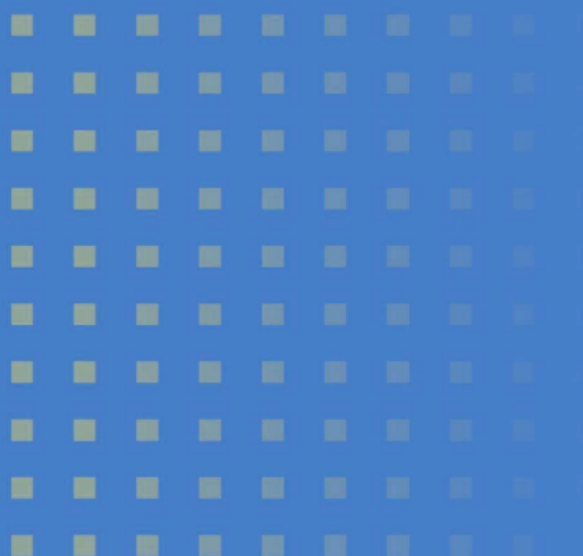


智能环境下基于音视频 多模态融合的身份识别

吴 迪 著



天津出版传媒集团

 天津科学技术出版社

图书在版编目 (CIP) 数据

智能环境下基于音视频多模态融合的身份识别 / 吴迪著. —天津: 天津科学技术出版社, 2018. 3
ISBN 978-7-5576-4599-1

I. ①智… II. ①吴… III. ①图象识别 IV.
①TP391.413

中国版本图书馆CIP数据核字 (2018) 第038366号

责任编辑: 胡艳杰
责任印制: 兰 毅

天津出版传媒集团
天津科学技术出版社出版

出版人: 蔡 颢

天津市西康路 35 号 邮编 300051

电话 (022) 23332397

网址: www.tjkjcs.com.cn

新华书店经销

天津印艺通制版印刷有限责任公司印刷

开本 787×1092 1/16 印张 13.5 字数 200 000

2018 年 3 月 第 1 版第 1 次印刷

定价: 30.00 元

前 言

近年来,随着人们对安全要求的逐渐提高以及远程视频会议系统的快速发展,智能环境下基于生物特征的身份识别技术成为模式识别领域的研究热点,其在智能视觉物联网、公共安全、金融服务和视频会议系统等众多领域有着广泛的应用。受数据噪音和识别系统本身的限制,基于单一生物特征的身份识别系统所能达到的准确率是有限的,为此,研究人员提出利用视听信息融合身份识别来提高识别的准确率,受到了广泛的关注。但是目前基于视听信息融合的身份识别主要局限于理想环境下的单模态识别以及在现有融合方法上音视频特征的简单融合,对于复杂环境下单模态生物特征的有效提取、高精度高普适性识别算法的构造与音视频特征在不同融合层级最优融合算法的确定少有考虑。从人的视听觉认知机制出发,本书从特征提取、识别算法和融合规则三个方面对视听信息融合身份识别进行了研究,以便为智能环境下的视听信息识别提供可行的解决方案,本书的主要工作和创新点如下。

1. 实现了复杂背景下人脸特征和语音特征的精确高表征提取

针对人脸图像 DCT 特征系数的最优提取问题,本书提出了一种基于鉴别能力分析的 DCT 系数提取方法,在分析 DCT 系数鉴别能力值的基础上提取那些鉴别能力值较大的 DCT 系数作为特征。在分析头发几何特性和颜色特性的基础上,本书将人体的 Hair 特征应用于人脸识别,扩展了人脸特征的多样性。人脸的 DCT 系数对噪声敏感,识别精度不高,而纹理特征对光照和噪声的适应性强,目前成为人脸特征的研究热点。基于此,基于信息的显著性和深度学习技术,本书先后探讨基于局部定向模式的人脸特征提取方法。

针对传统语音参数 MFCC 受噪声影响较大而且只能反映语音静态特性的缺点,本书基于能有效反映人耳听觉特性的 Gammatone 滤波器,提取了 Gammatone 滤波倒谱系数,并基于滑动差分倒谱,提取了能反映语音动态特性的 Gammatone 滑动差分倒谱系数。

2. 提出了可有效解决“高维小样本问题”的人脸识别算法

目前,基于子空间分析的方法由于描述能力强、可分性好、计算简单等优点,成为人脸识别的主流算法,但常常面临“高维小样本问题”,导致人脸识别系统泛化能力较差。本书结合子空间分析方法和核思想,先后提出核相关权重鉴别分析算法和核鉴别局部保持投影算法,一方面解决了“高维小样本问题”,另一方面解决了传统子空间分析方法由于其线性本质所导致的在处理高度线性不可分对象时能力差的缺点。

3. 解决了说话人识别 GMM 模型的建模问题

GMM 模型是目前说话人识别的主流算法,并且在此基础上衍生了一系列说话人识别算法。针对由于训练语料较短而导致 GMM 模型参数训练不充分、识别性能下降的问题,本书通过引入因子分析技术,实现了一种自适应均值的 GMM 模型。i-vector 说话人识别系统是在 GMM 模型和因子分析技术基础上产生的目前国内外说话人识别研究前沿的主流系统,本书通过改进局部保持投影算法,实现了 i-vector 说话人识别系统中

i-vector 矢量的有效降维。

4. 建立了不同层次音视频特征的最优融合规则

本书以信息熵理论、概率密度方法和决策科学为指导，建立最优的匹配层融合规则 and 解决 D-S 证据理论的证据冲突问题。首先在分析现有证据冲突问题解决方法的基础上，提出基于群体决策和多准则选择融合的证据组合方法，有效解决证据冲突问题；其次为避免对匹配分数密度进行估计，本书将总错误概率 TER 引入到匹配层融合，通过 TER 来刻画匹配分数的分布，并将不确定度量融合方法引入到多特征融合识别；然后采用高斯密度求解加性融合中的最优权值，并将其用于逻辑回归排序层融合；最后针对匹配分数密度融合密度函数的求解，引入 FAR 和 FRR 以求解信任度函数，并基于三角模算子融合信任度函数，有效规避加性融合中权值的求解。

