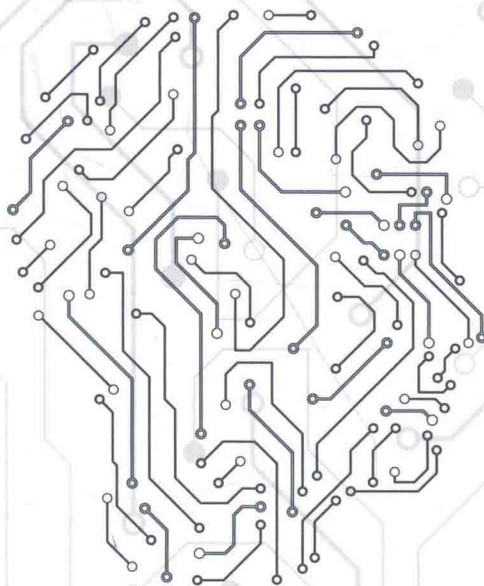


MACHINE INTELLIGENCE FACE ENGINEERING

机器智能 人脸工程

肖若秀 王志良 / 编著



人脸工程学包括人脸识别、表情识别和人脸合成三个部分
本书分别介绍了学科基础、面部运动测量
图像处理、检测跟踪、特征提取、面部特征模式识别
人脸合成、人脸识别、表情识别等技术



机械工业出版社
CHINA MACHINE PRESS

智能制造与装备制造业转型升级丛书

机器智能：人脸工程

肖若秀 王志良 编著



机械工业出版社

人脸工程学的研究内容主要包括人脸识别、表情识别和人脸合成三个部分。本书分别介绍了人脸识别、表情识别和人脸合成研究中用到的相关理论和算法,最后在总结作者所在课题组研究成果的基础上,给出了人脸识别、表情识别和人脸合成系统的设计与实现流程实例。

本书适宜从事计算机、自动化、模式识别、智能科学、人机交互技术的科技人员阅读,也可以作为高等院校相关专业的学生、研究生的教学参考书。

图书在版编目(CIP)数据

机器智能. 人脸工程/肖若秀,王志良编著. —2版.
—北京:机械工业出版社,2017.8
(智能制造与装备制造业转型升级丛书)
ISBN 978-7-111-57609-9

I. ①机… II. ①肖… ②王… III. ①人工智能
IV. ①TP18

中国版本图书馆CIP数据核字(2017)第185728号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:罗莉 责任编辑:罗莉

责任校对:李锦莉 任秀丽

责任印制:孙炜

北京京丰印刷厂印刷

2017年9月第2版·第1次印刷

184mm×260mm·19.5印张·479千字

0 001—3 000册

标准书号:ISBN 978-7-111-57609-9

定价:89.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线:010-88361066 机工官网:www.cmpbook.com

读者购书热线:010-68326294 机工官博:weibo.com/cmp1952

010-88379203 金书网:www.golden-book.com

封面无防伪标均为盗版

教育服务网:www.cmpedu.com

前 言

人脸研究是跨越人文科学与自然科学的新兴交叉研究领域，日益得到人们的重视。随着其研究成果在和谐人机交互、机器人、虚拟现实、身份识别等领域得到应用，希望了解学习人脸研究相关知识的人也越来越多。针对这种需求，作者在多年研究的基础上，整理编写了本书，目的是为对人脸研究感兴趣的学生和研究人员提供一本较全面的入门参考书籍。

人脸工程学是指以人类学、心理学、脑科学、人文科学、认知科学、信息科学、人工智能等学科为理论基础，利用工程的方法和技术（尤其是信息技术）对人脸（包括识别、建模与重构）进行研究的学问。本书详细介绍了人脸工程研究的相关理论和算法，并总结作者所在课题组的研究成果，给出了人脸识别、表情识别和人脸合成等实例系统。

全书共有 11 章。第 1 章主要介绍人脸工程学的研究内容、发展情况及应用领域；第 2 章论述了人类学、文学艺术、心理学与脑科学、动画领域关于人脸的研究；第 3 章给出了几种常见的面部动作编码系统和面部表情测量系统；第 4 章阐述了图像处理技术；第 5 章给出了几种常用的人脸跟踪检测技术；第 6 章主要介绍几种常用的特征提取算法；第 7 章叙述了几种常用的模式识别算法；第 8 章主要介绍常用的人脸合成技术；第 9 章给出了人脸识别系统的设计范例；第 10 章介绍了表情识别系统；第 11 章阐述了人脸合成系统。

本书的出版得到了机械工业出版社的大力支持，在此表示诚挚的感谢。同时感谢国家重点研发计划重点专项（2016YFB1001404）、国家自然科学基金重点项目（6143204）、国防科技创新特区项目（1716312ZT00201401）以及北京科技大学中央高校基本科研业务项目（FRF—TP—16—045A1）的资助。

