

北京交通大学

硕士学位论文

人脸识别算法研究和实现

姓名：马驰

申请学位级别：硕士

专业：模式识别与智能系统

指导教师：阮秋琦

20071201

## 中文摘要

**摘要:**人脸自动识别技术是用计算机分析人脸图像,从中提取有效的识别信息并通过分类器自动鉴别的一种技术。作为生物特征识别的一种,人脸识别更为直接、友好而自然。因此,人脸识别自上个世纪以来一直受到研究者的极大关注。

人脸识别技术可分为人脸特征提取和分类器设计两方面,两方面的技术相辅相成而又相互制约,我们最终的研究的目的就是提取出最具代表性的人脸特征,并通过分类性能优良的分类器进行识别分类。本文主要工作内容包括:经典算法实现、理论研究及改进、实验证明和最终人脸识别系统的设计。重点描述如下:

第一,特征提取和融合。目前,人脸特征提取的算法很多,每种算法都有各自的优点和缺点;因此近些年来,许多学者倾向于将几种算法混合起来达到较好的识别效果。本文也是从此角度出发,在对以有经典算法进行实现和筛选的基础上,又对特征融合方法做了大量实验和尝试。最终选定独立成分分析,线性判别,离散余弦变换三种特征融合的方法提取人脸特征。实验结果表明,此三种特征的融合可以更好的表征人脸图像的统计特征,相对每种方法单独使用而言,识别正确率的提升比较明显,而计算量并没有显著提高。

第二,支持向量机的研究。支持向量机是目前比较流行的模式识别方法,在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势。本文将支持向量机用于人脸识别,主要对支持向量机的核函数和支持向量机的参数选择进行了研究,提出了基于  $\cosh$  的核函数,实验表明,该核函数在人脸识别方面效果优于普遍使用的传统核函数。并且提出了基于离散微粒群算法的支持向量机参数选择方法,大幅度提高了核函数的参数选择速度,从而使得本文提出的基于  $\cosh$  的核函数的实用性大大提高。

第三,参与创建北京交通大学人脸识别系统。在 Visual Studio 2005 为实验平台,将独立成份分析算法和基于  $\cosh$  的核函数用于实时人脸识别系统,证明了以上算法的可行性和有效性,效果比较理想。

本文对人脸识别技术和人脸识别系统的实现有比较全面的论述,理论研究关注的重点在于人脸特征的提取和改进,以及支持向量机核函数的改善,一系列的工作为日后的研究奠定了基础也确定了方向。同时也充分认识到本文自身的局限性,并对后续研究提出了目标与希望。

**关键词:**人脸识别系统;独立分量分析;特征融合;支持向量机;核函数;微粒群优化算法;支持向量机参数选择

**分类号:**

## ABSTRACT

### ABSTRACT:

The automatic human face recognition is a technology that using computer to analyze the human face images and extracting effective features from the human face images, then to recognize them. Compared with other biometric methods, such as fingerprint, palm, iris, and DNA, human face recognition is more friendly, direct, and imitate to humane. Therefore, in recent years, face recognition is getting more and more attention

I did some researches on both face features extraction and features recognition. The total work of the paper research includes:

Firstly, researches on typical statistical features extraction methods, including PCA, ICA, LDA, DCT, then developed a feature combination of ICA, LDA and DCT.

Secondly, researches on SVM and its application on face recognition. Development of a kernel function based on cosh function, and its application on face recognition. Development of a characters selection method of SVM based on Discrete Particle Swarm Optimizer, which effectively accelerates the characters selection.

Thirdly, establishment of the real time face recognition system, using ICA for feature extraction and SVM for classification. Validation of the feasibility and activities of the related arithmetic.

### KEYWORDS:

Face Recognition; Real Time Face Recognition System; PCA; ICA; DCT; LDA; Feature Fusion; SVM; Kernel Function; PSO; Character Selection.

### CLASSNO.:

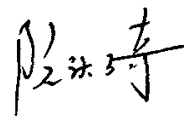
## 学位论文版权使用授权书

本学位论文作者完全了解北京交通大学有关保留、使用学位论文的规定。特授权北京交通大学可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：马驰

导师签名：



签字日期：2007年12月1日

签字日期：2007年12月21

## 独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京交通大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：

签字日期：

年 月 日

## 致谢

本论文的研究工作是在我的导师阮秋琦教授的悉心指导下完成的,阮教授严谨的治学态度,勤勉的工作精神以及科学的工作方法给了我极大的帮助和影响。阮教授对于我的科研工作和论文都提出了许多的宝贵意见,在此向阮老师表示衷心的感谢!

同时,在我攻读硕士学位的这三年来,信息所裘正定老师,苗振江老师,梁满贵老师和孙冬梅老师也给予了我很多的帮助,在此向几位受人尊敬的老师表示衷心的感谢!