5. 研究了基于多特征融合的视频目标跟踪算法

智能环境下视频目标的鲁棒跟踪问题，一直是计算机视觉领域的研究热点，广泛应用在视频监控、机器人导航、人机交互等领域。针对复杂环境下视频目标跟踪精确度低的问题，本书提出了一种基于混合迭代无迹粒子滤波和关联系数自适应融合的目标跟踪算法和基于视频场景复杂度多特征融合算法，以提高跟踪的精确性和自适应性。针对复杂环境下目标跟踪鲁棒性差的问题，定义了一种特征相似性度量方法；其次在特征相似性的基础上提出了一种新的多特征融合算法，并且结合特征的相似性将跟踪场景切分成三种模式，在不同的模式背景下采用不同的融合算法，以提高跟踪的精确性和自适应性。

综上所述，本书的研究内容有效提高了计算机对复杂感知信息的理解能力和对异构信息的处理能力，进一步拓展了多生物特征融合身份识别的适用条件和应用范围，有效提高了智能环境下基于音视频多特征融合身份识别的鲁棒性和识别率，对推动我国人机交互技术的发展具有重要的意义。

在本书的编撰过程中，作者研读了大量文献，参考了国内外专家、学者在相关领域的研究成果。在此对他们表示感谢。特别是要感谢兰州理工大学曹洁教授、湖南工程学院罗毅平教授、黄中华教授、黄峰教授、唐勇奇教授、胡慧博士和万琴博士等人。

本书的出版得到了湖南省自然科学基金（16JJ6025）、湖南省高校创新平台项目（15K032、16K024）、湖南省教育厅重点项目（15A044、17A048）、湖南省普通高等学校教学改革研究项目（湘教通[2014]247 号）、湖南省 2011 计划“风电装备与电能变换协同创新中心”、湖南省高校科技创新团队支持计划资助（湘教通[2014]103 号）等项目支持。在此致以诚挚的谢意。

由于作者学术水平有限，书中难免存在诸多不妥甚至错误之处，殷切期盼广大读者批评指正。

作者

2017 年 12 月

目 录

第 1 章 概述	1
1.1 课题的背景和研究意义	1
1.2 智能环境简介	3
1.3 国内外研究现状	6
1.4 智能环境下基于视听信息融合身份识别的研究难点	13
1.5 本书主要工作和结构安排	15
第 2 章 智能环境下音视频融合的基本方法与进展	24
2.1 音视频信息融合的基本方法	25
2.2 基于任务的音视频融合方法比较	31
2.3 存在的问题和发展方向	38
第 3 章 高鉴别能力人脸和语音特征提取研究	50
3.1 引言	50
3.2 基于鉴别能力分析的人脸 DCT 特征提取	50
3.3 基于 Gammatone 滤波器的静动态语音听觉特征提取	60
3.4 本章小结	64
第 4 章 基于 DCT 系数和非线性子空间的人脸识别	67
4.1 引言	67
4.2 基于核相关权重鉴别分析的多特征融合人脸识别	67
4.3 基于核鉴别局部保持投影的多特征融合人脸识别	76
4.4 本章小结	85
第 5 章 基于局部定向模式与深度神经网络的人脸识别	87
5.1 引言	87
5.2 基于显著性局部定向模式和深度学习的鲁棒人脸识别	87
5.3 基于显著性差值局部定向模式和深度卷积网络的鲁棒人脸识别	95
5.4 本章小结	104
第 6 章 基于静动态听觉特征融合与模型补偿的说话人识别	106
6.1 引言	106
6.2 基于静动态听觉特征融合与自适应 GMM 的说话人识别	106
6.3 基于改进的局部保持投影的 i-vector 说话人识别	111
6.4 本章小结	115
第 7 章 基于音视频信息决策级融合的多模态身份识别	118
7.1 引言	118
7.2 基于群体决策和多准则选择融合的证据组合方法	118
7.3 基于多义度和三角模算子的加权冲突证据组合	132

7.4 基于三角模与均方欧氏距离的加权证据组合	136
7.5 基于证据信息散度的冲突证据表示	139
7.6 本章小结	142
第 8 章 基于音视频信息多层级融合的多模态身份识别	145
8.1 引言	145
8.2 基于 TER 和不确定度量的匹配层融合算法	145
8.3 基于高斯概率权值求解的逻辑回归排序层融合算法	149
8.4 基于信任度函数和三角模的融合算法	153
8.5 基于 TER 和特征关联自适应的融合算法	156
8.6 实验分析	159
8.7 本章小结	178
第 9 章 基于多特征融合的视频目标跟踪	182
9.1 引言	182
9.2 基于混合迭代无迹粒子滤波和关联系数自适应融合的视频目标跟踪	182
9.3 基于视觉场景复杂度多特征自适应融合的目标跟踪	195
9.4 本章小结	204
第 10 章 结论与展望	207