由于作者的水平有限，书中肯定有不少的缺点和疏漏之处，敬请读者批评指正。

目 录

前言

第1章 绪论..... 1

1.1 人脸工程学..... 1

1.1.1 人脸工程学的研究内容..... 1

1.1.2 人脸工程学研究的意义与应用..... 2

1.1.3 人脸工程学研究在实用化过程中 的挑战..... 5

1.2 人脸工程学研究历程..... 7

1.2.1 人脸识别..... 7

1.2.2 表情识别..... 8

1.2.3 人脸合成..... 10

1.2.4 相关学术资源..... 13

1.3 人脸工程学的未来研究方向..... 16

1.4 本书的内容..... 17

参考文献..... 17

第2章 人脸工程学研究学科基础..... 20

2.1 人类学对人脸的研究..... 20

2.1.1 达尔文之前的研究..... 20

2.1.2 达尔文对表情的研究..... 21

2.2 文学艺术中对人脸的研究..... 23

2.2.1 人脸的美学研究..... 23

2.2.2 人脸表情在戏曲和舞台剧中的 表现..... 27

2.2.3 人脸在各种美术画法中的表现..... 30

2.3 动画中人脸的表现..... 31

2.3.1 人脸动画的应用..... 31

2.3.2 卡通动画中人脸的表现形式及 常用软件..... 32

2.3.3 二维动画中人脸的表现..... 32

2.4 情绪心理学关于表情的研究..... 34

2.5 认知心理学关于人脸的研究..... 36

2.6 脑科学关于人脸的研究..... 40

2.6.1 人脸识别的 ERP 研究..... 40

2.6.2 表情识别的 ERP 研究..... 43

参考文献..... 45

第3章 面部运动的测量技术..... 47

3.1 概述..... 47

3.2 面部动作编码系统..... 48

3.2.1 概述..... 48

3.2.2 FACS 的特点与应用..... 52

3.2.3 FACS 的扩展与改进..... 53

3.3 最大限度辨别面部肌肉运动编码 系统..... 54

3.3.1 伊扎德与 MAX..... 54

3.3.2 MAX 的主要内容..... 55

3.3.3 MAX 与 FACS 的比较..... 56

3.4 其他面部表情测量系统..... 57

3.4.1 表情识别整体判断系统..... 57

3.4.2 自我评估情绪编码系统..... 57

3.4.3 面部表情分析工具..... 57

参考文献..... 57

第4章 图像处理技术..... 59

4.1 图像处理的基本概念..... 59

4.1.1 数字图像的概念..... 59

4.1.2 数字图像处理的概念..... 59

4.2 图像处理的基本操作..... 60

4.2.1 图像的平移、旋转、放缩、镜像 变换、转置..... 60

4.2.2 图像的平滑、锐化..... 64

4.2.3 图像的腐蚀、膨胀和细化..... 69

4.2.4 图像的恢复与重建..... 73

4.3 图像处理的高级操作..... 74

4.3.1 图像的边缘检测..... 74

4.3.2 图像的 Hough 变换..... 78

4.3.3 轮廓的提取与跟踪..... 79

参考文献..... 81

第5章 人脸检测跟踪技术..... 82

5.1 人脸检测..... 82

5.1.1 人脸检测方法的分类..... 82

5.1.2 基于肤色的人脸检测..... 82

5.1.3 基于形状的人脸检测..... 86

5.1.4 基于特征的人脸检测..... 88

5.2 人脸跟踪..... 91

5.2.1 帧差法..... 91

5.2.2 基于运动目标预测的人脸跟踪	92	7.4.6 小结	152
5.2.3 基于模型的人脸跟踪	92	7.5 神经网络	152
5.2.4 基于人脸局部特征的人脸跟踪	95	7.5.1 概述	153
参考文献	96	7.5.2 神经网络集成	160
第6章 面部特征提取的算法	98	7.5.3 小结	163
6.1 概述	98	7.6 模糊模式识别	163
6.2 几何特征的提取	99	7.6.1 概述	163
6.3 统计特征的提取	101	7.6.2 模糊数学基本理论	164
6.3.1 主成分分析算法	101	7.6.3 模糊模式识别	166
6.3.2 二维主成分分析算法	104	7.6.4 小结	169
6.3.3 线性判别分析算法	105	参考文献	169
6.3.4 独立成分分析算法	106	第8章 人脸合成的方法与技术	173
6.4 频率域特征的提取	107	8.1 概述	173
6.4.1 小波技术	107	8.2 人脸合成技术的分类	174
6.4.2 Gabor 小波	108	8.3 人脸几何建模	177
6.4.3 离散余弦变换	109	8.3.1 人脸模型的表达形式	177
6.5 运动特征的提取	109	8.3.2 一般人脸模型	178
6.6 代数特征的提取	111	8.3.3 特定人脸模型	181
参考文献	112	8.4 纹理映射	183
第7章 面部特征的模式识别算法	115	8.5 人脸动画	185
7.1 线性判别分析	115	8.5.1 人脸动画技术	185
7.1.1 线性判别函数的基本概念	115	8.5.2 人脸动画驱动技术	189
7.1.2 Fisher 线性判别	116	8.6 MPEG-4 人脸动画原理	190
7.1.3 小结	119	8.6.1 FDP、FAP 与 FAPU 的定义	190
7.2 支持向量机	120	8.6.2 FAP 驱动人脸动画的基本原理	191
7.2.1 支持向量机基本原理	121	参考文献	193
7.2.2 SVM 分类器的设计	125	第9章 人脸识别系统	196
7.2.3 小结	125	9.1 概述	196
7.3 贝叶斯网络	127	9.2 人脸识别关键问题的研究	197
7.3.1 概述	127	9.3 人脸识别流程	199
7.3.2 贝叶斯网络概率基础	128	9.4 人脸识别系统的设计与实现	200
7.3.3 贝叶斯网络的构建	128	9.4.1 人脸识别系统的总体设计	201
7.3.4 贝叶斯网络推理算法	132	9.4.2 人脸识别系统的算法设计	202
7.3.5 贝叶斯网络分类器	134	9.4.3 人脸识别系统的实现	207
7.3.6 小结	138	参考文献	212
7.4 隐马尔可夫模型及其基本问题	138	第10章 面部表情识别系统	215
7.4.1 概述	138	10.1 概述	215
7.4.2 马尔可夫链模型	139	10.2 基于静态图像的面部表情识别系统	218
7.4.3 隐马尔可夫模型	141	10.2.1 系统的总体设计	218
7.4.4 隐马尔可夫模型的三个基本问题	143	10.2.2 系统的算法设计	219
7.4.5 隐马尔可夫算法实现中的基本问题	149		