并且,在实验室工作及撰写论文期间,孙雯玉、金一、白小曼、张百振、安高云、曹刚、张怀福、瞿国伟、侯迎宾、支瑞聪、杨旻、谢竞、曹溪渺、任苏亚等同学对我论文中的研究工作给予了热情帮助,在此向他们表达我的感激之情。

另外也感谢父母和亲友,他们的理解和支持使我能够在学校专心完成我的学业。感谢他们二十多年来对我学业的支持。

最后,向在这两年半以来所有关心过我帮助过我的人表示感谢。

## 序

人脸识别是生物识别技术的一种。人脸自动识别技术就是用计算机分析人脸图像，从中提取有效的识别信息并自动鉴别的一种技术。并且这也是一种最为方便、有效的识别方法，用人脸特征进行身份验证自然方便，易于为用户接受。人脸识别因其在公安、验证系统、档案管理、视频会议、人机交互系统等方面的巨大应用前景而越来越成为当前模式识别和人工智能领域的一个研究热点。

人脸识别的研究始于 20 世纪 60 年代末，经过了几十年的发展，人脸识别算法取得了很大的成就。从早期的提取几何特征的方法与模板匹配的方法到目前的两个主要研究：基于整体的研究方法和基于特征分析的方法。

本论文主要是研究人脸识别的算法，目的是通过研究实验，找到合适通过计算机来实现简单直观的人脸识别算法。

本课题是国家自然科学基金（No.60472033）、国家 973 项目（No.2004CB318005）以及教育部博士点基金资助项目（No.20030004023）。

## 1 综述

### 1.1 人脸识别的研究背景

随着社会的发展以及技术的进步,尤其是最近十年内计算机的软硬件性能的飞速提升,以及社会各方面对快速高效的自动身份验证的要求日益迫切,生物识别技术在科研领域取得了极大的重视和发展。由于生物特征是人的内在属性,具有很强的自身稳定性和个体差异性,因此是身份验证的最理想依据。在人与人的接触中,人脸所包含的视觉信息占据了重要地位,它是区分人与人之间差异的最重要特征之一。而人脸识别就是利用计算机从图像或图像序列中检测出人脸,并判断其身份。而相对于指纹、虹膜、掌纹、步态、笔迹、声纹等生物特征,利用人脸来识别具有不可比拟的优势,具体体现为<sup>[1][2]</sup>:

1. 操作隐蔽,特别适用于安全、监控和抓逃工作。
2. 非接触式采集,无侵犯性,容易接受
3. 方便、快捷、强大的实时追踪能力
4. 符合人类识别习惯,交互性强
5. 设备成本较低(摄像头)

计算机人脸识别的研究内容从广义上讲大概包括以下几个方面<sup>[3][4]</sup>:

1. 人脸检测(Face Detection),就是从各种不同的场景中检测出人脸的存在并确定其位置。这方面的研究主要受光照、噪声、面部倾斜度以及各种各样的遮挡的影响。

2. 人脸表征(Face Representation),就是采取某种表示方法表示检测出的人脸和数据库中的已知人脸。通常的表示法包括几何特征(例如欧式距离、曲率、角度等)、代数特征(例如矩阵特征矢量)、固定特征模版、特征脸等。

3. 人脸鉴别(Face Identification),就是通常所说的“人脸识别”,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。这个过程是选择适当的人脸表征方式与匹配策略。整个识别系统的构造与人脸的表征方式密切相关。

4 表情姿态分析(Expression Gesture Analysis),就是对待识别人脸的表情或者姿态信息进行分析,并对其进行分类。

5. 生理分类(Physical classification),就是对待识别人脸的生理特征进行分析,得出其年龄、性别等相关信息。

人脸识别技术具有广泛的应用前景,在国家安全军事安全和公共安全领域,智能门禁、智能视频监控、公安布控、海关身份验证、司机驾照验证等是典型的应用;在民事和经济领域,各类银行卡、金融卡、信用卡、储蓄卡的持卡人的身



份验证、社会保险人的身份验证等具有重要的应用价值；在家庭娱乐等领域，人脸识别也具有有一些有趣有益的应用，比如能够识别主人身份的智能玩具、家政机器人、具有真实面像的虚拟游戏玩家等等。主要有以下几类应用<sup>[5]</sup>：

1 刑侦破案。当公安部门获得案犯的照片之后可以利用人脸识别技术在存储罪犯照片的数据库里找出最相象的人为嫌疑犯还有一种应用就是根据目击证人的描述画出草图然后用这种图到库里去找嫌疑犯罪数据库往往很大由几千幅图像组成如果这项搜索工作由人工完成不仅效率低而且容易出错因为人在看了上百幅人脸图像后记忆力会下降而由计算机来完成则不会出现此问题。

2 证件验证。身份证驾驶执照以及其他很多证件上都有照片现在这些证件多是人工验证的如果用人脸识别技术这项任务就可以交给机器完成从而实现自动化智能管理当前普遍使用的另一项证件是用符号或者条形码标记的比如信用卡自动提款机等这类卡的安全系数比较低因为卡可能丢失密码也可能被遗忘或者窃取如果在这类卡上加上人脸的特征信息则可大大改善其安全性能。

3 入口控制。入口控制的范围很广它可以是设在楼宇单位或私人住宅入口的安全检查也可以是计算机系统或者情报系统等的入口检查在楼宇或某些安全部门的入口处比较常用的检查手段是核查证件当人员出入频繁时要求保安人员再三检查证件是很麻烦的而且安全系数也不高在一些保密要求非常严格的部门除了用证件还要用一些另外的识别手段如指纹识别手掌识别视网膜识别和语音识别等人脸识别与之相比具有直接方便和友好的特点当前计算机系统的安全管理也倍受重视通常使用由字符和数字组成的口令可能会被遗忘或者破解但是如果把人脸当作口令则又方便又安全。