第1章 概述

1.1 课题的背景和研究意义

近年来，随着人们对信息安全要求的逐渐提高以及远程视频会议系统的快速发展，智能环境下基于生物特征的身份识别技术成为模式识别和人机交互领域的研究热点。它在智能视觉物联网、以人为中心的普适计算、公共安全、金融服务和视频会议系统等众多领域有着广泛的应用，例如，在智能视觉物联网中，利用各类图像获取传感器，包括监控摄像机、手机、数码相机，获取人、车、物图像或视频，并采用智能分析技术对视觉信息进行处理，为后续利用提供支撑。在在以人为中心的普适计算中，为了提供上下文的感知服务，计算机需要具备感知外部环境的能力，尤其是对说话人身份的准确识别。在公共安全领域，根据犯罪现场留下的指纹信息与数据库中的指纹信息进行比对可快速锁定犯罪嫌疑人，在出入境口，根据人脸信息和指纹信息可有效鉴别访客的身份。在金融服务领域，储户可根据电子签名和虹膜信息不需携带银行卡也无需密码即可办理相关银行业务。在视频会议系统中，根据说话人的语音信息和人脸图像，可快速识别说话人的有效身份。

语音及视觉是人们相互交流的重要手段，也是人机交互过程中最为直接的方式。目前，基于音频信息的说话人识别技术和基于计算机视觉信息的人脸识别技术已经成为解决身份识别问题的主要手段。基于音频信息的说话人识别技术不受视觉杂波的影响，在说话人被其他物体完全遮挡的情况下仍然可以实现有效识别，但背景噪声与房间混响则制约着说话人识别的性能；基于计算机视频信息的人脸识别技术具有强友好性、直观性和非侵害性，不受音频噪声的影响，但是人脸识别方法容易受到视频遮挡、光照、面部表情和姿态变化等因素的影响。随着社会的发展进步，人们对安全的要求越来越高，对身份鉴定系统准确性及安全性的要求也日益提高，特别是在机场、银行、军队等关系到国家和社会安全的特定场所。这时，这种基于单模态信息进行分析的方法已经很难满足目前复杂环境及干扰情况下身份识别技术对系统整体性能精度的需求，已成为人工智能领域急需解决的关键问题之一。

随着多源信息融合理论的发展，研究人员借鉴人类大脑对听觉和视觉信息进行融合处理的方法，提出将音频信息和视频信息融合在一起，利用两者的时空相关性和互补性进行异类信息融合识别，即所谓的“视听信息融合身份识别”，如图 1.1 所示，较好的克服了单模态情况下的不足，目前成为提高身份识别系统精度与鲁棒性的重要手段。

“视听信息融合身份识别”将数据融合方法用于身份鉴别，结合多种生理和行为特征进行身份鉴别，不仅提高了识别的准确性，扩展了识别系统的覆盖范围，而且能够降低系统的风险，增强系统的实用性。由于其重要的理论研究价值和市场应用前景，吸引了国内外越来越多的研究组从事这方面的研究，目前已成为模式识别领域的一个重要研究方向，也是未来生物特征应用领域的必然趋势。

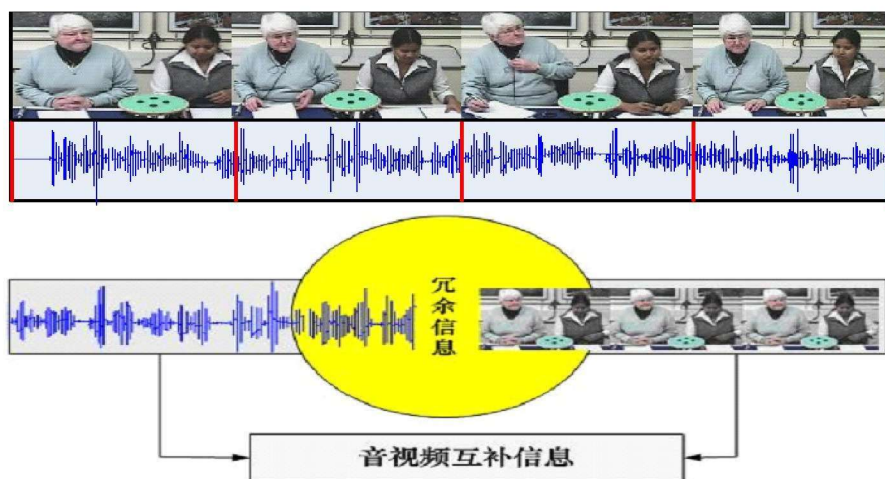


图 1.1 音视频信息相关性、互补性示意图

目前，“视听信息融合身份识别”已经受到发达国家的广泛关注，美国政府自从 2001 年“9·11”恐怖袭击事件后，成为使用该项技术解决安全问题的主要倡导者，美国联邦调查局投资近 10 亿美元，正在创建一个包括人体 DNA、语音、指纹等生物特征的新型多模态生物识别数据库，以便将其用于个人身份识别来保证国家的公共安全。德国联邦政府大力支持多生物特征识别技术，其市场份额由原来的 1200 万欧元增长到 3.77 亿欧元。澳大利亚政府将生物特征认证技术引入到外国人入境的 Smartgate 系统，用于加强监管外国人入境问题。就国内而言，政府对视听信息融合身份识别的开发与研究相当重视，北京大学成立了视觉与听觉信息处理国家重点实验室，在机器视觉与听觉信息处理领域开展了大量具有多学科交叉性质的基础与应用基础研究。2013 年，国家自然科学基金委发布《视听觉信息的认知计算重大研究计划 2013 年度项目指南》，指出以社会、经济和国家安全等领域中与人类视听觉信息相关的图像、语音和文本（语言）的认知机制和计算模型为研究对象，力争在视听觉信息处理的基础理论研究方面取得重要进展，包括视听觉信息早期处理（视听觉信息预处理与增强）、中期处理（降维、特征选择与提取）和后期处理（分类、识别与理解）等不同阶段的认知机制与可计算模型。