10.2.3 系统的实现	224	的变换	273
10.3 基于主动表观模型的实时面部表情 识别系统	225	11.1.3 纹理贴图	275
10.3.1 系统设计	225	11.1.4 面部表情的运动计算及表情 合成	276
10.3.2 基于肤色模型的人脸检测	226	11.1.5 实时表情信息获取与表情重构 的实现	278
10.3.3 人脸图像预处理	226	11.1.6 小结	283
10.3.4 特征点定位及特征提取	227	11.2 三维虚拟人脸模型	283
10.3.5 表情识别	230	11.2.1 三维人脸模型的建立	283
10.4 基于动态图像序列的面部表情 识别	231	11.2.2 特征点的选取	285
10.4.1 光流的基本计算方法	232	11.2.3 纹理映射	286
10.4.2 基于 Hessian 矩阵的改进光流 算法	235	11.2.4 特定虚拟人脸模型的实现	288
10.4.3 散度-旋度样条约束下的非刚 体光流算法	239	11.2.5 小结	289
10.4.4 基于改进 MMI 的 HMM 算法的 面部表情识别	244	11.3 虚拟人脸的年龄仿真	289
10.4.5 基于改进 MMD 的 HMM 算法的 面部表情识别	257	11.3.1 年龄老化特征的相关研究	290
参考文献	268	11.3.2 Dirichlet 自由变形算法	291
第 11 章 人脸合成实例	271	11.3.3 Dirichlet 自由变形算法在三维 空间中的应用	296
11.1 基于视频的人脸表情合成	271	11.3.4 应用 Dirichlet 自由变形算法 生成特定人脸模型	296
11.1.1 Candide 模型	271	11.3.5 虚拟人脸部年龄仿真的实现	298
11.1.2 标准人脸模型到特定人脸模型 的变换	273	参考文献	300
11.1.3 纹理贴图	275	附录 缩略语	302
11.1.4 面部表情的运动计算及表情 合成	276		
11.1.5 实时表情信息获取与表情重构 的实现	278		
11.1.6 小结	283		
11.2 三维虚拟人脸模型	283		
11.2.1 三维人脸模型的建立	283		
11.2.2 特征点的选取	285		
11.2.3 纹理映射	286		
11.2.4 特定虚拟人脸模型的实现	288		
11.2.5 小结	289		
11.3 虚拟人脸的年龄仿真	289		
11.3.1 年龄老化特征的相关研究	290		
11.3.2 Dirichlet 自由变形算法	291		
11.3.3 Dirichlet 自由变形算法在三维 空间中的应用	296		
11.3.4 应用 Dirichlet 自由变形算法 生成特定人脸模型	296		
11.3.5 虚拟人脸部年龄仿真的实现	298		
参考文献	300		
附录 缩略语	302		

第1章 绪 论

1.1 人脸工程学

1.1.1 人脸工程学的研究内容

人脸是人类情感表达和交流最重要、最直接的载体。通过人脸可以推断出一个人的种族、地域,甚至身份、地位等信息;人们还能通过人脸丰富而负责的细小的变化,得到对方的个性和情绪状态。人脸在人与人的交流中不但能表达友好、敌对、赞成和反对等语气上的信息,甚至可以对话语、语言等语义上的信息进行说明和补充。从古至今,各类艺术创作者一直使用神态各异的人物来表达自己的思想、展现故事的情节。尤其在电影创作中,往往演员的一个眼神就能够将人物的内心展现无遗。正因为人脸在人的情感表达中扮演着重要的角色,人们很早就意识到人脸的重要性。1872年,Charles Darwin就出版了《人与动物的情绪表达(The Expression of the Emotions in Man and Animals)》一书,开始了对人脸表情的研究。长期以来,科学界从计算机图形学、图像处理、计算机视觉、人类学等多个学科对人脸进行研究。在这些领域中,人脸的识别、获取、生成和模拟一直是难点和热点。人脸的识别、获取、生成和模拟正是人脸工程学研究的主要内容。

人脸工程学研究是指以人类学、心理学、脑科学、人文科学、认知科学、信息科学、人工智能等学科为理论基础,利用工程的方法和技术(尤其是信息技术),对人脸(包括识别、建模与重构及其应用等相关内容)进行研究。人脸工程学的研究内容主要包括人脸识别、表情识别和人脸合成三部分。

1. 人脸识别

广义的人脸识别是指分析待识别的人脸图像,从中提取出有效的信息,并与数据库中的已知人脸信息进行比较,从而得出决策或认证信息的一种技术。其研究内容包括以下五个方面:

(1) 人脸检测 从各种不同的背景中检测是否存在人脸,并确定其位置、大小、形状、姿态等信息的过程。它关系到后续识别工作能否正确进行,并保障最终识别结果的可靠性。

(2) 人脸表征 确定表示检测出的人脸和数据库中的已知人脸的描述方式。通常的表示方法包括几何特征(如欧氏距离、曲率、角度等)、代数特征(如矩阵特征矢量)和固定特征模板等。

(3) 人脸鉴别 即狭义的人脸识别,就是通常所指的将待识别的人脸与数据库中的已知人脸进行比较,得出相关信息。这一过程的核心是选择适当的人脸表示方式和匹配策略,系统地构造与人脸的表征方式密切相关。

(4) 表情/姿态分析 即对待识别人脸的表情或姿态信息进行分析,并对其加以归类。

(5) 生理分类 对待识别人脸的物理特征进行分类,得出其年龄、性别、种族等相关

信息, 或从几幅相关的图像中推导出希望得到的人脸图像, 如从父母的脸像推导出孩子的脸像等。

本书中的人脸识别主要是指狭义的人脸识别, 指将待识别的人脸与数据库中的已知人脸之间进行匹配的人脸鉴别。

人脸识别的目的是让计算机具有通过人脸的特征来鉴别身份的功能。基于人脸特征的身份识别主要涉及复杂场景中的人脸检测及识别技术, 是一种依托于图像理解、模式识别及计算机视觉、统计学和人工智能等高技术的研究方向。

2. 表情识别

表情的研究融合了多个学科、多个领域, 计算机人脸表情识别的定义是: 利用计算机对人脸的表情信息进行特征提取分析, 按照人的认识和思维方式加以归类和理解, 利用人类所具有的情感信息方面的先验知识使计算机进行联想、思考及推理, 进而从人脸信息中去分析理解人的情绪, 如快乐、惊奇、愤怒、恐惧、厌恶、悲伤等。计算机表情识别的主要目的是利用计算机进行人类表情的模拟和分析, 进而更好地理解面部表情在艺术、人际交往, 特别是非语言交流以及情绪加工中所起的作用。

3. 人脸合成

人脸合成可以分为两类: 基于模型的人脸合成和基于图像库的人脸合成。

基于模型的人脸合成研究的内容主要分为两个部分: 人脸建模和人脸动画。

人脸建模主要研究人脸模型的静态建模、合成人脸模型的几何外形特征和纹理特征。其中, 既包含了如何合成同一个人脸模型在不同的表情和动作状态下形状和纹理特征, 也包含了如何合成不同人脸模型的形状和纹理特征。

