

视觉语音情感识别

詹永照 毛启容 著
林庆 成科扬



科学出版社

视觉语音情感识别

詹永照 毛启容 著
林庆 成科扬

科学出版社

北京

内 容 简 介

情感识别是近年来模式识别和人机交互领域的研究热点之一,它对推动自然的人机交互和类人机器人等技术发展有着重要的作用。本书共8章,首先介绍当前视觉语音情感识别技术研究概况,接着介绍基于视觉信息的情感特征提取方法和情感识别方法、语音情感特征选择提取方法和情感类别识别方法,然后介绍视觉语音融合的情感识别方法,最后介绍情感分析在人脸动画生成和 E-learning 环境中的应用。

本书对于从事情感计算研究的科技工作者是一本系统的参考书,也可作为高年级本科生和研究生学习情感计算和新型人机交互等课程的教材。

图书在版编目(CIP)数据

视觉语音情感识别/詹永照等著. —北京:科学出版社,2013
ISBN 978-7-03-037321-2

I. ①视… II. ①詹… III. ①视觉信息-情感-识别 IV. ①B842.2

中国版本图书馆 CIP 数据核字(2013)第 078081 号

责任编辑:王 哲 王 苏 / 责任校对:朱光兰
责任印制:张 倩 / 封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

骏志印刷厂印刷

科学出版社发行 各地新华书店经销

*

2013年4月第 一 版 开本:B5(720×1000)

2013年4月第一次印刷 印张:16 1/4

字数:327 000

定价:65.00 元

(如有印装质量问题,我社负责调换)

前 言

计算机科学技术的迅速发展改变了人们的工作和生活方式。人机交互(human computer interaction, HCI)技术是突破计算机与人类交互瓶颈的重要技术,它有力地促进了计算机的普及应用。为了使人类与计算机之间进行更加智能、更加自然的交互,新型的人机交互技术正在逐渐成为研究热点。人们不仅希望能以更方便、更高效、更自然的方式操纵计算机,而且还希望计算机能理解人的情感,并提供有价值的增值服务。因此,情感识别已经成为人机交互领域亟待突破的关键技术之一。

所谓情感识别,就是利用计算机分析特定表演者的面部表情、姿势和语音信号及其变化过程,进而确定该表演者的内心情绪或思想活动,实现人机之间更智能、更自然的交互。情感分析在许多领域都有着潜在的应用价值,如心理学研究、图像理解、脸部动画合成、视频检索、机器人技术和虚拟现实技术等领域。随着多媒体技术的发展,基于音视频的情感分析、识别研究对增强计算机的智能化和人性化、开发新型人机环境,以及推动多媒体技术和信号处理等相关领域的发展有着重要的意义。

本书作者近几年来在国家自然科学基金项目(60673190、61003183、61272211)的支持下,开展了基于视频和音频的情感识别技术的研究,包括视频和音频的情感特征的提取、适合于音视频情感分析、分类和识别的算法或分类器、多分类器融合的情感识别方法以及情感分析、识别的应用等。本书是在这些项目的研究成果基础上,系统化地加以归类总结撰写而成的。

全书内容分为8章。第1章为绪论,简要地介绍了视觉语音情感识别产生的背景和意义、研究内容及其应用领域。第2章为视觉语音情感识别技术概况,包括情感识别的框架、情感的定义与分类、情感信号的预处理、常见的情感特征提取与选择、现有的情感识别方法,以及情感识别的难点及新动向。第3章为基于视觉信息的情感特征提取方法,内容包括基于视频、图像的表情特征提取的相关方法,这些方法是对已经提出的一些方法的改进,并提出自己所设计的方法。第4章为基于视觉信息的情感识别方法,内容包括了较有特色的相关表情分类、识别方法,如针对样本的不均衡性,提出并实现的最大间隔最小体积球形支持向量机的表情识别方法;针对细微表情的识别问题,提出并实现的混合特征结合分类树的细微表情识别算法;针对视频信息存在噪声和部分遮挡问题,提出并实现的基于模糊深隐马尔可夫模型的图像序列表情识别方法。第5章为语音情感特征选择提取方法,内