音视频信息融合技术的应用领域很多，图 1.2 展现了音视频融合技术当前的几个典型应用平台，主要包括智能会议场景、自然的人机交互场景、智能家居以及智能驾驶系统。智能会议场景是音视频信息融合技术最广泛的应用领域，其主要特点就是背景复杂、噪声干扰多、同时具有多个目标。针对该场景，目前主要的研究热点就是采用远场麦克风阵列和摄像机来定位说话人的位置并对其轨迹进行精确有效的跟踪，并进行有效的后续处理，例如：语音增强、语音识别、行为分析、身份识别和情感识别等。人机自然交互系统是音视频信息融合技术应用的另外一个热点领域，其主要特点就是场景中的目标具有合作性，因此对融合系统与传感器阵列没有太苛刻的要求，通常情况下易于采用自适应系统进行处理。智能家居和残疾人辅助生活系统是音视频融合技术的一个崭新的应用领域，其主要目的就是对特定场合下的特定目标进行状态监控并进行有效的交互处理。随着车联网技术的发展，基于音视频信息融合技术的汽车辅助驾驶系统开始受到汽

车行业的青睐，认为融合处理有效提升了系统的整体性能。在无人驾驶车辆技术中，音视频融合技术可在一定程度上解决安全问题，实现实现复杂交通环境下高性能长距离自主行驶。

尽管在实际生活中，音视频融合技术还没有被广泛应用，但是随着信息技术和物联网技术的发展以及“智慧地球”“感知中国”“平安政府”等一系列智能化项目的推进，该技术具有很好的市场前景和发展潜力，同时，也有很多关键技术需要进一步的深入研究和发 展。开展音视频融合身份识别技术的研究，不仅可以提高身份识别系统的精确性和鲁棒性，而且对推动我国人机交互技术的发展具有重要的意义。其研究成果不仅为智能控制、通信系统、金融经济和社会生活方式产生巨大的影响，而且将改变人们的生活和工作方式，产生相应的经济和社会效益，使人与计算机的交互如同人与人交流那样自然。



图 1.2 音视频融合系统应用平台

1.2 智能环境简介

所谓智能环境即通过安装在会议室四角的摄像机和安装在墙面和桌子上的麦克风阵列对会议做全程的跟踪，并由此对与会人员进行辨识。为了配合相关研究，国外的科研人员专门建立了许多智能的会议室环境，在这些会议室中录制了大量的视听会议，主要包含意大利特伦托大学特伦蒂诺迪研究所的 ITC-irst 环境、德国卡尔斯鲁厄大学人机

交互实验室的 UKA 环境、希腊雅典社会教育信息研究中心的 AIT 环境、美国 IBM 公司 T.J.Watson 研究中心的 IBM 环境和西班牙巴萨罗那 Politecnica 大学的 UPC 环境，如图 1.3 所示。



图 1.3 不同智能会议室环境

上述 5 个智能环境都是由欧洲的 GHIL (Computers in the Human Interaction, CHIL) 项目投资兴建，目前，由欧盟投资兴建的 AMI (Augmented Multimodal Interaction, AMI) 智能环境成为国际上进行智能环境下音视频联合身份识别的标准实验平台，它是一个可以容纳 4 人的小型会议室环境，如图 1.4 所示。在会议室中间，有一张会客桌和四把椅子，桌子上放置有一个位置固定的圆盘。圆盘的底部安装有一圈麦克风阵列，可以捕捉参会人的语音信息。圆盘的侧面每隔 90 度安装有一台近距离摄像机，每台摄像机镜头对准一把椅子，专门用来拍摄参会人的面部特征（如图 1.5 下面一排图片所示）。在会议室的前部，有一块白色黑板和一张幻灯片投影屏幕，可以供参会人演讲或者结合 PPT 进行讲解。在会议室侧面和后面的墙壁上都安装有远距离摄像机和麦克风组合，可以捕捉到整个会议室的人员活动情况（如图 1.5 上面一排图片所示）。当参会人离开座位走向讲台或者站在讲台上进行发言时，这三台远距离摄像机可以完成对演讲过程的跟拍，保证参会人不会从画面中丢失。

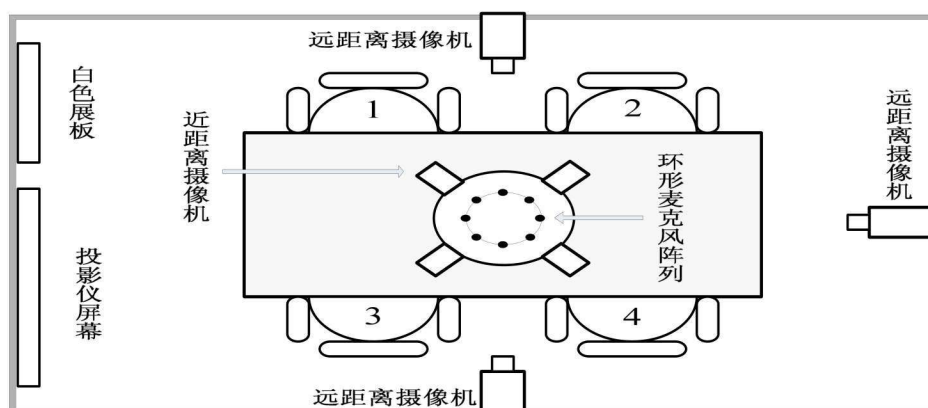


图 1.4 AMI 智能会议室环境平面图



图 1.5 AMI 智能环境摄像机所采集的画面

当前，美国和欧洲的许多实验室都在进行智能环境的研究，由瑞士政府资助的 IM2 项目目的是研究音视频多模环境下说话人的互动行为，欧盟资助的 CHIL 项目希望创建一个计算机集群，进而以一种非入侵的方式辅助人们更好地进行人与人的交流，同样 AMI 项目也是由欧盟资助，研究的重点是会议或会议类型的商务谈判。其中 AMI 项目所录制的语料库是迄今信息量最为庞大、功能最为全面的音视频语料库。研究人员对于智能环境的研究，也从最初的单模态说话人跟踪与识别，发展到现在的音视频多模态融合跟踪与识别、说话人手势检测、语音事件检测等，具体如图 1.6 所示。