人脸动画则研究人脸模型的动态过程, 研究人脸模型在表情和动作变化过程中, 几何外形和纹理的变化, 以及如何有效地驱动人脸模型做出复杂的表情和动作。

基于图像库的人脸合成步骤是: 从一个事先建立好的真人图像库中根据一定的规则挑选视频片断, 然后做相关的图像处理, 最后进行拼接, 从而达到合成的目的。

1.1.2 人脸工程学研究的意义与应用

总体来说, 有三个大方面的应用驱动着计算机人脸研究的进展: 其一是安全, 包括视频监控、生物识别和认证等, 相应的典型研究方向为人脸检测、人脸跟踪、人脸识别等; 其二是艺术, 可以应用在影视艺术、游戏等行业, 应用需求为: 如何使用计算机技术生成具有丰富表情的人脸, 其对应的典型研究方向为真实感/非真实感, 2D/3D 计算机人脸动画; 其三为人机交互, 主要研究方向为多通道人机接口(注视跟踪与识别)、用户认知建模、注意机制等, 应用需求为: 使计算机能够理解人的面部动作、行为、情感等, 并据此做出反馈。

下面分别从人脸识别、表情识别和人脸合成三方面来具体介绍人脸工程学研究的意义与应用。

1. 人脸识别

研究人脸识别在理论和技术上都具有重要的意义: 一是可以推进对人类视觉系统本身的认识; 二是可以满足人工智能应用的需要。采用人脸识别技术, 建立自动人脸识别系统, 用计算机实现对人脸图像的自动识别有着广阔的应用领域和诱人的应用前景。具体来说, 人脸识别技术的典型应用有:

(1) 身份鉴定（一对多的搜索） 在鉴定模式下，确定一个人的身份。人脸识别技术可以快速地计算出实时采集到的面纹数据与人脸图像数据库中已知人员的面纹数据之间的相似度，给出一个按相似度递减排列的可能的人员列表，或简单地返回鉴定结果（相似度最高的）和相对应的可信度。

(2) 身份确认（一对一的比对） 在确认模式下，面纹数据可以存储在智能卡中或数码记录中，人脸识别技术只需要简单地将实时的面纹数据与存储的相比对，如果可信度超过一个指定的阈值，则比对成功，身份得到确认。

(3) 监控 应用面像捕捉，人脸识别技术可以在监控范围中跟踪一个人和确定他的位置。

(4) 监视 可以在监控范围内发现人脸，而不论其远近和位置，能连续地跟踪他们，并将他们从背景中分离出来，将他们的面像与监控列表进行比对。整个过程完全是无须干预的、连续的和实时的。

(5) 面像数据压缩 能将面纹数据压缩到 84 字节，以便用于智能卡、条形码或其他存储空间有限的设备中。

2. 表情识别

人类的语言分为自然语言和形体语言两类。面部表情是形体语言的一部分，既是人们交往的一种手段，也是情绪表达的一种方式。在人类交往过程中，言语与表情经常是相互配合的。同时，表情比言语更能显示情绪的真实性。有时，人们能够运用言语来掩饰和否定其情绪体验，但是表情则往往掩饰不住内心的体验。情绪作为一种内心体验，一旦产生，通常会伴随相应的非言语行为，如面部表情和身体姿势等。一些心理学家研究发现，在人类进行会话交流传递信息时，言语内容占 7%，说话时的语调占 38%，而说话人的表情占 55%，所以说表情在人类交往活动中起到了重要的作用。

面部表情的分析与识别是实现智能化的人机接口必不可少的一个重要方面，是关于人工心理理论和情感计算研究的一个重要组成部分。计算机或机器人如果能够像人类那样具有理解和表达情感的能力，并能够自主适应环境，这将从根本上改变人机关系，使之能够更好地为人类服务。这就是研究人脸表情识别的理论意义。另外，人脸表情识别的研究还具有重要的学术价值。人脸是一类具有复杂细节变化的非刚性自然物体，研究人脸表情将为解决其他类似复杂模式的识别及其相关问题（例如人脸识别），提供重要的启示。同时，人脸表情识别的研究将对推动认知科学、生理学、心理学等相关学科的研究有着非常积极的影响。

在实际中，人脸表情识别有着广泛的应用前景，计算机的人脸表情识别能力对计算机视觉系统、建模和数据库的发展都有重要作用。当前，在人机交互中，结合进视觉、听觉以及更多的通道，如触觉、情感状态、表情等是必然趋势。未来计算机的发展将是以人为本为中心的，不仅计算机能通过多个通道、多种信息表现形式进行人机对话，而且能综合采用多个通道，使人能通过语言、文字、图像、手势、表情等与计算机以自然、并行、协作的方式进行人机对话，还能通过整合来自多个通道的输入来捕捉用户的交互意图，提高人机交互的自然性和高效性。此外，人脸表情识别也是视频人脸图像检索和低带宽人脸数据传输中的关键技术。当在图像流中检索人的动作以及情感状态的内容时，基于视觉的人脸表情识别是有用的。人脸数据的低带宽传输，可以通过利用对人脸动作的中间层和高层的视觉表示而更高效（例如，传送一个微笑和一些可以决定嘴巴动作的参数），这样在接收端就可以利用这些数

据恢复出人脸动作。在人类行为科学的研究方面，人们一直希望能揭开人脸表情的神秘面纱。由于人工进行表情动作编码太费时费力，人们希望能借助计算机来进行自动的人脸表情识别，并利用其结果来发现新的表情，以此作为人类行为研究的一个辅助手段。在语言学中，可辅助唇读；在行为学中，能帮助人们研究和建立交流中的可信度；另外，在其他方面，如面部神经瘫痪诊断、人脸图像合成与动画、智能机器人、智能监控等，表情识别都有着重要的应用价值。

3. 人脸合成

人脸是人类相互交流的渠道，在虚拟的计算机世界中，虚拟的人脸就是人脸表情和语言的载体。近几年来，人脸合成技术作为计算机图形学的一个独特分支越来越受到人们的重视。受电影和录像中高频率地出现虚拟人物、廉价的桌面处理能力和人机交互中需要一种全新的沉浸感的影响，人脸建模技术的研究成为近年来的一個新热点。

人脸建模系统具有广泛的应用。有些用于娱乐方面的，如电影的制作；有些使我们的生活变得更便利；有些则是用于特殊的目的，如电视电话技术。人脸合成的应用领域包括：

(1) 人机交互界面 现有的人机界面存在的普遍问题是单一、呆板和专业性强，用户往往要花费很长的时间才能学会使用一个新系统。若能给用户提提供熟悉的界面，比如会说话、会思考的人脸的形式出现，而不是面对单调的键盘，那么将会提高设备的可用性。