容包括了语音情感测试库的构建、个性化和基于多重分形理论的语音情感特征提取方法、语音情感特征有效选择方法,以及基于流形学习的语音情感特征降维等方法。第6章为语音情感识别方法,内容包括了基于选择性特征的决策树的语音情感识别方法和基于改进有向无环图的分层语音情感识别方法。第7章为视觉语音融合情感识别方法,内容包括基于D-S证据理论的多粒度语段融合情感识别方法、多分类器融合方法和具有噪声过滤功能的分类器协同训练半监督主动学习方法等。第8章为情感分析的应用,内容包括基于表情动作单元参数的逼真表情动画方法和E-learning环境中的情感分析应用方法。

本书较全面地总结了课题组近年来的有关视觉语音情感识别的研究成果,内容系统、深入浅出、方法有新颖性和创新性。适合从事音视频信息的处理,特别是从事智能人机交互、计算机视觉与听觉、模式识别与人工智能以及智能动画等领域的科技工作者阅读参考。本书也可作为情感计算、新型人机交互和智能信息处理等课程的研究生教材。

本书由詹永照总体负责,毛启容负责组织实施。本书的第1、2章由毛启容和成科扬撰写,第3章由林庆撰写,第5、6章由毛启容撰写,詹永照负责了其余各章的撰写和全书的统稿工作。

完成这些项目的老师和合作者还有张建明、文传军、叶敬福、曹鹏、周庚涛、陈亚必、刘娟、李婷、陆捷荣、张娟、胡敏灵、徐莉婷、刘云、孔建等,他们在课题研究中刻苦钻研,做了有意义的探索性工作,也为本书的完成做出了重要贡献。在此谨向他们表示衷心的感谢。本书的撰写过程也参考了国内外研究者的研究成果和资料,也一并向他们致谢。

情感计算属于交叉学科的新兴研究领域,由于我们的水平有限,书中难免有疏漏之处,敬请读者不吝指正。

詹永照

2012年10月于江苏大学

目 录

前言

第 1 章 绪论	1
1.1 视觉语音情感识别的产生背景	1
1.2 视觉语音情感分析的研究内容	2
1.3 视觉语音情感识别的应用领域	3
1.3.1 表情识别的应用	3
1.3.2 语音情感识别的应用	5
参考文献.....	6
第 2 章 视觉语音情感识别技术概况	7
2.1 情感识别框架	7
2.2 情感描述模型	7
2.2.1 情感的定义	7
2.2.2 情感的分类	8
2.3 视觉语音信号预处理.....	10
2.3.1 人脸表情图像预处理	10
2.3.2 情感语音信号预处理	12
2.4 情感特征提取.....	16
2.4.1 视觉信息情感特征提取	16
2.4.2 语音情感特征提取	22
2.5 情感特征选择.....	24
2.6 常用的情感识别模型.....	25
2.6.1 基于相似性的情感识别模型	25
2.6.2 基于连接机制的情感识别模型	26
2.6.3 基于概率模型的情感识别模型	32
2.6.4 基于集成学习的情感识别模型	43
2.7 视觉语音情感识别的挑战.....	43
2.8 视觉语音情感识别的新动向.....	44
参考文献	45
第 3 章 基于视觉信息的情感特征提取方法	48
3.1 概述.....	48