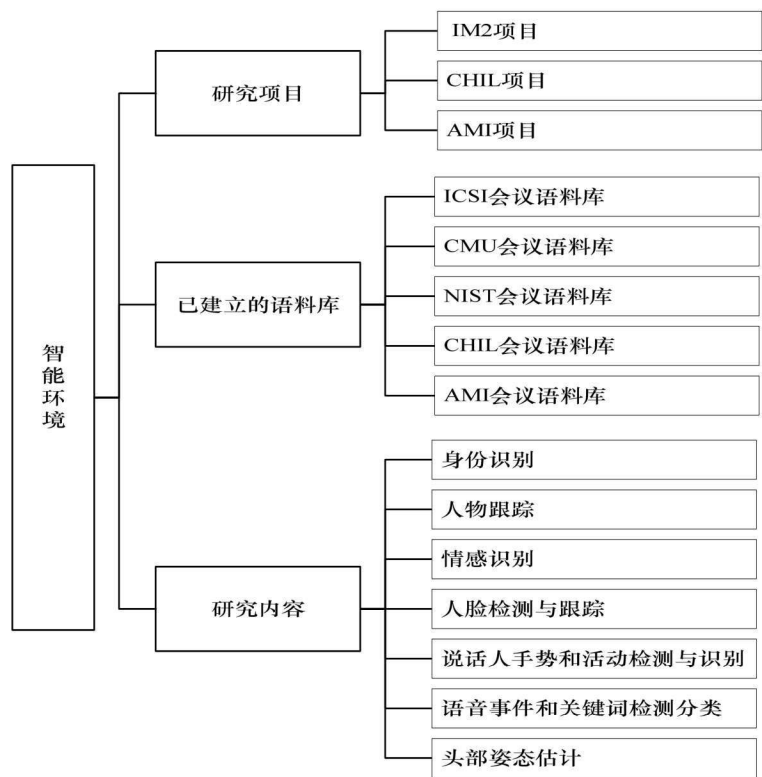


图 1.6 智能环境的研究项目与内容

1.3 国内外研究现状

基于视听信息多特征融合的身份识别技术是一个多学科交叉的前沿性研究课题，涉及现代信号处理、计算机视觉以及多源信息融合技术等多个学科领域。在过去的几十年里，以提高系统识别精度和鲁棒性为目的，广大研究人员从特征提取、识别算法和融合规则等方面进行了深入的研究，当智能环境的概念被提出后，人们又在此特定场景下开展了智能环境下基于生物特征的身份识别研究，取得了一定的进展。下面将从多生物特征融合身份识别和智能环境下身份识别两个方面介绍目前国内外的研究现状，并就其存在的问题进行分析。

1.3.1 多生物特征融合身份识别技术概述

多生物特征识别技术是指对人体的多个生物特征进行采集处理，接着进行生物特征的检测和相关特征的提取和数字化处理，判断和识别，然后在相应的融合层次上进行融合处理，最后得出综合的识别结果。自从多生物特征识别技术在身份识别系统中得到深入的研究和应用后，研究人员对于生物特征识别的研究进入了新的时代。下面分别从多生物特征融合身份识别系统的分类、特征提取、识别模型以及融合规则四个方面介绍多生物特征融合身份识别技术的发展情况。

1. 多生物特征融合身份识别系统简介

按照融合在哪个环节之后进行，多生物特征识别技术可以分为五类：数据层融合、

特征层融合、匹配层融合、排序层融合以及决策层融合；根据融合数据的来源，多生物特征识别技术可以分为五类：多采集仪系统、多单元系统、多样本系统、多算法系统和多模态系统。图 1.7 是生物特征识别系统的分类。