4 视频监视在许多银行公司公共场合等处都设有24小时的视频监控另外侦察员在破案时也要用摄像机对人进行跟踪在对图像进行集体分析时也要用到人脸的检测跟踪和识别技术。

除了这几部分的应用外人脸识别技术还可用于视频会议机器人的智能化研究以及医学等方面<sup>[6]</sup>。

## 1.2 人脸识别的研究历史

最早的关于人脸识别的研究可以追溯到十九世纪末 Sir Francis Galton 发表于 Nature 的两篇文章。Galton 在人脸侧面轮廓图像上做了一些工作<sup>[7]</sup>。

为了更好的对人脸识别研究现状进行介绍，本文将自动人脸识别的研究历史按照研究内容、技术方法等方面的特点大体划分为以下三个时间阶段：

- 1964-1990

- 1990-1997
- 1998-至今

### 1.2.1 1964 年到 1990 年

Bledsoe 的工作可以视为真正自动人脸识别研究的开端, 在二十世纪六十年代中后期, 他以人工提取的基准点及其空间关系的几何特征为基础<sup>[8]</sup>。

Bledsoe 是最早利用现代计算机技术研究人脸识别的学者之一。他在 1964 年提出了一个基于人脸直观特征的半自动人脸识别系统。Kelly<sup>[9]</sup>在 Bledsoe 的框架之上做了改进, 但一个明显的进步是它不需要人的干预。他的方法利用了人的身体和头部图像。其中脸部的测量包括头部的宽度、眼睛之间的距离, 头顶到眼睛的距离, 眼睛到鼻子的距离, 以及眼睛到嘴巴的距离。同样采用了最近邻分类法。后来很多学者对这一方法进行了发展。包括: Y.Kaya 和 K.kobayashi<sup>[10]</sup>, I.Craw<sup>[11]</sup>, T.Poggio<sup>[12]</sup>, F.Girosi<sup>[13]</sup>, R.Brunelli<sup>[14]</sup>, X.Jia, M.S.Nixon<sup>[15]</sup>, Nicholas Roeder, Xiaobo Li<sup>[16]</sup>等。

七十年代人脸自动检测与识别开始引起不少学者的浓厚兴趣, 从而形成了第一次研究高潮; 这一阶段研究的主要方向集中于人脸识别。无论是正面图像还是侧面图像, 基本上都是采用基于面部器官的一些几何特征, 通过标准的模式分类技术, 实现人脸识别。

其中, 比较有代表性的工作有 Kelly 和 Kanade<sup>[17]</sup>的两篇博士论文, Kanade 设计了一个高速且有一定知识导引的半自动回溯识别系统, 创造性地运用积分投影法从单幅图像上计算出一组脸部特征参数, 再利用模式分类技术与标准人脸相匹配。Goldstion、Harmon 和 Lesk 用几何特征参数来表示人脸正面图像, 他们采用 21 维特征向量表示人脸面部特征, 并设计了基于这一特征表示法的识别系统。Kaufman 和 Breeding 以及 Harmon 等人的研究工作, 对交互式人脸识别方法在理论和实践上进行了详细论述。

八十年代是 FRT 的休眠期。其中 Harmon 在 1981 年运用结构方法用 17 个几何特征对 112 人的脸型侧面图像获得了 90% 的识别精度<sup>[18]</sup>。Stonham 在 1984 年提出了一种单隐层的自适应神经网络来进行人脸识别、表情分析和人脸鉴定, 每一个人对应一个网络。该系统需要 200~400 个样本来训练每一个分类器训练样本存在平移和表情变化, 16 个人共需 16 个分类器。对于输入图像, 通过判断哪一个分类器可以得到最大的响应输出, 来完成人脸识别。该系统对人脸鉴定和人脸表情分析也作了进一步的研究。

### 1.2.2 1990 年到 1997 年

九十年代初期,研究者对 FRT 的研究兴趣又开始上升并且取得较大突破,提出了多种全自动识别系统。Turk 和 Pentland 于 1991 年提出了基于 KL 变换的特征脸方法<sup>[16]</sup>。纯粹从概率的角度来考虑人脸识别问题。该方法首先使用 Karhunen-Leove(KL)变换来表达和识别人脸,确定了特征脸之后,每一幅人脸图像都可以表示为一个权值矢量。Welsh 在 1992 年使用该方法来进行人脸特征编码<sup>[17]</sup>。而这些权值可以通过计算人脸图像和特征脸的内积而得到。所谓特征脸就是对应于人脸(模式)协方差矩阵的那些较大特征值的特征向量。由特征脸所张成的子空间在维数上比原模式空间大大减少,人脸检测和识别工作就在该子空间上进行。