3.2	基于小波分解和优选 VLBP 的表情特征提取方法	49
3.2.1	表情图像的小波分解	49
3.2.2	小波分解图像的情感特征提取	50
3.2.3	实验结果与分析	51
3.3	基于多频域 LBP-TOP 的人脸表情特征提取方法	53
3.3.1	LBP-TOP 算子	53
3.3.2	多频率图像分块 LBP-TOP 特征提取	56
3.3.3	实验结果与分析	57
3.4	基于 VLBP 与光流的混合情感特征提取	58
3.4.1	眼睛区域的小波分解分块 VLBP 特征提取	58
3.4.2	特征点自动标注的嘴部光流特征提取	59
3.4.3	基于混合特征的表情识别	62
3.4.4	实验结果与分析	62
3.5	基于 Gabor 变换的表情图像特征提取方法	63
3.5.1	小波变换与多分辨率分析	63
3.5.2	Gabor 变换	66
3.5.3	人脸表情图像的网格化	68
3.5.4	基于 Gabor 小波变换的表情弹性图的构造	69
3.5.5	实验结果与分析	71
3.6	基于积分图像的表情特征提取方法	72
3.6.1	积分图像的概念	72
3.6.2	积分图像表情特征提取	73
3.6.3	实验结果与分析	76
3.7	一种加权矩形提取表情特征的方法	77
3.7.1	矩形模板设计	77
3.7.2	基于加权矩形的表情特征提取	78
3.7.3	实验结果与分析	80
3.8	本章小结	81
	参考文献	82
第 4 章	基于视觉信息的情感识别方法	84
4.1	概述	84
4.2	最大间隔最小体积球形支持向量机	84
4.2.1	最大间隔球形支持向量机	84
4.2.2	最大间隔最小体积球形支持向量机	85

4.2.3	模型性能分析	89
4.2.4	基于最大间隔最小体积球形支持向量机的表情识别	91
4.3	混合特征结合分类树的细微表情识别算法	92
4.3.1	混合特征提取	92
4.3.2	基于分类树的表情识别	95
4.3.3	实验结果与分析	97
4.4	基于模糊深隐马尔可夫模型的图像序列表情识别方法	98
4.4.1	模糊深隐马尔可夫模型	99
4.4.2	模糊深隐马尔可夫模型的特性	101
4.4.3	基于模糊深隐马尔可夫模型的图像序列表情识别	106
4.4.4	实验结果与分析	107
4.5	本章小结	112
	参考文献	113
第5章	语音情感特征选择提取方法	114
5.1	概述	114
5.2	情感语音库的录制	114
5.2.1	语音情感数据库	115
5.2.2	音视频情感数据库	115
5.2.3	音视频情感的有效性分析	115
5.3	语音情感特征提取方法	116
5.3.1	传统声学语音情感特征分析与提取	116
5.3.2	基于导数的非个性化语音情感特征提取方法	124
5.3.3	基于多重分形理论的语音情感特征提取方法	134
5.4	语音情感特征选择方法	142
5.4.1	基于神经网络贡献分析的语音情感特征选择	142
5.4.2	基于遗传算法的语音情感特征选择和分类器参数优化方法	143
5.4.3	基于类集/类对的两级语音情感特征选择方法	145
5.4.4	基于扩展测地距离的语音情感特征有效性分析方法	146
5.4.5	实验结果与分析	148
5.5	基于流形学习的语音情感特征降维	153
5.5.1	Isomap 算法描述	153
5.5.2	ELE 算法描述	154
5.5.3	基于增量流形学习的情感特征降维方法	155
5.5.4	实验结果比较与分析	156

5.5.5 相关方法比较	157
5.6 本章小结	159
参考文献	159
第 6 章 语音情感识别方法	163
6.1 概述	163
6.2 基于选择性特征的 SVM 决策树的语音情感识别方法	163
6.2.1 情感混淆度	163
6.2.2 基于 SVM 一对一算法的语音情感识别	164
6.2.3 基于 SVM 决策树和选择性特征的语音情感识别	166
6.3 基于改进有向无环图的分层语音情感识别方法	173
6.3.1 基于 SVM 的有向无环图	173
6.3.2 改进有向无环图的构造算法	174
6.3.3 基于测地距离的待识别样本鉴别度量算法	175
6.3.4 基于 SVM 的改进有向无环图的构造	178
6.3.5 语音情感识别实验	179
6.3.6 相关工作比较	183
6.4 本章小结	184
参考文献	184
第 7 章 视觉语音融合情感识别方法	186
7.1 基于混合特征和多 HMM 融合的图像序列表情识别	186
7.1.1 多 HMM 融合的图像序列表情识别方法	186
7.1.2 实验结果分析与比较	188
7.2 基于 D-S 证据理论的多粒度语段融合语音情感识别	189
7.2.1 D-S 证据理论简介	189
7.2.2 多粒度情感语句分段方法	190
7.2.3 多粒度语段融合情感识别算法描述	192
7.2.4 实验结果分析与比较	193
7.2.5 相关工作比较	195
7.3 基于对象模糊密度赋值的决策级层次式融合算法	195
7.3.1 模糊测度、模糊密度和模糊积分	195
7.3.2 类模糊密度与混淆矩阵	196
7.3.3 对象模糊密度	197
7.3.4 基于对象模糊密度赋值的决策层融合识别算法	198
7.3.5 基于模糊密度的人脸表情融合识别	199

