45] T F Cootes, G J Edwards, C J Taylor. Active appearance models. *IEEE Trans*, 2001, PAMI-23(6): 681-685
- [46] D.Beymer. Vectorizing Face Images by Interleaving Shape and Texture Computation. AI Memo 1537, Artificial Intelligence Laboratory, MIT, 1995
- [47] I Matthews, S Baker. Active appearance models revisited. *International Journal of Computer Vision*, 2004, 60(2): 135-164
- [48] Kirby M, Sirovich L. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(1): 103-108
- [49] Ashish Kapoor, Yuan Qi, Rosalind W.Picard. Fully Automatic Upper Facial Action Recognition. In: *Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*. Nice, France: IEEE Computer Society, 2003, 195-202
- [50] JJ Lien, T.Kanade, J.Cohn et al. Subtly Different Facial Expression Recognition and Expression Intensity Estimation. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Santa Barbara, CA, USA : IEEE Computer Society, 1998, 853-859
- [51] Y.L.Tian, T.Kanade, J.Cohn. Recognizing action units for facial expression analysis. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 2001, 23(2):1-19
- [52] L.Smith. A tutorial on Principal Components Analysis. <http://kybele.psych.cornell.edu/%7Eedelman/Psych-465-Spring-2003/PCA-tutorial.pdf>, Feb 26, 2002

## 致 谢

在本人读研期间，一直得到了导师邹北骢教授的悉心指导。从论文的选题、研究、审查直至完成，都是在邹老师的精心指导下完成的，在此谨向邹老师表示崇高的敬意和最衷心的感谢。邹老师高深的学术造诣、严谨的治学态度、诲人不倦的敬业精神以及高度的责任感使我受益匪浅，一直是我前进的导航灯。两年多来，邹老师和师母刘波老师在学习和生活上都给予了我无微不至的关怀和帮助，至此学位论文完成之际，衷心感谢他们对我悉心地关怀和谆谆地教诲。

同时感谢李仁发教授、孙家广教授、朱理副教授的悉心指导和帮助。感谢潘华伟、王磊、黄昆博士及我的师兄龚理、张保国、肖秀春，同学朱岳、刘洁、周洁琼、黄丽达、周浩宇、吕格峰、张华、郭勇、杨磊、王梓展、张锦、张白妮、王艳琴、卢新国、李闻、李文嘉，感谢他们在学习和生活上给予我支持和帮助。还感谢我的师弟朱江、叶宁宁、陈宇、辛国江、傅红普、王涛和师妹汤丽君、李燕、陆欣星、张玲、康晶、冯莲、刘丽丽等给予我的建议和帮助，也非常感谢他们乐意作为本文人脸表情库采集的志愿者。

感谢我亲爱的父母，感谢他们对我的养育、关心和支持，使我能够有信心和勇气面对困难，迎接挑战，顺利完成硕士学业。感谢我的女朋友周凌，是她在我情绪低落的时候给予了我最有力的鼓励和支持。

感谢国防科技大学《计算机工程与科学》杂志社编委对本文第四章内容的悉心指导和建议。

感谢各位专家百忙之中对本文的审阅和提出的宝贵意见！

2005 年 5 月

## 附录 A 攻读学位期间完成的论文和参加的项目

攻读学位期间完成的论文:

- [1] 应伟, 邹北骢, 周凌, 孙家广. 表情特征区域归一化误差分析及矫正. 计算机工程与科学. (已录用)
- [2] 周凌, 邹北骢, 应伟, 孙家广. 采用多尺度掩模消除干扰的反向组合算法. 计算机工程与科学. (已录用)
- [3] 应伟, 邹北骢, 周凌, 孙家广. 用与研究. (已投稿)

攻读学位期间参加的项目:

- [1] 智能小区车牌自动识别与管理系统, 省第五届“挑战杯”长沙电信大学生课外学术科技作品, 2002
- [2] 基于视频的人脸表情动画技术研究, 湖南省自然科学基金项目(编号为 02JJY3049), 2002

## 附录 B FACS 系统基本形变单元及肌肉图示

## B.1 基本形变单元描述

表 B1 单个形变单元 (Single Action Units)