(2) 虚拟现实游戏 一般的游戏中的角色往往离现实很远，使用了快速人脸建模和动画技术的游戏，可以让人们在虚拟场景中用自己的虚拟形象进行体验和交互。

(3) 远程视频通信 直接传输视频图像在现有的网络带宽下不仅不现实，而且不必要。很多视频压缩算法借助人脸造型和动画的技术，能够快速地压缩人脸的视频信息，使其适合在网络上传输。

(4) 医疗研究 医疗中脸部建模的主要应用是外科手术和心理领域。建立好的人脸模型可预演校正外科手术和牙科诊治的过程。不断发展的人脸建模系统为心理学家进行人脸表情的研究提供了依据。心理学家可以用人脸运动的计算机模型来实现他们的研究，这比以前他们必须使用刺激肌肉来做研究更方便。

(5) 辅助教学 使用脸部建模的另一个应用是作为教学的工具。教人正常发音是一项艰辛而又单调的工作，而当学生在说话和听力方面有障碍时，更是如此。相反，学生可以在任意地方，跟着计算机模拟的老师学习，而不必一对一地教学，那将减少很多工作量。另外，将一个学习的过程做成游戏的形式，学生的学习积极性将大大提高，学习过程充满了乐趣，利用计算机，老师可以省去大量重复性的工作，并且一次教学能容纳更多的学生。更重要的是，学生可以随时随地学习。

(6) 公安鉴别 身份鉴别方法的不断提高有助于案件的侦破。通过操作三维人脸模型适配到目击者的描述。这样做的主要优点是三维图形给出犯罪分子更精确的描述，特别是当目击者没有看到嫌疑犯的正面时，该系统的价值更能得到体现。

(7) 考古人物复原 在考古研究中，经常会出土大量未知身份的颅骨，通过计算机辅助颅骨面貌复原技术有助于人物身份、种族的判定。如2001年，根据出土的老山汉墓墓主人的尸骨，利用颅骨面貌复原技术重建出了老山汉墓主人的生前面貌形象。在2016年举行的第39届数字考古国际会议上，北京师范大学信息科学与技术学院的周明全教授团队利用3D颅面复原技术，向世人展示了1200多年前唐高祖李渊第五代孙女李倕复原后的容颜。

(8) 电影特效和动画产品 从早期的《玩具总动员》《真实的谎言》《变脸》到最近的《阿凡达》和《变形金刚》等电影,都体现了人脸造型和动画技术的发展。电影制作者利用最先进的技术,把人的特点赋予到其他的动物和物体上,给这些角色带来鲜活的拟人个性。

1.1.3 人脸工程学研究在实用化过程中的挑战

人脸是日常生活中人们最为熟知的对象之一,相对于一般绘制对象,人脸具有如下重要特性:

(1) 生理结构复杂 面部的生理结构十分复杂,包括表皮、肌肉、骨骼三层,基本形状由最内层的骨骼决定,肌肉末端附着于骨骼上,肌肉和表皮间由韧带相连。整个头部骨骼通常称为头颅,由楔状骨、上颞骨、次鼻骨、颊骨、颞骨、眼眶骨、鼻犁骨、眉骨、颅骨、下颞骨、筛骨组成。肌肉的缩张驱动表皮组织产生运动,导致面部表现形式的变化,所有面部肌肉运动综合作用就产生了丰富多彩的表情。上述生理解剖学的原理是人脸建模的基础和依据。

(2) 形态内容丰富 心理学研究表明,人脸能够产生大约 55000 种不同的表情,其中有多种能够用人类自然语言词汇区别开来。

(3) 结构、表情上共性明确 所有人的面部结构和表情变化都有着明确的相似性。生理结构上都由口、眼、鼻、耳、眉等五官组成,头颅结构也完全相似;表情表达上,所有人脸都存在着共性,甚至动态的变化过程也十分相似。

(4) 个性因素繁多 人脸存在共性的同时,又有着千差万别的个性。例如,人眼睛虹膜近乎相同的概率是百万分之一,人耳朵形状的差别更大。不同的人种具有不同的肤色、五官特征。表情的细节也各有特点,没有两个人的笑容完全相同。

(5) 易受环境影响 摄取的人物的图像、视频随着周围光照环境的不同,差别很大;因为面部的形状不是严格的凸结构,所以有时会出现光照上的遮挡;人们有时会佩戴眼镜。这些都会给计算机处理带来很大困难。

下面分别从人脸识别、表情识别和人脸合成三个方面来具体介绍人脸工程学研究在实用化过程中的挑战。

1. 人脸识别的难点

目前人脸识别的难点主要存在于以下几个方面:

1) 光照变化是影响人脸识别性能的最关键因素,对该问题的解决程度关系着人脸识别实用化进程的成败,在人脸图像预处理或者归一化阶段,尽可能地补偿乃至消除其对识别性能的影响。

2) 成像角度及成像距离等因素的影响,即人脸的姿态变化,会使垂直于图像平面的两个方向的深度旋转,会造成面部信息的部分缺失。

3) 不同年龄的人脸有着较大的差别。身份证是以前照的,在逃犯的照片也是以前的,因此在公安部门的实际应用中,年龄问题是一个最突出的问题。

4) 采集人脸图像的设备较多,主要有扫描仪、数码相机、摄像机等。由于成像的机理不同,形成了同类人脸图像的识别率较高,而不同类别间人脸图像的识别率较低的情况。随着人脸识别技术的发展,这一问题也将逐步得到解决。

5) 人脸的图像数据量巨大。目前出于计算量的考虑,人脸定位和识别算法研究大多使

用尺寸很小的灰度图像。一张 64×64 像素的 256 级灰度图像就有 4096 个数据, 每个数据有 256 种可能的取值。定位和识别算法一般都很复杂, 在人脸库较大的情况下, 计算量很大, 很多情况下速度令人难以忍受。而灰度数据事实上是丧失了色彩、运动等有用信息的。如果要使用全部的有用信息, 计算量就更大了。

