中国自动化学会青年工作委员会系列丛书

INGYONG MO SHI SHIBIE JISHU DAO L

应用模式识别技术导论

人脸识别与语音识别

REN LIAN SHIBIE YU YU YIN SHIBIE

苏剑波 徐 波 编著

上海交通大学出版社

YING YONG MO SHI SHIBIE JI SHU DAO LUN REN LIAN SHIBIE YU YU YIN SHIBIE

YING YONG MO SHI SHIBIE JI SHU DAO LUN REN LIAN SHIBIE YU YU YIN SHIBIE

YING YONG MO SHI SHIBIE JI SHU DAO LUN REN LIAN SHIBIE YU YU YIN SHIBIE

自动化理论、技术及应用丛书

应用模式识别技术导论

——人脸识别与语音识别

苏剑波　徐 波　　编著

上海交通大学出版社

内 容 提 要

本书是中国自动化学会青年工作委员会组织编写的自动化理论、技术及应用丛书之一，内容包括目前模式识别领域研究的两大热点技术——人脸识别和语音识别。全书分上下两篇，分别对人脸识别和语音识别这两个研究领域的基本问题、研究思路和方法、经典的算法和技术全方位地做了深入系统的介绍，突出反映了这两个研究领域国内外最新的研究进展和成果，特别对一些尚未解决的问题给出了评点。

本书可作为高等学校有关专业的研究生、高年级本科生、研究院所和有关单位广大科技工作者和工程技术人员的参考书。

图书在版编目(CIP)数据

应用模式识别技术导论：人脸识别与语音识别/苏剑波，徐波编著. —上海：上海交通大学出版社，2001
ISBN 7-313-02665-X

Ⅰ.应… Ⅱ.①苏…②徐… Ⅲ.模式识别
Ⅳ.TP391.4

中国版本图书馆 CIP 数据核字(2001)第 14706 号

应用模式识别技术导论
——人脸识别与语音识别
苏剑波 徐波 编著
上海交通大学出版社出版发行
(上海市番禺路 877 号 邮政编码 200030)
电话：64071208 出版人：张天蔚
常熟市印刷二厂印刷 全国新华书店经销
开本：890mm×1240mm 1/32 印张：7.625 字数：218 千字
2001 年 5 月第 1 版 2001 年 5 月第 1 次印刷
印数：1～550
ISBN 7-313-02665-X/TP·453 定价：15.00 元

前　言

四年前，中国自动化学会青年工作委员会即形成计划，编著出版一套自动化理论、技术及应用丛书，目的是向国内的青年自动化工作者和高年级本科生、硕士生、博士生有计划、有步骤地展示国内外自动控制和自动化领域各个研究方向最新的研究动态和研究成果，同时也把所存在的问题全面地展现出来，以便于青年自动化工作者和研究生们能很快地了解自动化发展的概貌。希望通过这套丛书，不仅能给热爱自动化事业的青年人以鼓舞，而且能帮助青年自动化工作者为自己选择一个能持续发展的研究方向，便于研究生们很快地掌握各个研究方向的前沿并比较容易地选择自己的研究方向，同时也希望有更多的本科生了解自动化领域的勃勃生机，从而为将来从事自动化领域的研究和技术开发打下扎实的基础。

由中国自动化学会、中国自动化学会青年工作委员会主办的中国自动化学会青年学术年会已成功地举办多届，已逐渐成为青年自动化工作者和自动化学科青年学子一年一度的盛会。很自然地，青年工作委员会就设想在每年的青年年会举办之际，由主办单位牵头，组织编写出版这套丛书中的一本，因此从１９９９年的第１４届青年学术年会开始，由北方工业大学和中科院自动化所组织国内十数家大学和研究所，联合编著出版了本丛书的第一本——《先进机器人和集成制造技术》，以小专题的形式展现了目前我国在国家８６３自动化领域的两大主题——机器人和ＣＩＭＳ的最新研究成果。本书为此丛书的第二本。

中国自动化学会第１５届青年学术年会于２０００年７月在上海召开，由中国自动化学会、中国自动化学会青年工作委员会主办，上海交通大学承办，上海自动化学会协办。因此本系列丛书的第二本