AU Number	FACS Name	Muscular Basis
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae, Depressor Supercilli, Corrugator
5	Upper Lid Raiser	Levator Palpebrae Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palebralis
8	Lips Toward Each Other	Orbicularis Oris
9	Nose Wrinkler	Levator Labii Superioris, Alaeque Nasi
10	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
11	Nasolabial Furrow Deepener	Zygomatic Minor
12	Lip Corner Puller	Zygomatic Major
13	Cheek Puffer	Caninus
14	Dimpler	Buccinator
15	Lip Corner Depressor	Triangularis
16	Lower Lip Depressor	Depressor Labii
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris; Incisivii Labii Inferioris
20	Lip Stretcher	Risorius
22	Lip Funneler	Orbicularis Oris
23	Lip Tightner	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris

25	Lips Part	Depressor Labii, or Relaxation of Mentalis or Orbicularis Oris
26	Jaw Drop	Maseter, Temporal and Internal Pterygoid Relaxed
27	Mouth Stretch	Pterygoids, Digastric
28	Lip Suck	Orbicularis Oris
38	Nostril Dilator	Nasalis, Pars Alaris
39	Nostril Compressor	Nasalis, Pars Transversa and Depressor Septi Nasi
41	Lid Droop	Relaxation of Levator Palpebrae Superioris
42	Slit	Orbicularis Oculi
43	Eyes Closed	Relaxation of Levator Palpebrae Superioris
44	Squint	Orbicularis Oculi, Pars Palpebralis
45	Blink	Relaxation of Levator Palpebrae and Contraction of Orbicularis Oculi, Pars Palpebralis
46	Wink	Orbicularis Oculi

表 B2 联合形变单元 (More Grossly Defined AUs)

AU Number	FACS Name	AU Number	FACS Name
19	Tongue Out	33	Cheek Blow
21	Neck Tightener	34	Cheek Puff
29	Jaw Thrust	35	Cheek Suck
30	Jaw Sideways	36	Tongue Bulge
31	Jaw Clencher	37	Lip Wipe
32	Lip Bite		

表 B3 头部及眼睛方位变换描述单元

AU Number	FACS Name	AU Number	FACS Name
51	Head turn left	57	Head forward
52	Head turn right	58	Head back
53	Head up	61	Eyes turn left
54	Head down	62	Eyes turn right
55	Head tilt left	63	Eyes up
56	Head tilt right	64	Eyes down