Nakamura 于 1991 年提出了一种等灰度线的人脸识别方法<sup>[18]</sup>。

Yuille 于 1992 年提出了一种可变模板法来检测和描述人脸特征<sup>[19]</sup>。

洪子泉(1992)年和杨静宇(1994)年最早,提出了代数特征抽取的方法。认为图像的代数特征反映了图像的内在属性,提出了人脸图像的奇异值特征<sup>[18]</sup>。

Cheng 和 Hong<sup>[20][21]</sup>于 1991 年中提出了基于 SVD(奇异值分解方法)的人脸识别方法,针对 45 个特征向量样本,该方法得到了最佳分类面和二次分类器。该分类器能够识别 45 个训练样本,新的测试样本由 13 幅照片构成,该方法存在 42.67%的错误率。但有人认为这是由于训练样本太少所导致的。

Manjunath 在 1992 年和高希奇在 1996 年将小波变换应用在人脸识别研究中。Manjunath 使用对每幅人脸图像使用 Gabor 小波分解,提取特征点。

Brunelli 和 Poggio<sup>[22]</sup>在 1993 年使用了一个基于几何特征的人脸识别器,该识别器在一个较大的由 47 个被测试者的数据库中测试,其正确识别率大约为 90%。然而使用一个极为简单的模板匹配方法在同样的数据库中却取得了高达 100%的识别率。

Liu 于 1993 年提出了基于最佳鉴别准则抽取人脸识别方法研究是人脸识别的主流。

Comon 于 1994 年提出了独立成份分析(ICA)概念。并且 Hyvarner 和 Karhunen 于 1997 年提出了不同的 ICA 算法<sup>[23]</sup>。

Samaria 在 1993 年最早将 HMM 用于人脸识别。Nefian 发展了 Samaria 的方法,提出了基于 2D-DCT 特征提取的方法。在一定程度上解决了 Samaria 的大存储量的缺陷,但识别率提高不大<sup>[24]</sup>。

Starner 于 1996 年将隐马尔可夫模型(HMM)应用与人脸表情识别。

Kanade 于 1997 年提出了一种使用几何参数的方法来识别人脸。该系统抽取 16 个正面人脸几何特征(随后减少到 13 个)。使用的数据库为每个测试者有一

个训练图像和一个测试图像,共有 20 个被测试者,正确识别率为 45%到 75%之间<sup>[25]</sup>。

与此同时,基于图像序列的人脸识别技术也得到了发展。

Chellappa 于 1995 年从图像序列中分割运动目标,结构估计,三维人脸目标与非刚体运动分析<sup>[26]</sup>。

Bichsel 和 Pentland 于 1994 年将基于模板的方法用于在图像序列中寻找头部,利用所谓的人脸系统拓扑学来分析人脸的集合变化如平移、旋转、缩放等<sup>[27]</sup>。

Horn 于 1981 年和 Barron 于 1994 年将光流场分析的方法应用于人脸识别。尽管精确的光流计算仍然是一个未解决的问题,光流场分析方法受到了人们广泛的重视,可用于观察者偏离运动的检测、运动目标的分割与检测等。

Darrell 于 1996 年和彭辉<sup>[28]</sup>以及郭跃飞<sup>[29]</sup>于 1997 年提出基于 KL 变换的特征脸方法与以奇异值特征为代表的代数特征方法<sup>[29]</sup>。

Juell、Schofield、Intrator<sup>[30]</sup>、Yoon、Rowley、Ranganath 等人提出了人工神经网络的方法。

### 1.2.3 1998 年—至今

近年来,人脸识别的算法研究不断出新,并取得了长足的进步。Scholkopf B 等人在 1999 年提出非线性子空间法<sup>[16]</sup>,通过引入核函数,在不提高计算复杂性的条件下,将人脸图像由原始空间非线性映像到一个高维的特征空间,从而更有利于模式的分类。2000 年左右以来,研究方法从线性子空间向非线性子空间过渡成为一种趋势。微软亚洲研究院 Face Group 在其研究报告中指出,线性子空间分析法与传统方法相比,有着基于事例、特征学习、维数降低和降低线性映射空间的特点<sup>[17]</sup>;而非线性子空间分析法则具有更加有效建模和固有特征维发现的特点。

## 1.3 人脸识别技术的发展

人脸识别系统的研究涉及模式识别、图像处理、生理学、心理学、认知科学等多种学科,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。一个完整的人脸识别系统应完成下列任务<sup>[6]</sup>:

(1) 对于任意图像,确定其中是否存在人脸。若有人脸存在,确定人脸的数目及各自的位置和大小,并对人脸图像做预处理。

(2) 提取人脸图像特征。

(3) 对特征进行分析处理,与已知人脸库内的人脸图像进行比较,识别,确认身份。

其中的(1)为人脸检测的内容,(2)、(3)是人脸识别与理解的内容。如图1.1所示,人脸自动识别过程包括三个主要环节,首先是人脸检测和定位,既从输入图像中找到人脸及人脸存在的位置,并将人脸从背景中分割出来,然后是对归一化的人脸图像进行特征提取与识别。

人脸识别系统三个环节中,特征提取与识别相对人脸检测更为关键,相关研究也更加深入。目前,静止图像中的人脸识别大致有三个研究方向:(1)基于几何特征;(2)基于统计特征;(3)基于连接机制。