另外, 人脸识别还涉及图像处理、计算机视觉、模式识别以及神经网络等学科, 也和人脸的认识程度紧密相关。这些因素使得人脸识别成为一项极富挑战性的课题。

2. 表情识别的难点

人类用肉眼识别人脸信息时几乎没有什么困难, 但对计算机视觉系统来说则非常不容易。用计算机来分析识别面部表情是一个非常复杂的、极富挑战性的问题, 它的挑战表现在:

- 1) 人脸是一个塑性变形体而不是刚体, 难以对人脸表情进行建模。
- 2) 表情对应于人脸部骨骼和肌肉的运动, 其表现是细微多变的, 可以缓和也可以激动, 所以非常复杂。
- 3) 光照对于人脸图像的影响也很大, 同一个人的同一个表情图像, 如果光照不同, 可能得到完全错误的分类结果。
- 4) 表情识别课题研究涉及多学科的发展, 比如图像处理、计算机视觉、模式识别、应用心理学、生命科学、认知科学等, 而其中的大部分学科都是刚刚兴起, 处于探索阶段, 理论和方法仍有待完善; 现有的经典的图像处理技术的局限性也是一个不容忽视的问题。

此外, 没有统一的表情库, 也使得难于对各种识别方法进行比较和判断。目前可供研究的数据中, 数据库表情带有很强的人为色彩, 而现实生活中人们的表情可能并不夸张, 因而实验室中采用的方法将难以应用到实际生活中。

3. 人脸合成的难点

在计算机技术飞速发展的今天, 用计算机便捷、逼真地模拟人脸的造型及其运动仍然面临着巨大挑战, 这主要有几个方面的原因:

(1) 人脸生理结构复杂 人脸面部由表皮、肌肉和骨骼三层结构构成, 其基本形状由内层的骨骼决定, 肌肉末端附着于骨骼上, 表皮与肌肉紧密相连。面部的长相由骨骼和肌肉共同决定, 面部的表情变化是骨骼、肌肉以及表皮和皮下层的物理运动综合作用的结果。另外, 表皮的颜色和纹理的细微变化都会影响人脸的视觉效果。

(2) 人脸的几何结构复杂 人脸不但生理结构复杂, 几何结构也很复杂, 而且人与人之间的几何结构均不相同。人的头发、嘴巴、耳朵、眼睛的几何结构很难用一个恰当的模型来表示, 其几何数据也很难获得, 人脸运动时的几何结构变化则更难模拟。人脸的运动包括刚性的人脸姿态变化和非刚性的人脸器官和肌肉运动, 这些由人脸的皮下肌肉和人脸皮肤的力学特性决定的运动很难用模型来表示。

(3) 人脸表情丰富 人们通常把人脸大致分成六种表情: 高兴、生气、害怕、惊奇、难过和厌恶。而且人的心境往往同时包含多种情绪, 因此实际的表情就更要复杂得多。

(4) 个性因素 人脸在具有相同的拓扑关系的同时, 具有非常明显的个性因素, 如种族、形状、颜色以及纹理的不同。除了外观差异外, 表情动作的细节也是千差万别。

(5) 光照的影响 人脸的皮肤存在着皮下散射和折射, 眼睛的光照特性很难获得, 人脸的复杂几何以及皱纹、眉毛和脸部细毛使人脸的光照特性非常复杂。目前还没有较好的方

法来直接测量人脸的光照特性。

对于基于图像的方法而言,建立一个性能优良的图像库存在较大难度。一方面,为了使得系统能够合成尽可能多的表情和姿态,应该使图像库包含尽可能多的图像样本。但是由于图像本身具有较大的数据量,因此大量的图像的存储需要消耗很多的资源,而且在合成阶段,过多的图像样本会对搜索速度带来负面影响,从而难以达到实时要求;另一方面,图像库中的样本过少,又使得合成备选单元的数量不足,影响最终的合成质量。怎样使数据库在数量和质量之间取得某种折中是一个难题。

1.2 人脸工程学研究历程

计算机对于面部的自动分析和识别可以分为两个方面:一是永久面部特征分析;二是运动面部特征分析或者称瞬间面部特征分析。下面就分别介绍这两种面部分析的起源、发展、研究现状以及应用等。

第一种,永久面部特征分析,主要是针对面部的轮廓、五官、纹理等特征,对应地称为面像识别或人脸自动识别技术,就是利用计算机对于输入的人脸图像或者视频,首先分析其中是否存在人脸,如果存在人脸,则给出人脸的位置和大小,然后从检测到的人脸提取人脸中的有效识别信息,并将其与已知人脸图像库中的人脸进行比对,这是用来辨认身份的一种技术。

第二种,运动面部特征分析,主要就是表情的分析、编码与识别。让未来的计算机会思考、有感情、能说话、可以与人类平等交流一直是人类的美好愿望,通过计算机面部分析,但愿有那么一天,计算机能够理解我们的表情。当一位老人在 ATM 屏幕前眯着眼辨认字体时,字体尺寸立刻加倍。当一名妇女对着购物中心的旅游广告微笑时,会促使设备打印出一张旅游折扣表给她。这些不是幻想,一切都会实现的。现在许多研究机构都看好表情机器识别技术,相信此技术具有很大的科研价值和商业价值。

1.2.1 人脸识别

人脸识别的研究始于 20 世纪 60 年代末,最早的研究见于参考文献[44],Bledsoe 以人脸特征点的间距、比率等参数为特征,建成了一个半自动的人脸识别系统。而且早期人脸识别研究主要有两大方向:一是提取人脸几何特征的方法,包括人脸部件归一化的点间距离和比率,以及人脸的一些特征点,如眼角、嘴角、鼻尖等部位所构成的二维拓扑结构;二是模板匹配的方法,主要是利用计算模板和图像灰度的自相关性来实现识别功能。Berto 在 1993 年对这两类方法做了较全面的介绍和比较后认为,模板匹配的方法优于几何特征的方法。

目前的人脸识别研究也主要有两个方向:其一是基于整体的研究方法,它考虑了模式的整体属性,包括特征脸(Eigenface)方法、SVD 分解的方法、人脸等密度线分析匹配方法、弹性图匹配(Elastic Graph Matching, EGM)方法、隐马尔可夫模型(Hidden Markov Model, HMM)方法以及神经网络的方法等;其二是基于特征分析的方法,也就是将人脸基准点的相对比率和其他描述人脸脸部特征的形状参数或类别参数等一起构成识别特征向量。这种基于整体脸的识别不仅保留了人脸部件之间的拓扑关系,而且也保留了各部件本身的信息,而基于部件的识别则是通过提取出局部轮廓信息及灰度信息来设计具体识别算法。有文献认