## B.2 人脸面部肌肉图示

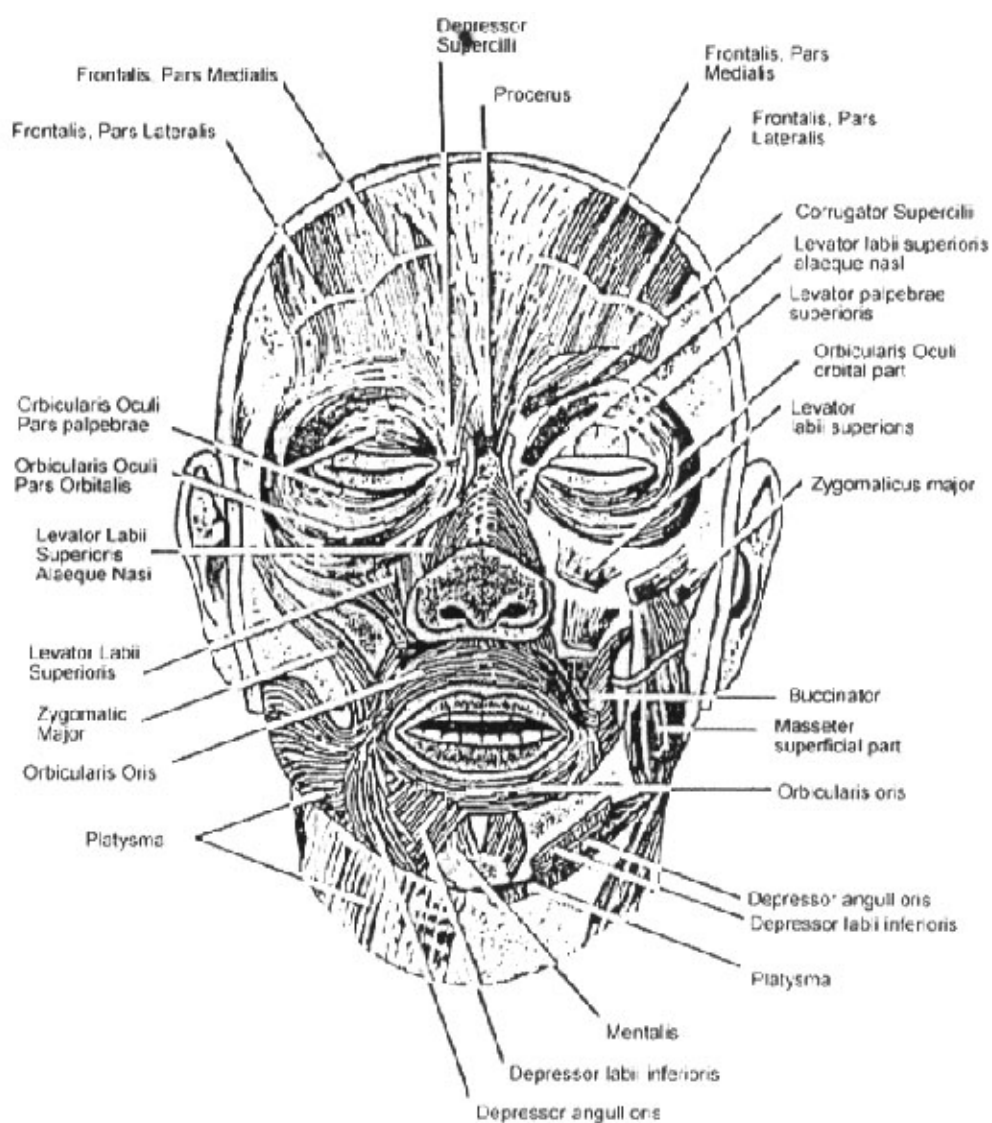


图 B1 与基本形变单元相关的面部肌肉图



作者: [应伟](#)  
学位授予单位: [湖南大学](#)  
被引用次数: 2次

参考文献(47条)

1. [Maja Pantic, Rothkpantz L J M](#) [Toward an affect-sensitive multimodal human-computer interaction](#) 2003 (09)
2. [R W Picard](#) [Affective Computing](#) 1997
3. [M Suwa, N Sugie, K Fujimora](#) [A Preliminary Note on Pattern Recognition of Human Emotional Expression](#) 1978
4. [Terzopoulos D, Waters K](#) [Analysis and Synthesis of Facial Images Using Physical and Anatomical Models](#) 1990
5. [陶建华](#) [查看详情\[会议论文\]](#) 2003
6. [Jianhua Tao](#) [The First International Conference on Affective Computing & Intelligent Interaction](#) 2005
7. [杨磊](#) [首届全国无处不在的计算与第二届全国可穿戴式计算学术会议](#) 2003
8. [Dan Ferber](#) [The Man Who Mistook His Girlfriend for a Robot](#) 2003
9. [Y Dai, Y Shibata, K Hashimoto](#) [A New Method of Facial Action Extraction and Expression Recognition of Person without Language](#) 2000
10. [Haisong Gu, Qiang Ji](#) [An Automated Face Reader for Fatigue Detection](#) 2004
11. [P Ekman, W V Friesen](#) [Constants Across Cultures in the Face and Emotion](#) 1971 (02)
12. [C Darwin, J Murray](#) [The Expression of the Emotions in Man and Animals](#) 1872
13. [P Ekman, E Rosenberg, J C Hager](#) [Facial Action Coding System Affect Interpretation Database \(FACSAID\)](#) 1998
14. [I Essa, A Pentland](#) [Coding, Analysis, Interpretation and Recognition of Facial Expressions](#) 1997 (07)
15. [Hiroyuki Konishi, Mitsuteru Kokubun, Kirokazu Iguchi](#) [System for analysis of involuntary facial expressions](#) 1998
16. [P Ekman, W V Friesen](#) [Facial Action Coding System: A Technique for the Measurement of Facial Movement](#) 1978
17. [P Ekman](#) [An argument for basic emotions](#) 1992 (03)
18. [高文, 金辉](#) [面部表情图像的分析与识别\[期刊论文\]-计算机学报](#) 1997 (9)
19. [Beat Fasel, Juergen Luetttin](#) [Automatic Facial Expression Analysis: A Survey](#) 2003 (01)
20. [林维训, 潘纲, 吴朝晖, 潘云鹤](#) [脸部特征定位方法\[期刊论文\]-中国图象图形学报A辑](#) 2003 (8)
21. [Rabiner LR](#) [A tutorial on hidden Markov models and selected applications in speech recognition\[外文期刊\]](#) 1989 (02)
22. [H Kobayashi, F Hara](#) [Dynamic recognition of basic facial expressions by discrete-time recurrent neural network](#) 1993
23. [M J Lyons, J Budynek, S Akamatsu](#) [Automatic Classification of Single Facial Images\[外文期刊\]](#)