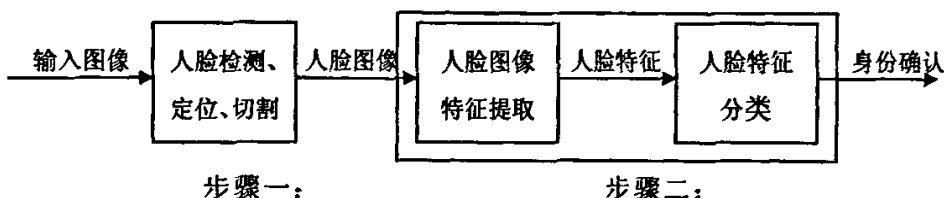


图 1.1 人脸识别系统

Figure 1.1 Face Recognition System

### 1.3.1 基于几何特征的人脸识别技术

人脸的几何特征包括各个面部器官的形状、灰度以及各个器官之间的结构关系。基于几何特征的人脸正面图像识别方法,是通过人脸面部拓扑结构几何关系的先验知识,利用基于结构的方法在知识的层次上提取人脸面部主要器官特征,将人脸用一组几何特征矢量表示,识别归结为特征矢量之间的匹配。基于欧氏距离的判决是最常用的识别方法。正面识别所采用的几何特征是以人脸器官的形状和几何关系为基础的特征矢量,基分量通常包括人脸指定两点间的欧氏距离、曲率、角度等。

侧影识别<sup>[1]</sup>也是早期基于几何特征人脸识别的一个重要方法,其基本原理是从人脸的侧影轮廓线上提取特征点,将侧影转化为轮廓曲线,从中提取基准点。根据这些点之间的几何特征来进行识别。由于侧影识别相对较简单且应用面小,对侧影识别的研究较少。

基于几何特征识别的方法具有如下优点:①符合人类识别人脸的机理,易于理解;②对每幅图像只需存储一个特征矢量,存储量小;③对光照变化不太敏感。该方法缺点是:①从图像中抽取稳定的特征比较困难;②对强烈的表情变化和

姿态变化的鲁棒性较差；③一般几何特征只描述了部件的基本形状与结构关系，忽略了局部细微特征，更适合于粗分类。

总体来讲，基于几何特征的人脸识别方法，特征提取不精确，而且由于忽略了整个图像的很多细节信息，识别率较低，所以近年来已经很少有新的发展。

### 1.3.2 基于统计特征的人脸识别技术

基于统计特征的方法把一幅人脸图像的像素点灰度值矩阵看作随机向量或矩阵（2D 方法），从而用统计方法来分析人脸模式，以反映矩阵某些性质的数值特征作为人脸的特征，这类方法有着完备的统计学理论支持，得到了较好地发展，出现了一些较成功的算法，目前常见的几种技术介绍如下。

1) 特征脸法<sup>[32]</sup> Turk和Pentland在1991年发展了一种人脸识别的方法，这就是著名的特征脸方法(Eigenface)。所谓特征脸就是对应于人脸(模式)协方差矩阵的那些较大特征值的特征向量，特征脸方法是从主成分分析(PCA)导出的一种人脸识别和描述技术。这种方法将包含人脸图像区域看作一种随机向量，因此可以采用KL变换得到正交变换基，对应其中较大的特征值的基底具有与人脸相似的形状，即特征脸。算法利用这些基底的线性组合可以描述、表达人脸和逼近人脸，因此可以进行人脸的识别和重建。识别过程就是把待识别人脸映射到由特征脸张成的子空间中，特征脸所生成的子空间在维数上比原模式空间大大减少，人脸检测和识别工作就在该子空间上进行。

识别时可采用主分量作正交基的主分量方法<sup>[31]</sup> (PCA)，也可采用次分量作正交基的次分量方法。与较大特征值对应的正交基（也称主分量）可用来表达人脸的大体形状，与小特征值对应的特征响应（也称次分量）可用来描述具体细节。用次分量作为正交基的原因是所有人脸的大体形状和结构相似，真正用来区别不同人脸的信息是那些用次分量表达的高频成分。

后来Pentland等人进一步扩展了特征脸方法，将类似的思想运用到面部特征上，分别得到了本征眼、本征鼻、本征嘴，并且将它们结合起来进行人脸识别。实验结果表明，这样比单独使用特征脸效果更好。特征脸方法从能量压缩和重建误差最小化的角度来讲，PCA是最优的方法。但它对于外界因素所带来的图像差异和人脸自身所造成的差异是不加区分的，因此外界因素(例如光照、姿态)变化会引起识别率的降低。

2) 线性判别分析法 特征脸方法使用由各个特征脸扩展的空间来表示人脸，虽然可以有效地表示人脸信息，但是并不能有效鉴别和区分人脸。很多研究者提出了使用其他线性空间来代替特征脸空间以取得更好的识别效果。此中线性判别

分析(Linear Discriminate Analysis, LDA)方法(也叫Fisher脸方法)利用了类别归属信息,它选择类内散布正交的矢量作为特征脸空间,从而压制了图像之间与识别信息无关的差异,强调了不同人脸之间的差别,同时弱化了同一人脸由于光照、视角和表情而引起的变化,获得了比特征脸更好的识别效果。