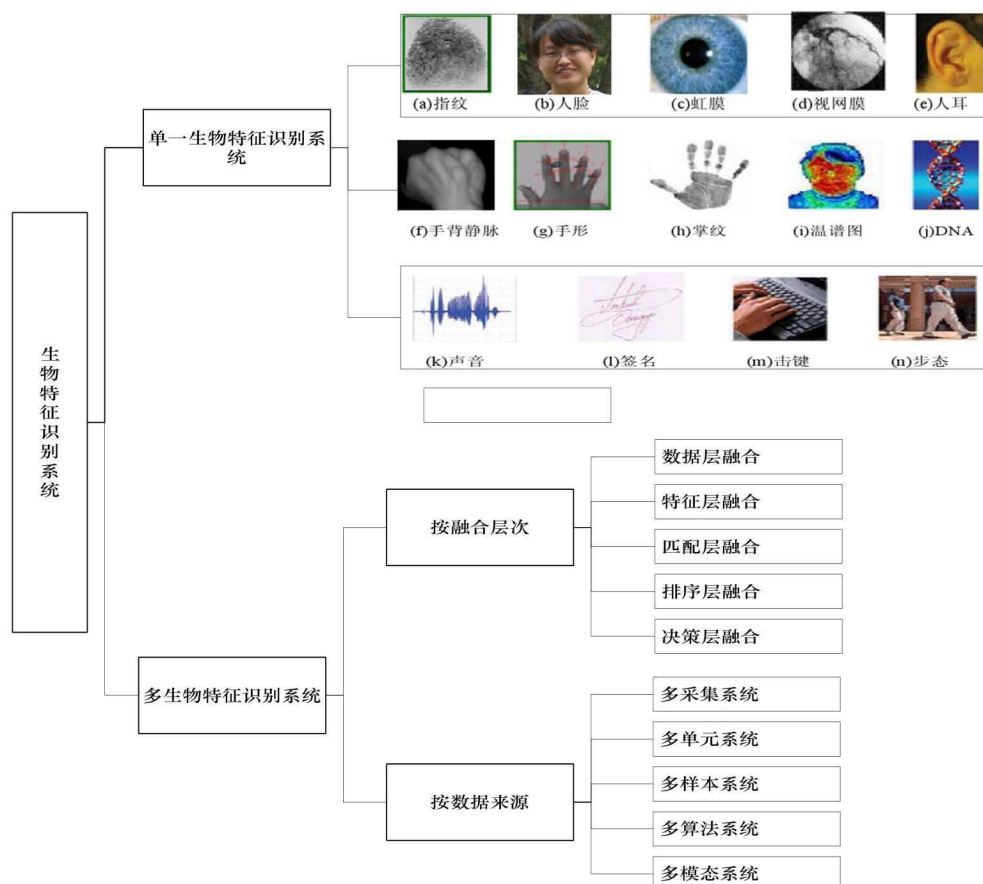


图 1.7 生物特征识别系统分类

数据层融合是在特征提取前对数据进行融合，也称为传感器级融合或图像级融合。特征层融合是指输入数据经过前端处理后，分别得到生物特征的特征描述向量，然后经过特征融合处理，将多个低维的特征向量融合形成更高维的联合特征向量参数。匹配层融合是将单项生物特征分别进行独立的匹配处理，得到各个匹配分数，通过融合算法将各匹配分数进行综合，得到最终决策结果。排序层融合是介于匹配层和决策层之间的一种融合算法，其对分类器输出匹配对象的可能进行排列，然后对排列进行融合。决策层融合是在最高层上进行的融合，其输入是单项生物特征识别的逻辑输出，然后将各逻辑输出按照一定逻辑进行判断得到最终结果。

多采集仪系统时利用数据间的相关性和互补性通过不同的采集设备采集相同生物特征，然后在不同层次进行融合。多单元系统是对同一特征的不同单元进行融合，例如融合左眼和右眼的虹膜识别系统。多样本系统时多次采集同一生物特征的多个样本，对多个样本进行融合，例如同一个指纹的多幅图像，同一语言的多个样本等。多算法系统时

采用相同的生物特征、相同的传感器，但采用不同的算法进行特征提取和匹配。多模态系统时通过对不同类型的生物特征进行融合识别，从而使最终的生物识别系统能够取得最优的性能。

2. 人脸特征和语音特征提取的研究现状

一般来说，人脸描述特征可以分为全局特征和局部特征两大类。全局特征是指其特征向量的每一维都包含了人脸图像上所有部分（甚至所有像素）的信息，因此反映的是人脸的整体属性。在人脸识别的研究中，两种类型的全局特征被广泛使用：①基于子空间的方法，比如主成分分析、线性判别分析、独立分量分析等；②基于空频变换的方法，比如傅里叶变换、离散余弦变换等。与全局特征不同，局部特征的每一维都只对应人脸图像上的一个局部区域，因此侧重于提取人脸的细节特征。近些年来，局部特征被认为对人脸的光照、表情和遮挡等变化不敏感，因此也被越来越多地用于人脸表示。比较常用的局部特征包括局部二值模式、Gabor 小波、局部非负矩阵分解等。

语音特征提取是指从语音信号波形中获取一组能够描述语音信号特征参数的过程，由于语音信号由多种因素综合产生，至今也没有有效的方法将语义内容信息和说话人个人特征信息完全区分开来。所以，现有的特征提取方法只能对这些参数做适当地选择与分析，尽量排除语音内容变化引起的干扰。最早应用于语音识别的特征是 Pruzansky 利用滤波器组从语音频谱中提取的滤波器组系数。德国仪器公司的 Doddington 采用共振峰分析取代滤波器组进行特征提取，与此同时，语音识别中一个重要的问题，说话人自身特征变异问题也成为说话人识别的研究重点之一，并对说话人模型表示产生了巨大影响。Furui 提出了基于倒谱系数和差分特征的电话信道说话人识别方法，极大地提高了识别系统的鲁棒性。此后不久，一系列基于短时谱分析的倒谱特征相继被提出，例如：线性倒谱系数、Mel 频率倒谱系数和感知线性预测系数等，目前为止，这些倒谱特征仍是语音识别中提取短时谱信息的主要手段。

3. 人脸识别算法和语音识别算法的研究现状

目前的人脸识别算法主要包含五种方法：基于几何特征的方法、基于神经网络的方法、基于弹性图匹配的方法、基于三维模型的方法和基于子空间的方法。

基于几何特征的方法是提取人脸的几何特征如眼睛、眉毛、鼻子等的形状和结构关系作为特征进行识别，基于几何特征的人脸识别方法具有直观、识别速度快、对光照有一定鲁棒性等优点，但由于受面部器官轮廓和姿态变化影响较大，且忽略了整个图像的细节信息，近年来少有发展，不过可以作为其他特征的有益补充。人工神经网络在自学习、自组织和容错等方面具有较强的能力，并且在学习的过程中具有自动提取特征的能力，可以通过反复学习将人脸识别中的规律和规则隐性的表示出来，因此在特征提取后进行分类识别时人工神经网络具有得天独厚的优势。基于弹性图匹配的方法考虑了人脸的局部细节，保留了人脸的空间分布信息，由于在一定程度上能够容忍从三维到二维投影所引起的变形，故可很好的解决姿态和表情的变化问题，但由于没有考虑人脸图像的尺寸变化以及图像在平面内的旋转，因而当人脸大小、位置、光照和表情变化明显时，其性能会大幅度下降。人脸本身是三维物体，用三维模型能够更好地表征人脸，也能够更好的处理光照和姿态两大棘手问题，但由于计算复杂度高，采集不易等缺点在实际应用中受到一定程度的限制。子空间分析在于通过一定的性能目标函数来寻找使该目标函