7.3.6 实验结果与分析	200
7.4 具有噪声过滤功能的分类器协同训练半监督主动学习算法	201
7.4.1 具有噪声过滤功能的协同训练半监督主动学习算法	201
7.4.2 基于 NF-CT-SSAL 训练算法的半监督主动学习人脸表情识别	205
7.4.3 实验结果与分析	206
7.5 基于可分度和支持度的模糊密度赋值融合识别算法	209
7.5.1 分类器的初始模糊密度	210
7.5.2 分类器的自适应模糊密度赋值	211
7.5.3 算法描述	213
7.5.4 算法在人脸表情识别中的应用	214
7.5.5 实验结果与分析	214
7.6 本章小结	218
参考文献	218
第 8 章 情感分析的应用	220
8.1 概述	220
8.2 人脸表情动画中的逼真人脸重构	220
8.2.1 一般人脸模型	220
8.2.2 特定人脸模型重构	221
8.2.3 实验结果与分析	231
8.3 基于表情动作单元参数的逼真表情动画方法	233
8.3.1 ERI 介绍	234
8.3.2 图像的对准方法及分部位表情比率图库的建立	235
8.3.3 表情动画中细微纹理生成	237
8.3.4 实验结果与分析	240
8.4 情感分析在 E-learning 环境中应用	242
8.4.1 基于规则的本体推理	242
8.4.2 E-learning 环境中的本体规则推理情感激励	245
8.4.3 情形分析及情感激励实例	246
8.4.4 相关工作比较	246
8.5 本章小结	247
参考文献	247

第 1 章 绪 论

1.1 视觉语音情感识别的产生背景

计算机科学技术的迅猛发展对人类社会产生了巨大的影响,一场信息革命正在酝酿。人们认知自然、改造自然的方式也随之发生变化。在众多领域,计算机正在逐渐代替人类完成一些极具挑战性的任务。为了使人类与计算机之间进行更加智能、更加自然的交互,新型的人机交互技术正逐渐成为研究热点。人类自然形成的与自然界沟通的认知习惯和形式必定是人机交互的发展方向。一方面要求计算机能看、能说、能听、能感觉,即能够智能地感知使用者的意图;另一方面,使用者可以不必坐在计算机前通过鼠标和键盘操作计算机,而是在三维空间中以更加自然和人性化的方式同计算机交流。总之,新型的人机交互环境要求计算机必须具有自然化、智能化和人性化等特点。

实时的多媒体信息在新型的人机交互环境中扮演着举足轻重的角色,计算机应该能够通过采集用户的图像信息和语音信息,形成计算机视觉和听觉,然后处理采集到的多媒体信息,并进行识别,从而感知使用者的意图,完成人机间的自然交互。实时的视频图像传递更加丰富的信息,这对于感知使用者的意图,判断其行为极具价值,因此,目前国内外众多研究学者将信号处理技术应用于实时的视频信息,使用计算机处理视觉语音信息,研究并开发新型的人机交互系统。在该研究领域,基于数字图像的人脸检测技术、人脸识别技术、手势识别技术、语音识别技术和基于音视频信息的情感识别技术等陆续被提出,相关的应用系统也将逐渐改变人们的工作与生活。