LDA的目的就是从高维特征空间里提取出最具有判别能力的低维特征<sup>[34]</sup>。这些特征能帮助将同一个类别的所有样本聚集在一起。不同类别的样本尽量分开。即选择使得样本类间离散度和类内离散度的比值最大的特征。

3) 独立分量分析法 独立分量分析(Independent Component Analysis, ICA)最早是作为一种新的盲源分离技术而出现,目前已被引入到模式识别领域中,作为特征提取的方法<sup>[35][36]</sup>。ICA法假设:目标信号由各个互相独立的源信号现行混合而成,而我们要做的就是根据混合后的信号矩阵来获得源信号。对于人脸识别,我们可以利用ICA方法获得训练人脸集的独立分量,即独立分量脸,进而将人脸图像(包括训练集和检测集)投影到由独立分量脸张成的独立分量空间,投影系数作为特征向量用于后续的训练和识别过程。

ICA理论及其分离算法的关键在于如何定量地判别分离结果的独立性。由中心极限定理,随机量如由许多相互独立的随机量之和组成,只要各独立的随机量具有有限的均值和方差,则不论各独立随机量为何种分布,该随机量必接近高斯分布。据此可知,由许多相互独立的随机量之和组成的随机变量,比其中任一独立的随机变量的分布更具有高斯性。因此我们可以在分离过程中,通过对分离结果非高斯性的度量来监测分离结果间的相互独立性。当非高斯性度量达到最大时,就表明已完成对各独立分量的分离。以信息理论一个基本的结论是,在所有具有等方差的随机变量中,具有高斯分布的随机变量信息熵最大,非高斯性越强,信息熵越小。这意味着熵能用来作为非高斯性的测量,分布明显的集中于某个值的熵很小,非高斯性测量中,高斯变量应该为零,而它总是非负,有人对熵的定义作了修改,称为负熵,负熵非负, $y$ 的非高斯性越强;负熵越大,当且仅当随机向量 $y$ 为高斯分布时,负熵为零<sup>[37]</sup>。

当前估计ICA模型的主要方法有非高斯的最大化;信息的最小化;大似然函数估计。由Aapo Hyvärinen, Ekki Oja提出的Fast Fixed-Point ICA的方法,大大提高了ICA方法的计算速度,使得处理高维数据的ICA问题成为可能<sup>[38][39]</sup>。

ICA方法同PCA方法相比,不同之处在于,PCA方法是最小均方误差意义上的最优维数压缩技术,而且PCA方法所抽取特征的各分量之间是不相关的。这种方法只基于数据的二阶统计信息(即基于相应协方差矩阵)进行分析,而忽略其高阶统计信息。ICA方法假设数据由若干独立源信号混合而成,并力求分析出这些独立源信号。在独立分量分析中数据的二阶和高阶统计信息都能得到利用,而且还

能消除高阶相关性。ICA方法在盲信号分离上得到了成功的应用，在抽取人脸图像特征与人脸识别上也取得了较好的效果。

### 1.3.3 基于连接机制的人脸识别技术

1)弹性图匹配法 弹性图匹配(Elastic Bunch Graph Matching)是一种基于人脸的结构特征和灰度分布信息的 Gabor 小波纹理分析相结合的识别算法<sup>[40][41]</sup>。它在图像空间中依据人脸图像的灰度分布信息选定特征点,以对于通常的人脸变形具有一定不变性的距离为边,采用属性拓扑图来代表人脸。拓扑图的任一顶点均包含一个特征向量,用来记录人脸在该顶点位置附近的信息,如图 1.2。拓扑图的顶点采用了小波变换特征,它对于光线、尺寸、角度具有一定的不变性。弹性图匹配能够容忍表情的变化,并在一定程度上能够容忍视角的变化。

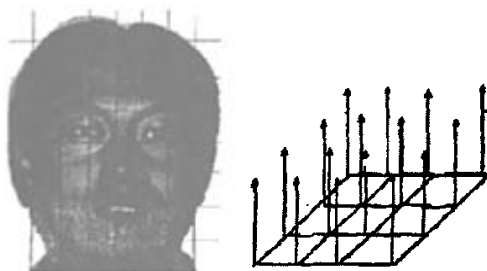


图 1.2 弹性匹配方法示意图

Figure 1.2 Elastic Bunch Graph Matching

由于该算法不但较好的兼顾到了人脸的整体信息和局部特征,而且利用了 Gabor 小波变换的结果与人眼视网膜上低层次的细胞对空间的响应相似的特性,因而具有良好的识别效果。但时间复杂度高,实现复杂。特征点的准确定位是影响识别速度的关键因素,因此改进特征点的定位方法,降低特征点的定位的时间耗费是提高识别速度的有效途径。人脸弹性图的设计是弹性匹配人脸识别的前提和基础,主要考虑两个方面的问题:(1)特征点的选择及特征表征;(2)边的选择及赋值。因此,一个好的弹性图既要能正确选择特征点和边的位置和数量,又要能够对特征点和边选择合适的表达方式。边的选择要根据特征点的情况进行,因此对特征点选择是弹性图设计的前提和基础。