数达到最优的空间变换，利用该空间变换实现从高维图像空间到低维人脸空间的映射，由于其描述能力强、可分性强、计算简单等优点，近年来，逐渐成为人脸识别的主流算法，具体如表 1.1 所示。

表 1.1 人脸识别算法概括

人脸识别方法	代表性算法
基于几何特征的方法	①bledsoe: 人脸特征点的间距、比列为参数 ②Brunelli: 眉毛、眼睛、脸宽等 35 个参数
基于神经网络的方法	①Kohonen: 利用网络的联想能力 ②Lee: 模糊反向传播网络 (BPN) ③Ranganath: 径向基函数网络 (RBFN)
基于弹性图匹配的方法	①Buhmann: 动态链接结构模型 ②Wiskott: 弹性束图匹配 (EBGM)
基于三维模型的方法	①Blanz: 三维形变模型 (3DMM) ②Nese: 基于三维曲率的表示方法
基于子空间的方法	①主成分分析 (PCA)
	②线性鉴别分析 (LDA)
	③独立成分分析 (ICA)
	④典型相关分析 (CCA)
	①核主成分分析 (KPCA)
核方法	②核 Fisher 鉴别分析 (KFDA)
	③核独立成分分析 (KICA)
	④核典型相关分析 (KCCA)
	①等距映射 (Isomap)
流形学习	②局部线性嵌套 (LLE)

对于说话人识别模型的发展，早期使用粗糙且性能较差的统计模板作为识别模型。20 世纪 80 年代，动态时间规整技术和矢量量化技术被引入到说话人识别中，并使说话人识别系统的性能进一步提高。与此同时，人工神经网络和 HMM 在语音识别领域也取得了巨大的成功，由于 HMM 能够很好地描述语音信号的时变性和短时平稳性，同时具有把声学—语言学到句法等统计知识全部集成在一个框架的优点，成为文本相关说话人识别的主要建模方法。随后 Rose 提出了 GMM，它是一个鲁棒的参数化模型，识别率与总的混合数有很强的关系，但与状态数无关，即意味着不同时间的转移信息对说话人识别系统是没有用的，因此，GMM 建模在说话人识别研究中得到了越来越多的重视。20 世纪 90 年代中期，Vapnik 等人提出了 SVM 方法，由于它在解决小样本、非线性和高维模式识别中表现出许多特有的优势，从而迅速成为机器学习领域的研究热点。鲁棒性问题是说话人识别走向实用化的核心问题之一，因此各种针对说话人领域的鲁棒技术相继被提出，如基于 SVM 模型的扰动属性投影技术和基于 GMM 的因子分析技术。这些模型域鲁棒技术都充分利用了模型中说话人个性信息与信道信息在高维空间中的分布特性，提高了说话人识别系统的鲁棒性。

4. 音视频异类信息多层级融合的研究现状

在获得音频特征和视频特征后，选择在哪个层次上进行融合就是融合策略问题，也是研究融合规则时非常重要的问题，到目前为止，还没有哪种融合规则能够适用于所有情况，因此，融合规则的选择就显得尤为重要。匹配层融合屏蔽了各种生物特征数据底

层的多样性和识别过程的复杂性，同时又保存了特征相似性的度量；D-S 证据理论属于决策层融合，在区分不确定与未知过程中显示出了极大的优势，适合于无先验信息的融合。两者成为目前研究最多的融合规则。

匹配层融合规则是将不同特征对应分类模型输出的相似度作为匹配分数进行融合，主要分为三类：基于归一化的融合方法、基于密度的融合方法和基于分类器的融合方法。

(1) 基于归一化的融合方法先将匹配分数进行归一化处理，然后利用一些固定的融合规则对归一化后的匹配分数进行融合处理从而得到一个新的匹配分数，最后做出最终决策。基于归一化的融合方法有两个需要考虑的因素：一是归一化函数，二是融合规则。目前常用的归一化函数主要是 Min-Max、Z-Score、Tanh 和 Sigmoid 等。对于融合规则，文献提出了归一化融合的一般理论框架，并在此框架下推导出了 Sum、Product、Min、Max 和 Media 五种基本的融合规则；文献提出了加权融合方法，权值根据各个简单分类器的 EER 来计算；文献基于均值和方差提出了降低大分数影响的归一化方法。

(2) 基于密度融合的方法是首先将匹配分数转换成后验概率，然后根据贝叶斯判决准则做出最后的决策。Duda 等指出匹配分数的概率密度可以通过参数化或者非参数化方法求得。Snelick 等假设匹配分数服从高斯分布，采用一种参数化方法来求取匹配分数的条件概率密度。Jain 等提出使用基于 Parzen 窗的非参数化方法求取概率密度。Dass 等提出了一种基于通用概率密度的匹配分数融合方法。Nandakumar 等使用有限混合高斯模型求取匹配分数密度进行多生物特征识别融合，取得了较好的效果。