所谓情感识别,就是利用计算机分析特定表演者的面部表情、姿势或语音信号及其变化过程,进而确定该表演者的内心情绪或思想活动,实现人机之间更智能、更自然的交互。情感分析在许多领域都有着潜在的应用价值,如心理学研究、图像理解、脸部动画合成、视频检索、机器人技术和虚拟现实技术等^[1]。随着多媒体技术的发展,基于音视频的情感分析、识别研究对于增强计算机的智能化和人性化、开发新型人机环境,以及推动多媒体技术和信号处理等相关领域的发展有着重要的意义。

表情是人类交流中信息传递的主要媒介,在人类的日常生活中扮演着重要的角色,是语言交流的重要补充和非语言交流的最主要方式之一。表情包含了丰富

的情感信息,是情绪的外在表现^[2],即情绪在生理上、心理上和外在行为上所表现的一切变化或活动,同时也是人的社会行为的外在表现^[3]。由于人的脸部表情是反映人的内心情绪和想法的最自然最直接的方式,因此,基于视觉信息的人脸表情识别技术的研究是继人脸检测与识别之后的又一研究热点。

语音是人类表达情感的重要通道之一,语音信号中包含了大量的情感信息。近年来,从语音信号的角度研究说话人情感状态的研究越来越多,已成为情感识别的一个重要分支。

1.2 视觉语音情感分析的研究内容

视觉语音情感分析的主要研究内容概括为六个方面^[4]。

1) 情感机理和描述

情感机理和描述即研究人的情感状态的判定以及生理和行为的关系,并以一套形式化或模型的方式加以表达,这方面的研究涉及心理学、生理学及认知科学等,其研究成果将为情感分析提供理论基础。人类情感的研究是一个非常古老的话题,心理学家、生理学家已经在这方面做了大量的工作。任何一种情感状态都可能伴随几种生理或行为特征的变化,而某些生理或行为特征也可能起因于数种情感状态。因此,确定情感状态与生理或行为特征之间的对应关系是情感分析理论的一个基本前提,这些对应关系目前还不十分明确,需要进一步探索和研究。

2) 视觉语音情感信号的获取和量化

视觉语音情感信号的获取和量化即通过摄像头、麦克风等设备记录情感变化信号,如语音、面部表情、手势及姿势等声音和体态语的变化信息,并结合心理学、生理学及认知科学等关于情感状态的表征,研究各种情感状态特征的提取方法,便于计算机进行分析处理,力求能更准确分析、判断和理解人类的各种情感状态。

3) 视觉语音情感信号的分析、建模与识别

一旦由各类有效传感器获得了情感信号,并针对情感状态可能的表征提取并量化出数字化信息,接下来的任务就是将情感信号与情感机理相应方面的内容对应起来,这里要对所获得的情感信息进行建模和识别。视觉语音情感信号的分析、建模与识别即对所获取的情感信息进行分析,基于情感机理建立各种情感状态模型,并获取欲分析识别的视觉语音所对应的情感信息,通过事先或训练得到的情感状态模型识别该视觉语音所对应的情感状态。由于情感状态是一个隐含在多个生理和行为特征之中的不可直接观测的量,不易建模,所以部分可采用诸如情感信息变化规则、隐马尔可夫模型、贝叶斯网络模式及支持向量机(support vector machines, SVM)等模型。情感信息的非线性关联关系的建模更能使模型具有泛化性

和鲁棒性。如 MIT 媒体实验室给出了一个隐马尔可夫模型,可根据人类情感概率的变化推断得出相应的情感走向。因此,研究如何度量人类情感的深度和强度的、定性和定量的情感度量的理论模型、指标体系、计算方法及测量技术,将是视觉语音情感信号的分析、建模与识别的一个重要研究方向。

4) 情感理解和反馈

情感理解和反馈即通过对情感的获取、分析与识别,计算机便可了解其所处的情感状态,同时理解情感信号识别的结果,分析人的情感产生的原因,并做出合理恰当的情感反应。情感分析的最终目的是使计算机在了解用户情感状态的基础上,做出适当反应,去适应用户情感的不断变化。因此,这部分主要研究如何根据情感信息的识别结果,对用户的情感变化做出最适宜的反应。在情感理解的模型建立和应用中,应注意以下事项:情感信号的跟踪应该是实时的,并保持一定时间间隔进行记录;情感的表达是根据当前情感状态进行适时的表达;情感模型是针对个人生活的并可在特定状态下进行更新;情感模型应具有自适应性,可通过理解情况适当反馈调节识别模式。