为, 基于整个人脸的分析要优于基于部件的分析, 理由是前者保留了更多的信息, 但是这种说法值得商榷, 因为基于人脸部件的识别要比基于整体的方法来得直观, 它提取并利用了最有用的特征, 如关键点的位置以及部件的形状分析等, 而对基于整个人脸的识别而言, 由于把整个人脸图像作为模式, 那么光照、视角以及人脸尺寸会对人脸识别有很大的影响, 因此如何能够有效地去掉这些干扰非常关键。虽然如此, 但对基于部件分析的人脸识别方法而言也有困难, 其难点在于如何建立好的模型来表达识别部件。近年来一个趋势是, 将人脸的整体识别和特征分析的方法结合起来, 如 Kin Man Lam 提出的基于分析和整体的方法, Andreas Lanitis 提出的利用可变形模型 (Flexible Model) 来对人脸进行解释和编码的方法。

计算机人脸识别技术是近 20 年才逐渐发展起来的, 20 世纪 90 年代成为科研热点。近年来关于人脸识别的研究取得了很大的进步, 国际上发表有关论文的数量大幅增长, 仅 1990~1998 年之间, 工程索引 (Engineering Index, EI) 可检索到的相关文献就多达数千篇。美国电气电子工程师学会 (Institute of Electrical and Electronics Engineers, IEEE) 的 PAMI (模式分析与机器智能) 汇刊还于 1997 年 7 月出版了人脸识别专辑, 每年的国际会议上关于人脸识别的专题也屡屡可见。特别是, 随着近年来深度学习在模式识别等相关领域中获得广泛的应用, 研究基于深度学习的人脸识别技术将是未来一段时间内较为热门的研究方向。由于人脸识别实验所采用的人脸数据库通常不大, 最常见的人脸数据库仅包括 100 幅左右的人脸图像, 如 MIT (麻省理工学院)、Yale (耶鲁大学)、CMU (卡内基梅隆大学) 等人脸数据库均为小型库, 且由于不同的人脸数据库之间的输入条件各异, 因此不同的识别程序之间很难进行比较。为促进人脸识别算法的深入研究和实用化, 美国国防部发起了人脸识别技术 (Face Recognition Technology, FERET) 工程, 它包括一个通用人脸数据库和一套通用测试标准。该 FERET 人脸数据库可用于各种人脸识别算法的测试比较。1997 年, FERET 人脸数据库存储了取自 1199 个人的 14126 幅图像, 其中同一个人的图像差异, 包括不同表情、不同光照、不同头部姿态以及不同时期 (相隔 1 个月以上) 拍摄差异等。如今 FERET 人脸数据库仍在扩充, 并定期对各种人脸识别程序进行性能测试, 其分析测试结果对未来的工作起到了一定的指导作用。由于 FERET 人脸数据库中包括军人的图片, 不能在美国以外获得, 因此其他国家的研究只能采用本地的人脸库, 如英国的 Manchester (曼彻斯特) 人脸数据库。

1.2.2 表情识别

计算机人脸表情研究开始于 20 世纪 70 年代, 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, 进行表情分析。当然这还只是初步的尝试, 表情分析、编码与识别作为多模式人机交互的重要组成部分而受到关注, 而真正发展是在 20 世纪 90 年代。日本的东京大学、京都大学、ATR 研究所 (Advanced Telecommunications Research Institute International, 国际电气通信基础技术研究所); 美国的麻省理工学院 (MIT) 媒体实验室、卡内基梅隆大学 (CMU) 机器人研究所、马里兰 (Maryland) 大学、佐治亚工学院; 欧洲的剑桥大学、曼彻斯特大学等都投入很多人力物力来进行这方面的研究。

在日本, 20 世纪 90 年代初, 东京大学的 Hiroshi Kobayashi 和 Fumio Hara 利用神经网络

方法对基本表情和混合表情进行机器识别, 最终将表情分解为惊奇、恐惧、厌恶、愤怒、高兴和悲伤, 并将这 6 个基本表情中的 2 或 3 个表情进行合成。他们借助来自眉毛、眼睛、嘴三个区域的 60 个面部特征点 (Facial Characteristic Point, FCP) 的坐标来获取面部表情信息, 并将其作为神经网络的输入进行训练, 训练图像来自用录像带记录的 30 个对象的 172 幅具有混合表情的面部图像。最后, 将识别结果与人类对表情的识别结果进行了比较, 已经得到 70% 相对较好的识别率。日本 ATR 媒体整合与通信研究实验室的 Kenji Mase 提出利用光流来跟踪运动单元, 从而进行表情识别。Mase 的表情分析思想分从上至下和从下至上两个方向。①从上至下: 假设脸部图像被分解成肌肉单元 [对应于动作单元 (Action Unit, AU)], 把肌肉单元集成成矩形, 在矩形中计算光流。这种方法严重依赖于矩形的定位。②从下至上: 在矩形区域中计算光流, 量化成 4 个方向, 每个窗口提取了一个主要的肌肉收缩的方向。定义并提取一个 15 维的特征向量用来表征表情序列中光流变化最活跃的点。实验中对高兴、愤怒、厌恶、惊奇四种表情进行了分类, 用基于 K 最近邻的方法, 识别率达到了 80%。数据来源是作者本人的若干组各种表情图像序列: 20 种作为样本数据 (每种表情 5 组数据), 30 组作为测试数据, 图像大小为 256×240 像素。