就由上海交通大学组织编写。

本书的主题为模式识别，内容包括人脸识别和语音识别这两个目前在国内外都处于热点中的研究方向。全书分上下两篇，上篇为人脸识别，包括7章，由上海交通大学苏剑波撰写；下篇为语音识别，包括8章，由中科院自动化所国家模式识别实验室徐波撰写。上海交通大学的黄福珍也参与了第一部分的撰稿工作。全书由苏剑波统稿。

由于本书由相对独立的上下两篇组成，内容涉及两个学科的前沿和全方位的研究，作者学识有限，再加上本书出版比较仓促，内容上会有不足之处；书中所述观点均是作者个人看法，也会有片面之处，而且对这种形式的丛书的编著经验还不足，缺点和差错在所难免，恳请广大读者给予批评指正。

新时代赋予了自动化学科无限机遇，而IT时代的到来又使自动化学科受到前所未有的冲击。这是一个转折期，正是青年自动化人只争朝夕，大显身手的好机会。让我们携起手来，继承和发扬老一辈所开创的美好事业，脚踏实地，兢兢业业，为新世纪自动化事业的繁荣和中国自动化事业的辉煌贡献自己的聪明才智。

作 者

2001年4月

于上海交通大学

目　录

上篇　人脸识别

下篇 语音识别

上篇　人脸识别

第 1 章　人脸识别概述

人脸因人而异，绝无相同，即使一对双胞胎，其面部也一定存在着某方面的差异。虽然人类在表情、年龄或发型等发生巨大变化的情况下，可以毫不困难地由脸而检测和识别出某一个人，但要建立一个能够完全自动进行人脸识别的系统却是非常困难的，它牵涉到模式识别、图像处理、计算机视觉、生理学、心理学以及认知科学等方面的诸多知识，并与基于其他生物特征的身份鉴别方法以及计算机人机感知交互领域都有密切联系。与指纹、视网膜、虹膜、基因、掌形等其他人体生物特征识别系统相比，人脸识别系统更加直接、友好，使用者无任何心理障碍，并且通过人脸的表情/姿态分析，还能获得其他识别系统难以得到的一些信息。

20 世纪 90 年代以来，随着需要的剧增，人脸识别技术成为一个热门的研究话题。虽然在这方面的研究已经取得了一些可喜的成果，但在实际应用中仍面临着许多严峻的问题，人脸的非刚体性，表情、姿态、发型以及化妆的多样性都给正确识别带来了困难。要让计算机像人一样方便准确地识别出大量的人脸，尚需不同学科研究领域的科学家共同作出不懈的努力。

1.1　人脸识别的研究内容

人脸识别（Face Recognition）一般可描述为：给定一静止或动态图像，利用已有的人脸数据库来确认图像中的一个或多个人。从广义上讲，其研究内容包括以下五个方面[63,86]：

（1）人脸检测（Face Detection）：即从各种不同的场景中检测出人脸的存在并确定其位置。这一任务主要受光照、噪声、头部倾斜度

以及各种遮挡的影响。

（2）人脸表征（Face Representation）：即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）、固定特征模板、特征脸、云纹图等。

（3）人脸鉴别（Face Identification）：即通常所说的人脸识别，就是将待识别的人脸与数据库中的已知人脸比较，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与匹配策略。

（4）表情分析（Facial Expression Analysis）：即对待识别人脸的表情进行分析，并对其加以分类。

（5）物理分类（Physical Classification）：即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。