(3) 基于分类器的融合方法是将多个匹配分数作为一个特征向量来看待，将匹配分数融合的问题巧妙的转换成为一个特征向量的分类问题。刘红毅等研究了基于改进 ENN 算法的指纹和语音特征的多生物特征融合算法，主要思想是先将匹配分数降维，然后进行分类。Wang 等研究了基于支持向量机的人脸和虹膜的多生物特征识别融合问题，通过遍历参数来寻找最优的支持向量机参数。Kumar 等研究了基于粒子群优化算法，融合了人脸和语音特征，动态选用最优融合策略来满足不同的安全需求。Tronci 等提出了匹配分数融合的理想选择器，并在此基础上提出了一种匹配分数动态选择方法。

D-S 证据理论研究的热点是证据冲突问题，目前对于证据冲突问题的解决方法可以分为三类：一是针对传统 D-S 证据理论模型框架进行修改；二是修改 Dempster 合成规则的方法，主要解决如何将冲突重新分配和管理的问题；三是基于修改原始证据源的方法，即首先对冲突证据进行预处理，然后再用 Dempster 规则合成证据。

(1) 针对传统 D-S 证据理论模型的框架进行修改主要是 Jean Dezert 于 2002 年提出的 DS_mT 理论（似是而非理论），它是传统 D-S 证据理论的扩展，主要是处理不确定、高冲突和不精确的证据源的融合问题。

(2) 基于修改 Dempster 规则的方法又可分为证据可靠和不可靠两种情况，但都是考虑如何对证据间的冲突进行分配，包括冲突分配到识别框架幂集的哪些命题元素及冲突在各元素上分配的比例问题。Yager 认为所有冲突均不能提供有用信息，因此把证据冲突全部赋予未知项 $m(\Theta)$ ，该方法将证据间的冲突全部否定，过于保守，而且会使合成后证据的不确定性增大，不符合证据推理的目的。孙全等假设证据具有同样的可信度，认为虽然证据间存在着冲突，但也是部分可用的，通过计算证据中两两冲突的平均值定

义了证据集的有效性系数,将总冲突 k 按一定比例分配给各个合成命题。邓勇等进一步改进了上述方法,以证据不可靠为前提,引入证据距离定义各证据的可信度,先对两两证据的支持度代数平均并与证据间的相对可信度线性加权,得到局部冲突的可信度,再与各局部冲突线性加权。张山鹰将证据的局部冲突指派给产生冲突的焦元中 BPA 值较大的焦元(与证据可靠性无关),称之为吸收法,其理由是受扰动的焦元一般比未受扰动的焦元的 BPA 值小,把冲突值分配给 BPA 值较大的焦元是合理的,这种局部冲突的分配策略虽然有些偏激,但在应用上简单,不过失去了多源数据融合的可交换性,其结果与组合顺序有关。

(3) 基于修改原始证据源的方法则认为 Dempster 合成规则本身没有错,但要求所有参与合成的证据具有相同的重要程度,若直接利用该规则,首先要考虑参与合成的各证据的重要程度,所以,在证据高度冲突时,应该首先对冲突证据进行预处理,然后再使用 Dempster 组合规则。Murphy 将参与合成证据的 BPA 值进行平均,得到新的证据模型后,再利用 Dempster 规则组合证据,由于只是对证据进行简单平均,没有考虑到参与合成各证据的重要度,因此结果并不理想。杨善林等根据证据源存在的主观差异,提出了一种基于技术同步和信息不对称的动态合成法则,该方法忽略了同一时间序列专家群体中也应具有权重分配,一定程度上将问题简单化。Fan et al 通过建立和分析证据关系矩阵来识别异常证据,以减少它们对合成结果的影响。Xu et al 根据证据序列的灰色关系系数矩阵,提出一种自动识别和分配冲突的预处理方法。

5. 现有多生物特征融合识别技术存在的问题

尽管国内外对于多生物特征认证系统的研究取得了一定成果,但多生物特征融合识别技术仍处于研究探索阶段,技术还不成熟,仍存在许多亟待解决的问题,综合来说包括为以下几点。

(1) 在人脸特征提取中,尽管全局特征和局部特征都得到了广泛的应用,但是在人脸识别研究中一直存在关于全局特征和局部特征孰轻孰重的争论;人脸有着丰富的表情变化,因此严格的特征提取方法并不奏效,有时在某些数据样本上这些特征提取方法或许是有效的,但不能保证对整个数据库的所有样本都最佳。

语音特征提取在使用倒谱特征后,语音识别系统性能有了大幅度提高,大大促进了自动语音识别技术性能的提升。然而大量研究表明,在低信噪比的噪音环境下,特别是在非平稳噪音环境下,自动语音识别技术的性能与人类的听觉能力相比,尚有相当大的差距。究其原因,声学特征的鲁棒性较差是导致这种情况的关键因素之一。因为目前的语音特征存在着自身的缺陷:一是容易受环境噪声和信息畸变的影响,二是携带信息量有限,目前大多数语音特征只能反映语音的静态特性,无法充分描述说话人的个性信息。研究表明,仿真人耳的听觉机理有助于提升特征的鲁棒性,然而这方面的研究工作尚不完善,听觉系统在鲁棒性方面的相关机理还未能得到充分地挖掘。

(2) 现有的人脸识别算法虽然能够在一定程度上可以精确识别人脸,但是人脸识别技术也面临着许多困难和挑战。首先,由于人脸图像的高维非线性流形分布,这样将高维数据通过线性映射投影到低维空间会因无法保持人脸的非凸特性而不能很好地区分个体,从而使识别算法的应用受到极大的限制;其次由于人脸样本向量的维度远远大于样本数目,此时人脸识别系统很难具备良好的泛化能力,即目前人脸识别算法面临的