5) 情感合成与表达

情感合成与表达即通过计算机交互设备向用户表达所对应对象的情感。上述的研究技术是从声音或行为特征来推断情感状态,而情感合成与表达则是研究其反过程,即给定某对象某一情感状态,研究如何使这一情感状态在该对象的声音或行为特征中表征出来。例如,如何在语音合成和面部表情合成中得以体现所交互的对象情感,使所交互的对象富有情感,能够方便与用户进行交流。情感的表达提供了用户与虚拟对象情感交互和交流的可能,对于单个用户,情感的交流主要包括人与人、人与机、人与自然以及人与虚拟对象的交互、交流。

6) 情感人机交互的实现

情感人机交互的实现即形成一个完整的情感交流过程,实现自然和谐的人机交互。情感人机交互的实现主要在情感表达的基础上,进一步研究如何在计算机或机器人中,通过视觉和语音通道,分析、理解、模拟或生成情感模式,设计实现虚拟或实体的情感机器人或具有人工情感的计算机及其应用系统的情感生成、表达的方法和技术。

1.3 视觉语音情感识别的应用领域

1.3.1 表情识别的应用

人脸表情识别在许多领域都有着潜在的应用价值,这些应用领域包括智能人机交互、安全领域的智能监控、医疗领域与日常监护、数字娱乐等^[5-7]。

1) 智能人机交互

在以人为中心的交互方式中,设计、构建自然和谐的智能人机交互界面,让计算机通过观察人脸表情,识别和理解人的情感,并能针对人的情感做出智能、灵敏、友好的回应,更好地服务人类。显然,在人机交互的许多应用场合,人脸表情识别是系统中十分重要的一环。例如,美国麻省理工学院和英国剑桥大学的开发人员联合开发一种“情绪意识”计算机,这种技术可以识别用户的面部表情,通过对人面部运动的综合分析,得出人的心理情绪活动并在电脑上做出相应的回应,这种技术将在广告、汽车驾驶和网络上具有广泛的用途。

2) 安全领域的智能监控

对于核电站管理、长途汽车司机等涉及重大安全的工作岗位,如果工作人员疲劳工作,很容易发生险情。表情识别非常适合这些安全领域的智能监控,通过动态监测工作人员的表情状态,一旦出现疲劳、瞌睡的征兆,识别系统会自动及时发出警示避免险情发生。表情识别还可用于公共场合安全监控,通过人们脸部表情分析,在发现有人出现精神异常或过激表情时,监控系统及时予以提示,预防犯罪或危险事件的发生,维护公共安全。例如,美国伊利诺斯大学的 Thomas 教授一直致力于通过面部表情的识别读取人物情绪状态的研究。他和他的学生于 2005 年年底建立的一套名为“声音、面部、情绪”的识别系统,已经引起了美国国土安全部的兴趣,准备在机场应用来识别恐怖分子。

3) 医疗领域与日常监护

在心理疾病治疗中,表情分析可作为辅助手段,帮助医生分析病人的精神状态,做出正确的诊断。在远程家庭、幼儿园和老人园的日常监护中,通过对儿童、老人监护或需监护人员的脸部表情变化,及时发现其情绪和身体状况的变化,避免不必要事故的发生。