一种弹性匹配的改进方法是 将 KL 变换应用于小波变换,来生成二维网格中顶点的矢量串,以减少其维数,从而大大减少了表达一幅人脸所需要的特征数量,而识别率不会明显下降。另一种方法是采用特殊点网格,而不是采用矩形网格作为拓扑图的节点,依据人脸结构选择稳定性较好且能表达人脸各部件信息的一组



特点,选择稳定性较好的邻点连线为边,采用特殊网格拓扑图来代表人脸。如图 1.3 所示。



图 1.3 特殊点弹性匹配示意图

Figure 1.3 Feature Point Elastic Bunch Graph Matching

弹性匹配方法采用小波变换特征来描述人脸的局部信息,和人眼视网膜对图像的响应相似,一定程度上容忍光线等干扰,对细微表情也不敏感。而且弹性匹配中的人脸模型还考虑了局部人脸细节,并保留了人脸的空间分布信息,且它的可变匹配方式一定程度上能够容忍人脸从三维到二维的投影引起的变形。但是,弹性匹配方法在实现时,需要考虑具体的参数选择,如二维网格的大小、小波变换参数的选择等,这些参数都会影响识别的效果。

2) 人工神经网络方法 基于神经网络的方法是最近几年比较活跃的一个研究方向。神经网络进行人脸的特征提取和分类器的设计,有比较成熟的人脸特征提取方法,如多主元分量提取算法——自适应主分量神经网络提取算法等。

Valentin 的方法是首先提取人脸的 50 个主元,然后用自相关神经网络将它映射到 5 维空间中,再用一个普通的多层感知器进行判别,对一些简单的测试图像效果较好。Intrator 等提出了一种混合型神经网络来进行人脸识别,其中非监督神经网络用于特征提取,而监督神经网络用于特征分类。Lee 等将人脸的特点用六条规则描述,然后根据这六条规则进行五官的定位,将五官之间的几何距离输入模糊神经网络进行识别,效果较一般的基于欧氏距离的方法有较大改善。Laurence 等采用卷积神经网络方法进行人脸识别,由于卷积神经网络中集成了相邻像素之间的相关性知识,从而在一定程度上获得了对图像平移、旋转和局部变形的不变性,因此得到非常理想的识别结果。Lin 等提出了基于概率决策的神经网络方法,其主要思想是采用虚拟样本进行强化和反强化学习,从而得到较为理想的概率估计结果,并采用模块化的网络结构 网络的学习,这种方法在人脸检测、人脸定位和人脸识别的各个步骤上都得到了较好的应用。

神经网络方法在人脸识别上的应用有一定的优势,因为对人脸识别的许多规律和规则进行显形的描述是相当困难的,而神经网络方法则可以通过学习的过程

获得对这些规律和规则的隐性表达，它的适应性更强，一般也比较容易实现。

### 1.3.4 人脸识别的核方法

模式分析算法的发展，经历了三次革命。20 世纪 60 年代，引入了在向量机集内检测线性关系的高效方法，并分析了这些计算行为和统计行为。20 世纪 80 年代，模式分析经历了一场非线性革命，几乎同时引入了后向传播多层神经网络和高效的决策数方法。20 世纪 90 年代中期，出现了新的被称为基于核的学习方法的模式分析方法<sup>[42]</sup>，使得高效的分析非线性问题成为可能。基于核的学习方法，首先以支持向量机（Support Vector Machine, SVM）的形式出现，用来摆脱计算和统计上的困难，然而很快就产生了基于核的算法，它能够解决分类以外的问题，如特征提取等。

#### 1) 由核定义的非线性特征映射

继支持向量机(SVM)发明之后，研究者们又不断提出新的核学习方法，近年来又出现了如：核主分量分析(KPCA)、核Fisher判决分析(KFDA)等。目前，在机器学习领域，对核学习方法及其应用研究非常活跃。

核(kernel)是一个函数 $k$ ，这类核学习算法的基本思想是：对于原空间中线性不可分的数据，首先经过一个非线性映射 $\Phi$ ，将原空间的数据映射到一个维数可以无穷大的高维的特征空间(核空间) $F=\{\Phi(X), X \in R^n\}$ 中，如图1.4所示。只要选择满足Mercer条件的核函数 $k$ ，就可以在这个特征空间中隐含地进行运算，实现数据在高维空间中的线性分类(或近似线性分类)，这样就可以利用一些线性算法来实现相对于原空间为非线性的算法，从而提高算法的性能。利用核函数 $k$ 代替原空间中的内积，就对应于将数据通过一个映射，映射到某个高维的特征空间中，高维特征空间是由核函数定义的。选定了一个核函数，也就对应地定义了一个高维特征空间。特征空间中所有的运算都是通过原空间中的内积核函数来隐含实现。即：任何(线性)只用到标量内积的算法都可以通过一个核函数在高维特征空