1999(12)

24. [G L Donato, M S Bartlett, J C Hager](#) [Classifying Facial Actions](#)[外文期刊] 1999(10)
25. [A Lanitis, C J Taylor, T F Cootes](#) [Automatic Interpretation and Coding of Face Images using Flexible Models](#)[外文期刊] 1997(07)
26. [J Cohn, G S Katz](#) [Bimodal Expression of Emotion by Face and Voice](#) 1998
27. [Sim T, Baker S, Sato M](#) [The CMU Pose, Illumination and Expression \(PIE\) database](#) 2002
28. [Kanade T, Cohn J, Tian Y](#) [Comprehensive database for facial expression analysis](#) 2000
29. [Turk M, Pentland A](#) [Eigen faces for recognition](#) 1991(01)
30. [Graham D B, Allinson N M](#) [Characterizing virtual eigensignatures for general purpose face recognition, Face Recognition: From Theory to Applications](#) 1998
31. [Loui A C, Judice C N, Liu S](#) [An image database for benchmarking of automatic face detection and recognition algorithms](#) 1998
32. 吴丹, 林学闯 [人脸表情视频数据库的设计与实现](#)[期刊论文]-[计算机工程与应用](#) 2004(5)
33. [Freund Y, Schapire R E](#) [A decision-theoretic generalization of on-line learning and an application to boosting](#)[外文期刊] 1997(01)
34. [Horn B, Schunck B](#) [Determining optical flow](#) 1981
35. [Kevin Murphy](#) [Hidden Markov Model \(HMM\) Toolbox for Matlab](#) 2003
36. [Intel Corporation](#) [Open Source Computer Vision Library](#) 2000
37. [Yongmian Zhang, Qiang Ji](#) [Facial Expression Understanding in Image Sequences Using Dynamic and Active Visual Information Fusion](#) 2003
38. 赵慧琳, 王林泉, 葛元 [人脸图像定位与标准化算法](#)[期刊论文]-[计算机工程](#) 2003(22)
39. 朱长仁, 王润生 [基于单视图的多姿态人脸识别算法](#)[期刊论文]-[计算机学报](#) 2003(1)
40. [T F Cootes, C J Taylor, D Cooper](#) [Active Shape Models—Their Training and Application](#) 1995(01)
41. [T F Cootes, G J Edwards, C J Taylor](#) [Active appearance models](#) 2001(06)
42. [D Beymer](#) [Vectorizing Face Images by Interleaving Shape and Texture Computation](#) 1995
43. [I Matthews, S Baker](#) [Active appearance models revisited](#)[外文期刊] 2004(02)
44. [Kirby M, Sirovich L](#) [Application of the Karhunen-Loève procedure for the characterization of human faces](#) 1990(01)
45. [Ashish Kapoor, Yuan Qi, Rosalind W Picard](#) [Fully Automatic Upper Facial Action Recognition](#) 2003
46. [JJ Lien, T Kanade, J Cohn](#) [Subtly Different Facial Expression Recognition and Expression Intensity Estimation](#) 1998
47. [L Smith](#) [A tutorial on Principal Components Analysis](#) 2002

#### 引证文献(2条)

1. 汤丽君 [人脸面部表情识别方法研究](#)[学位论文]硕士 2006
2. 李燕 [基于视频的面部表情强度度量方法研究](#)[学位论文]硕士 2006

本文链接: [http://d.g.wanfangdata.com.cn/Thesis\\_Y832921.aspx](http://d.g.wanfangdata.com.cn/Thesis_Y832921.aspx)