本篇主要介绍狭义的人脸识别方法，不涉及表情识别和物理分类方面。一个人脸自动识别系统包括三个主要技术环节，如图 1-1 所示：

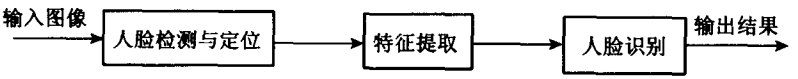

图 1-1 人脸自动识别系统构成

首先是人脸检测与定位，即检测图像中有没有人脸，若有，将其从背景中分割出来，并确定其在图像中的位置。在某些场合，拍摄图像的条件可以控制，比如警察拍罪犯的照片时要他们将脸的某一部分靠近标尺，这时人脸的定位很简单。普通证件照片上的头部占据了照片中央的大部分地方，定位也比较容易。在另一些情况下，人脸在图像中的位置预先是未知的，比如在一些复杂背景中拍摄的照片，这时人脸的检测与定位将受以下因素的影响：① 人脸在图像中的位置、旋转角度和尺度不固定；② 发型和化妆会遮盖某些特征；③ 图像中出现的噪声。

其次是特征提取。特征提取之前一般需要做几何归一化和灰度归一化的工作。其中前者是指根据人脸定位结果将图像中的人脸变化到同一位置和大小；后者是指对图像进行光照补偿等处理，以克服光照

变化的影响。具体的特征形式随识别方法的不同而不同。比如在基于几何特征的识别方法中，这一步主要是提取特征点，然后构造特征矢量；在统计识别方法中，特征脸方法是利用图像相关矩阵的特征矢量构造特征脸，而隐马尔可夫方法则是对多个样本图像的空间序列训练出一个隐马尔可夫模型，它的参数就是特征值；模板匹配法用相关系数做特征；而大部分神经网络方法则直接用归一化后的灰度图像做为输入，网络的输出就是识别结果，没有专门的特征提取过程。

最后是人脸识别。数据库里预先存放了已知的人脸图像或有关的特征值，识别的目的就是将待识别的图像或特征与库里的进行匹配。识别的任务主要有两个：一个是人脸辨认，即确定输入图像为库中的哪一个人，是一对多的匹配过程；另一个是人脸证实，即验证某个人的身份是否属实，是一对一的匹配过程。根据输入图像的性质，可以将人脸识别分为静态图像的人脸识别和动态图像序列的人脸识别两大类。前者主要是用静态图像如从证件照片、罪犯照片、场景照片上扫描的图像进行识别；后者则是用摄像机摄取的时间图像序列进行识别。

1.2 人脸识别的研究历史[12, 28, 31, 63, 86, 87]

人脸识别的研究已有很长的历史，早在 19 世纪后期，Francis Galton 就曾对此问题进行了研究，他用一组数字代表不同的人脸侧面特征来实现对人脸侧面图像的识别。一直到 20 世纪 90 年代以前，典型的人脸识别技术始终是用人脸正面或侧面的特征点之间的距离量度，而且早期的人脸识别多集中于对侧影图像的研究。

Harmon 等人利用与 Galton 类似的方法识别人脸，他采用 9 个基准点表征侧影，在此基准点上导出一组特征，如基准点之间的距离和角度、由基准点形成的三角形区域的面积等，然后利用特征之间的归一化欧氏距离进行识别。其后期的工作又增加了两个基准点和一些新的特征，而且人脸侧影轮廓曲线可从侧影图像中自动抽取得到。Kaufman 和 Breeding 也设计了一个对人脸侧影进行识别的系统，他们

采用基于特征的方法，其中特征为极坐标形式的自相关函数的系数，他们同时对动量不变性特征也进行了实验。Baylou 等人选择 10 个特征点对人脸侧影进行识别，Wu 和 Huang 则采用三次 B 样条函数抽取 6 个侧影基准点，利用从中导出的 24 个特征对东方人的侧影人脸图像进行匹配识别。Lapreste 等人利用距离探测器来获得人脸侧影图像并从中抽取特征点，然后用欧氏距离对人脸进行匹配。Lee 和 Milios 同样利用距离图像来匹配两幅人脸侧影的相似特征。