4) 数字娱乐

微软亚洲研究院正在通过对人脸脸部表情的分析研究,完成一系列卡通人物模型,准备集成到下一代 Office 产品中,使人们在紧张的工作之余享受到快乐。另外,在游戏中,如果可以根据游戏者的喜、怒、哀、乐来做出相应的反应,那么这样的游戏肯定比那些传统游戏更加吸引人。动漫等影视产品,通过人脸表情识别、合成技术为角色赋予丰富的表情变化,比采用手工方式设计大大提高了制作效率。随着游戏产业和其他数字娱乐产业以惊人的速度迅猛发展,市场潜力巨大,表情识别、模拟技术具有相当大的应用价值。例如,索尼公司开发的 AIBO 狗可以通过同人进行交流来“学习”某些动作,对周围环境做出反应,表露某种感情如高兴、生气等。索尼公司的 T200 数码相机具有“笑脸快门”(smile shutter)的功能,在识别到被摄人微笑的表情时,可以自动记录下图像。

1.3.2 语音情感识别的应用

从语音中自动识别说话者的情感状态不仅可以应用于人机交互系统,在其他方面也具有广阔的应用前景^[8],其中包括如下几个方面。

1) 交互电影中自发(spontaneous)交互的判定

在传统的被动式电影观看模式中,观众所能亲身经历、感受和学习到的东西是非常有限的。随着数字媒体的丰富和计算机技术的发展,这种电影模式已经不能满足观众的需求。交互式电影将为人们提供一种全新的感受。在这种电影中,人们不再只是预测其中人物的命运,而是能亲身经历日常生活中从未有过的剧情和故事。通过这种方式将为大家提供学习不同本领和课程的大好机会。交互电影的关键因素之一,就是能与作为主要角色的参加者进行交互。日本ATR媒体集成和通信研究实验室,通过用情感识别将自发交互能力引入到了交互电影中^[9],在研究过程中,他们同时使用语音识别和情感识别来决定是进行预先设计好的演出方式,还是进入自发交互方式。

2) 辅助残疾人讲话

失语症是很多脑部疾病带来的直接后遗症,这些患者有着与健康人同样的情感却苦于无法表达。为此,有很多研究机构开始了辅助残疾人说话的相关产品的研究和开发,如VAESS(Voices Attitudes and Emotions in Synthesis Speech)工程开发研制了一种既能够辅助残疾人说话又能帮助他们表达情感的一种便携式情感语音合成器。文献[10]中给出了一个辅助残疾人进行情感语音输出的系统CHATAKO,该系统可根据用户输入的文本信息和所选择的情感选项,输出有情感的合成语音。

3) 情感语音将使人机界面更具人性化

随着计算机技术和网络技术的发展,人们可以通过互联网获取大量的信息,在电子购物、网上医疗、网上聊天、电子会议和有声电子邮件等应用上,人们希望听到的已不再是枯燥的机器音,而是更具“人情味”的语音。传统的利用菜单命令和目录方式进行的信息管理已不能完全满足人们的需求,人们需要更自然、更智能、更人性化的人机界面。在不久的将来,人机之间的交互不仅依靠键盘和鼠标,简单、易学、更具“人性化”的语音操作界面应当更符合人们的实际需求。这种人机之间的交互离不开情感语音识别技术和情感语音合成技术。

4) 情感语音与其他多媒体技术相结合

将情感语音配以相应面部特征的视频来传达情感,通过将一些视觉效果,包括人的头部建模、唇形同步技术和表情因素等视频信息加入,使声音、表情同步,这就是当前比较热门的“视觉语音”(visual speech)技术。视觉语音可以使输出效果更具表现力和感染力,虚拟主持人的实现就是利用了这一技术。

此外,情感语音识别还可以用于自动远程电话服务中心,及时发现客户的不满情绪^[11];用于远程教学和婴儿教育,及时识别学生的情绪并进行适当处理,从而提高教学质量;用于辅助临床精神分裂症的诊断和治疗;也可以用于刑事侦查中自动地检测犯罪嫌疑人的心理状态,辅助测谎^[12]。语音情感识别还可以用于信息查询系统、电子商务、娱乐游戏和虚拟人对话等领域。