在美国, 同样是 20 世纪 90 年代, 佐治亚工学院的 Irfan A. Essa 教授和 MIT 媒体实验室的 Alex Pentland 教授^[15], 用图像序列作为输入的计算机视觉系统来观察面部的运动单元。视觉观察与感知是通过优化估计光流方法与描绘面部结构的几何、物理肌肉模型相结合得到的。这种建模方法产生了一个随时间变化的面部形状的空间模板和一个独立的肌肉运动群的参数化表征。这些肌肉运动模板可以被用于分析、解释与合成表情。数据库中图像的大小为 450×380 像素, 来源于 7 个对象的 52 幅表情序列, 产生所“需要”的表情, 与真正意义下的情绪并无关系。识别的表情或表情动作是: 高兴、惊奇、愤怒、厌恶和抬眉毛, 识别率达到了 98%。马里兰大学的 Yaser Yacoob 和 Larry Davis, Black 和 Yacoob 的表情识别方法都是基于面部动作编码系统 (Facial Action Coding System, FACS) 编码的。Yacoob 和 Davis 集中于分析与嘴、眼睛和眉毛边缘的相关运动, 把光流的方向场量化成 8 个方向。同时建立了一个 Beginning-Apex-Ending 时间模型, 规定每种表情的整个过程以中性表情作为开始和结束, 并定义了变化中每个阶段的开始与结束的规则。识别方法使用简化的 FACS 规则来识别六种表情。他们的数据库来自 32 人的 105 种各种表情, 图像大小为 120×160 。对各种表情的识别率分别是: 高兴为 86%, 惊奇为 94%, 厌恶为 92%, 愤怒为 92%, 恐惧为 86%, 悲伤为 80%。Mark Rosenblum 和 Yaser Yacoob 等人用径向基函数网络 (Radial Basis Function Network, RBFN) 结构, 学习脸部特征与人类情绪之间的相关性, 在最高一级识别情绪, 在中间一级决定脸部特征运动, 在最低一级恢复运动方向。特征提取中不关注脸部的肌肉运动模型, 而是关注特征部件边缘的运动。此系统的识别率达到了 88%。

在欧洲, 剑桥大学计算机实验室的 Rana El Kaliouby 和 Peter Robinson 的研究主要针对: 自动识别用户实时的精神状态 (包括认知状态) 的接口, 以及发展表情在人机交互中的潜在用途。首先截取视频流中的头肩序列, 接着对图像进行运动单元分析, 最后利用 HMM 分类器分析头部运动和表情。实验测试在 207 个图像序列中进行, 其中包括 90 个基本表情和 107 个混合表情, 系统对于 6 种基本表情的识别率达到了 86%, 而混合表情的识别率为 79%。

曼彻斯特大学的图像科学与生物工程系的 David Cristinacce 和 Tim Cootes 采用 AdaBoost

算法和哈尔小波 (Haar Wavelet) 变换来提取面部特征, 该算法无论对于高分辨率还是低分辨率的图像, 都有稳定的人脸定位性能。

在国内, 北京科技大学王志良教授首次提出了人工心理的概念, 对人的心理活动 (包括情感、意志、性格、创造等) 进行人工模拟, 确立了人工心理理论结构体系 (目的、法则、研究内容、研究方法、应用范围), 并把这一理论应用于情感机器人、商品选购系统等实际生活中, 取得了较好的效果, 目前该课题组正在情感建模与计算、表情的分析与合成等方面进行深入探讨。此外, 哈尔滨工业大学、中国科学院计算技术研究所、中国科学院自动化研究所也开展了相关的研究。

1.2.3 人脸合成

人脸建模与动画的研究开始于 20 世纪 70 年代初, 1972 年, Parker 使用摄影测绘技术得到人脸各种表情的多边形网格模型^[18], 对于不同的表情的网格模型进行几何插值来产生人脸动画, 这种方法称为关键表情插值法。关键表情插值法需要建立每一种关键表情的网格模型, 关键表情可能有很多种, 因此使用这种方法很不方便。

1974 年, Parker 完成了第一个参数化的人脸模型, 1982 年又对参数化人脸模型进行了改进。参数化模型的基本思想是通过少量控制参数来生成各种不同模型的人脸和人脸的各种表情, 这样就可以方便地建立人脸模型、产生人脸动画。理想的参数化模型只需要选择合理的参数值就能够生成任何可能的人脸和任何可能的表情, 实际的参数化模型还远远不能达到这个目标。用于人脸建模与动画的参数可以分成两类, 即构造参数与表情参数。这两种参数在某种程度上已有重叠的部分, 但是在概念上还是有区别的。构造参数, 包括下颏宽度、前额形状、鼻梁的长度和宽度、脸颊形状等, 用于产生人脸的各种形状。表情参数, 包括眼睑张开、眉毛拱起、下巴旋转、眼睛凝视等, 用于产生人脸的各种表情。

1980 年, Platt 提出了第一个基于物理的肌肉控制人脸模型 (Physically Base Muscle-controlled Facial Model)。在该模型中, 人脸表面皮肤的多边形网格点之间通过模拟弹簧相互连接。这些网格点同时也与人脸模型的底层骨骼结构之间通过模拟肌肉相连。模拟肌肉具有弹性, 能够产生收缩力。将肌肉力作用在弹性网格上, 可以产生各种表情。使用的肌肉动作是按照 Ekman 和 Friesen 提出的面部动作编码系统 (FACS) 中的动作单元 (AU) 来设计的。

1985 年, Bergeron 和 Lachapelle 的动画影片《Tony de Peltrie》是人脸建模与动画历史上的一个里程碑, 该影片是第一个采用人脸表情和语音动画讲述故事的计算机动画片。

1987 年, Waters 提出了一种新的肌肉模型, 该模型包含两类肌肉: 一类是固定在骨骼上, 另一端连接在皮肤上的、可以拉伸的线性肌; 另一类是绕着中心点放大或缩小的括约肌。Waters 仍然采用 Platt 的质点弹簧模型来模拟皮肤和肌肉, 但是 Waters 模型中的肌肉具有独立于底层骨骼结构的方向 (向量) 性质, 这些向量使得模拟的肌肉与特定人脸的拓扑无关。每个肌肉都有一个影响区域, 肌肉的影响随着骨骼上的固定点的径向距离的增加而减少, Waters 模型的肌肉控制参数也是基于脸部动作编码系统来设计的。脸部运动是骨骼、肌肉、皮下组织和皮肤共同作用的结果, 其运动机理非常复杂。与肌肉模型不同, 伪肌肉模型不精确模拟人脸的解剖细节, 而是通过少量控制参数来模拟人脸的基本动作。

1998 年, Magnenat-Thalmann 等提出的抽象肌肉动作模型就是一种伪肌肉模型。抽象肌肉动作与脸部动作编码系统的动作单元类似, 但是不完全相同, 脸部动作编码系统的动作单