由于侧影识别对获取图像的约束较多，人们逐渐转向对正面人脸的识别研究。最早的半自动正面人脸识别系统由 Bledsoe 于 20 世纪 60 年代提出。在该人脸识别系统中，首先由操作员定出面部特征点并将其位置输入计算机，给定这些特征点之间的距离，采用最近邻原则或其他分类规则即可识别出待测试的人脸。由于特征提取是由人来完成的，该系统对人脸的旋转、倾斜等变化不太敏感。Kelly 对 Bledsoe 的系统进行了改进。他采用一种从上到下的分析方法从图像中自动抽取头部和身体的轮廓，然后应用一些启发式方法搜索眼睛、鼻子和嘴的位置。在这个人脸识别中主要用到的距离测度有头的宽度和两眼之间、头顶与眼睛之间、眼睛与鼻子之间以及鼻子与嘴之间的距离。Kaya 和 Kobayashi 采用统计识别方法，用欧氏距离来表征人脸特征，用人脸的 9 个显著特征组成特征向量，包括内眼宽度、外眼宽度、鼻子宽度、嘴的宽度、鼻尖处脸的宽度、鼻子与眼睛中间处脸的宽度、下唇与下巴之间的距离、上唇与鼻子的距离以及嘴唇的高度，然后基于这些特征及其估计建立分类器，对每个特征根据其统计行为确定一个变化阈值，根据图像特征向量之间的绝对范数进行分类。Buhr 采用 32 个原始特征和 12 个辅助特征进行识别，其中原始特征包括 21 个距离特征、4 个坐标差分特征、4 个三角形面积特征、2 个距离比特征以及 2 个特殊特征（两眼的面积），然后利用线性判决树决定最佳匹配。其他采用非欧氏距离表征人脸特征的还有：Campbell 用最小二乘法实现最佳匹配，Ricca 用聚类技术进行最佳匹配等。

Kanade 设计了一个高速的且有一定知识引导的识别系统，他创造性地运用积分投影法从单幅图像上计算出一组人脸几何特征参数，再

利用模式匹配技术与标准人脸相比较。Kanade 的特征点定位工作由两个阶段组成，首先利用粗分辨率差分图像的积分投影确定眼睛、鼻子和嘴的大致位置，然后在高分辨率图像上将人脸分为左眼、右眼、鼻子和嘴四个区域，从中抽取 16 个脸部特征组成特征向量。为消除尺度变化的影响，在识别前还对得到的特征向量进行了归一化处理。相比之下，Baron 所做的工作较少为人所知，他先将图像灰度归一化，再利用四个掩膜（眼、鼻、嘴及眉毛以下的整个脸部）表示人脸，然后分别计算这四个掩膜与数据库中的每幅标准图像的相应掩膜之间的互相关函数，以此作为判别依据。

总的来说，早期的人脸识别方法都需要利用操作员的某些先验知识，仍然摆脱不了人的干预。20 世纪 90 年代以来，随着高速度高性能计算机的出现，人脸识别方法有了重大突破，进入了真正的机器自动识别阶段，人脸识别研究也得到了前所未有的重视。国外有很多大学在此方面取得了很大进展，他们研究涉及的领域很广，其中有从感知和心理学角度探索人类识别人脸机理的，如美国 Texas at Dallas 大学的 Abdi 和 Toole 小组，主要研究人类感知人脸的规律，如漫画效应、性别识别与人脸识别的关系、种族效应等；由 Stirling 大学的 Bruce 教授和 Glasgow 大学的 Burton 教授合作领导的小组，主要是研究人类大脑在人脸认知中的作用，并在此基础上建立了人脸认知的两大功能模型，他们对熟悉和陌生人脸的识别规律以及图像序列的人脸识别规律也进行了研究。也有从视觉机理角度进行研究的，如英国 Aberdeen 大学的 Craw 小组，主要研究人脸视觉表征方法，他们对空间频率在人脸识别中的作用也进行了分析；荷兰 Groningen 大学的 Petkov 小组，主要研究人类视觉系统的神经生理学机理并在此基础上发展了并行模式识别方法。更多的学者则从事利用输入图像进行计算机人脸识别的研究工作。

在用静态图像或视频图像做人脸识别的领域中，国际上形成了以下几类主要的人脸识别方法：基于几何特征的人脸识别方法，主要代表是 MIT 的 Brunelli 和 Poggio 小组，他们采用改进的积分投影法提取出用欧氏距离表征的 35 维人脸特征矢量用于模式分类；基于模板匹配