总之,自动语音情感识别的研究,不仅可以推动计算机技术的进一步发展,也将大大提高人们的工作和学习效率,更高效地帮助人们解决问题,同时也将进一步提高人们的生活质量。

参 考 文 献

- [1] 王文成. 基于局部特征分析的人脸表情识别问题研究[博士学位论文]. 济南: 山东大学, 2008.
- [2] Soleymani M, Pantic M, Pun T. Multi-modal emotion recognition in response to videos. *IEEE Transactions on Affective Computing*, 2011, 99(1): 267-276.
- [3] Chin S, Kim K Y. Emotional intensity-based facial expression cloning for low polygonal applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2009, 39(3): 315-330.
- [4] 移动 Labs. 情感计算及分析. http://labs.chinamobile.com/mblog/382108_72248[2010-11-29].
- [5] Dubussion S, Devoine F, Masson M. A solution for facial expression representation and recognition. *Signal Processing: Image Communication*, 2002, 17(9): 657-673.
- [6] Otte M, Nagel H H. Estimation of optical flow based on higher-order spatiotemporal derivatives in interlaced and non-interlaced image sequences. *Artificial Intelligence*, 1995, 78(1/2): 5-43.
- [7] Chen X W, Huang T. Facial expression recognition: a clustering-based approach. *Pattern Recognition Letters*, 2003, 24(9/10): 1295-1302.
- [8] 韩纪庆, 邵艳秋. 基于语音信号的情感处理研究进展. *语音技术*, 2006, 5: 58-67.
- [9] Nakatsu R, Nicholson J, Tosa N. Emotion recognition and its application to computer agents with spontaneous interactive capabilities. *Knowledge-Based Systems*, 2000, 13: 497-504.
- [10] Iida A, Campbell N, Higuchi F. A corpus-based speech synthesis system with emotion. *Speech Communication*, 2003, 40: 161-187.
- [11] Batliner A, Fischer K, Huber R. How to find trouble in communication. *Speech Communication*, 2003, 40(1/2): 117-143.
- [12] Cowie R, Douglas C E, Tsapatsoulis N, et al. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 2001, 18(1): 32-80.

第 2 章 视觉语音情感识别技术概况

2.1 情感识别框架

情感识别研究涉及的关键技术大致包括四个方面,即情感库的建立、基于视觉语音信号的情感特征提取、情感特征降维和情感识别方法,如图 2.1 所示。情感特征提取是从经过预处理以后的视觉语音情感信号中提取能够有效地反映情感变化的情感特征,它是情感识别的基础。特征提取是否准确直接影响情感识别的效果。情感特征降维是指经过一定的降维算法,去除所提取的大量情感特征中的冗余特征,在不影响情感识别率的前提下,减少所提取情感特征的个数,改善情感识别的实时性。在情感识别方法方面,现有的识别方法很多,从是否需要带情感标签样本对分类器进行指导的角度可分为无监督的、有监督以及半监督的情感识别方法。本书主要对有监督和半监督情感识别方法进行讨论。本章的其余部分将对情感数据库的建立、基于视觉语音的情感特征提取、情感特征降维以及情感识别方法四个方面的关键技术以及发展概况进行介绍。

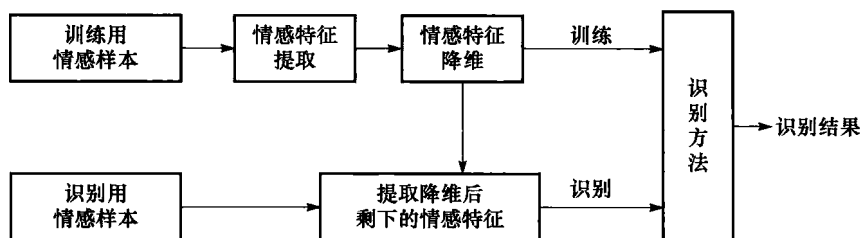


图 2.1 情感识别框架图

2.2 情感描述模型

2.2.1 情感的定义

究竟什么是情感?这是研究情感识别首先需要回答的问题。但这个问题比较复杂,虽然科学家对情感产生的机理进行了广泛的研究,但到目前为止,对于情感的定义仍然没有一个统一的说法。表 2.1 为《现代汉语词典》中情感、感情、情绪三个词汇的解释^[1]。