的人脸识别方法，主要代表是 Harvard 大学 Smith-Kettlewell 眼睛研究中心的 Yuille，他采用弹性模板来提取眼睛和嘴巴的轮廓，Chen 和 Huang 则进一步提出用活动轮廓模板（即 Snakes 模型）提取眉毛、下巴和鼻孔等不确定形状；基于 K-L 变换的特征脸方法，主要研究者是 MIT 媒体实验室的 Pentland 小组，在此基础上还出现了各种改进方法，如 Yale 大学的 Belhumeur 提出的 Fisher 脸方法等；隐马尔可夫模型方法，主要代表有 Cambridge 大学的 Samaria 小组和 Georgia 技术研究所的 Nefian 小组；神经网络识别方法，如 Poggio 小组提出的 HyperBF 神经网络识别方法，英国 Sussex 大学的 Buxton 和 Howell 小组提出的 RBF 网络识别方法等；基于动态链接结构的弹性图匹配方法，主要研究者是由 C. Von der Malsburg 领导的德国 Bochum 大学和美国 Southern California 大学的联合小组；利用运动和颜色信息对动态图像序列进行人脸识别的方法，主要代表是 Queen Mary 和 Westfield 大学的 Shaogang Gong 小组。其他有影响的从事人脸识别研究的单位还有：Carnegie Mellon 大学、Stanford 大学、Maryland 大学、George Mason 大学以及 Illilois 大学等。

近几年来，国际上发表有关人脸识别方面的论文数量大幅度增加，从 1990 年到 2000 年之间，EI 可检索到的相关文献多达数千篇，IEEE 的 PAMI 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，而且 IEEE 还专门召开了四次人脸和手势识别国际会议（1995，1996，1998，2000 年）。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库和一套通用测试标准，用于定期对各种人脸识别算法进行性能测试，其分析测试结果对未来的工作起到了一定的指导作用。另外，从 1994 年开始，一些科研单位和公司开始将研究成果转移为实用产品，如 Miros 公司的 TrueFace，Visinocs 公司的 FaceIt，以及 Zn Bochum Gmbh 公司研制的 ZN-Face 等。

1.3 国内人脸识别研究现状

国内关于人脸识别的研究始于20世纪80年代，主要是在国际上流行方法基础上作了发展性工作。

四川大学周激流等实现了具有反馈的人脸正面识别系统，运用积分投影法提取面部特征的关键点并用于识别，获得了较为满意的效果。他们同时尝试了“稳定视点”特征提取方法，为使系统中包含3D信息，他对人脸侧面剪影识别做了一定研究，并实现了正、侧面互相参照的识别系统[86,94,95]。

中国科技大学杨光正等[79,99]提出一种基于镶嵌图的人脸自动识别方法，采用基于知识的三级金字塔结构对人脸进行分割和定位，前两级建立在不同分辨率的镶嵌图基础上，用于对人脸进行基本定位，第三级用一种改进的边缘检测方法进一步检测眼睛和嘴。基于这些器官的匹配就可进行人脸识别。

清华大学张长水等[100]对特征脸的方法做了进一步发展，提出采用类间散布矩阵作为产生矩阵，进一步降低了产生矩阵的维数，在保持识别率的情况下大大降低了运算量。他们对多模板的人脸检测问题也进行了研究[115]。

南京理工大学杨静宇等[93]主要是采用奇异值分解方法进行人脸识别研究，如用Daubechies正交小波变换对人脸图像作预处理，得到它在不同频带上的4个子图像，对它们分别提取奇异值，然后用最近邻方法进行分类，同时设计一种适用于多分类结果融合的群体决策算法，并且对分类结果有选择的进行融合。他们还研究了基于Fisher最佳鉴别矢量的人脸识别方法，并对神经网络用于人脸识别也进行了研究[103,104,121]。

上海交通大学李介谷等[88~91,109]则专门研究了人脸斜视图像的集合特征提取与恢复。他们的实验建立了人脸斜视图像的数学模型，并对如何从斜视图像特征中恢复出标准特征做了一定研究，对如何消除