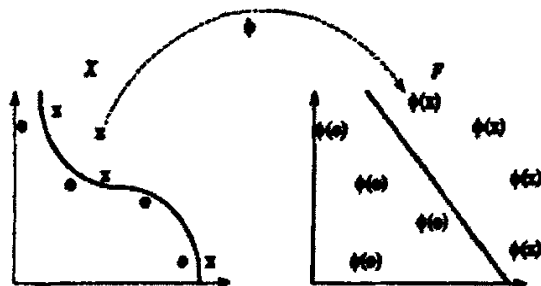


图 1.4 非线性映射

Figure 1.4 Non-linear Mapping

间中隐含地进行运算。我们可以利用此思想,在特征空间中实现一般的线性算法,但却能实现相对于原空间来说是非线性的算法,这将会大大地提高学习算法的效率。其优点如下:

(1) 核方法以统计学习理论为指导,具有坚实的理论基础。

(2) 利用核方法所训练的学习机器具有非常好的推广能力,因为它遵守了结构风险最小化原则。

(3) 核学习方法的抗干扰能力较强。

(4) 核学习方法具有强大的非线性和高维处理能力,利用核函数在高维空间中处理非线性问题时,很好地解决了高维空间中维数灾难问题。

## 2) 支持向量机

支持向量机<sup>[43][44][45][46][47]</sup>(Support Vector Machine SVM)是建立在统计学习理论基础之上,由Vapnik等人于20世纪年代提出的一种核学习机器,由于其在解决小样本问题时所表现出的卓越性能以及具有强大的非线性和高维处理能力,而倍受关注,已成为克服“为数灾难”和“过学习”等传统困难的有力手段。SVM算法是一种有监督学习方法,不仅可以进行样本的分类,而且可以进行函数的回归。

SVM是遵循结构风险最小化(SRM)原则的学习机器。SVM首先是从两类分类问题提出的,其目标是求得一个线性分类超平面,不仅使得两类分开(保证经验风险为0),而且使得两类的分开间隔最大(保证置信范围最小),以保证期望风险最小。

在实现支持向量机算法时,统计学习理论使用了与传统方法完全不同的思路,它不是像传统方法那样首先试图将原输入空间降维,进行特征选择和特征变换,而是设法将输入空间升维,使问题在高维空间中变得线性可分或接近线性可分。因为升维后只是改变了内积运算,并没有使算法的复杂性随着维数的增加而增加,而且在高维空间中的推广能力并不受维数影响,因此这种支持向量机方法才是可行的。支持向量机的这一特点提供了解决算法可能导致的“维数灾难”问题的方法:在构造判别函数时,不是对输入空间的样本作非线性变换,然后在特征空间中求解;而是先在输入空间比较向量(例如求点积或是某种距离),对结果再作非线性变换。这样,大的工作量将在输入空间而不是在高维特征空间中完成。同时,支持向量机的推广性也是与变换空间的维数无关的,只要能够适当地选择一种内积定义,构造一个支持向量数相对较少的最优或广义最优分类面,就可以得到较好的推广性<sup>[48]</sup>。

总的来说,支持向量机方法的优点主要包括以下几点:

(1) 算法适用于小样本问题,而不仅仅是样本数趋于无穷大时的才有最优值;

(2) 算法最终转化成为一个约束条件下的二次型寻优问题, 从理论上说, 得到的将是全局最优解, 解决了在神经网络方法中无法避免的局部极值问题;

(3) 算法具有强大的非线性和高维处理能力。保证了较好的推广能力, 并解决了维数问题, 其算法复杂度与样本维数无关, 只取决于支持向量的个数;

(4) 算法在很大程度上解决了模型选择、过学习、非线性、维数灾难等问题。

### 3) 核主成份分析法

Turk 等人提出的主分量分析法在进行人脸特征提取时, 只考虑了图像数据中的二阶统计信息, 未能利用数据中的高阶统计信息, 忽略了多个像素间的非线性相关性<sup>[49]</sup>。研究表明, 一幅图像的高阶统计往往包含了图像边缘或曲线的多个像素间的非线性关系。而核主成份分析(KPCA)是基于输入数据的高阶统计, 它描述了多个像素间的相关性, 所以 KPCA 能够捕捉这些重要信息, 从而取得更好的效果。同时, KPCA 的另一个优点就是可以把在输入空间不可线性分类的问题变换到特征空间实现线性分类, 简化了分类器的设计<sup>[50]</sup>。

Kernel PCA利用对偶表示把PCA应用到一个由核定义的特征空间, 基本思想是通过一个线性变换 $\Phi$ 把输入的向量化人脸图像数据 $X$ ,  $X \in R^n$ 映射到一个高维的特征空间 $F$ , 再在其中进行线性的主分量分析, 从而提取特征。因此特征空间中线性的PCA对应着输入空间的非线性PCA, 它可以把在输入空间无法线性分类的数据变换到特征空间来实现线性分类。

### 4) 核Fisher判决法

线性Fisher判决分析算法(Fisher Discriminant Analysis FDA)是R. A. Fisher于1936年提出的一种旨在降低特征维数的监督学习算法, FDA建立了一个子空间(由所有的投影轴构成), 所有样本在这个子空间内满足类内散度最小, 类间散度最大。所有样本在这些投影轴上的投影系数可以作为样本的特征向量, 利用这些特征向量, 就可以进行样本的分类识别。核Fisher判决分析算法(Kernel Fisher Discriminant Analysis KFDA)是将核学习思想引入到FDA中所产生的一种新的非线性特征提取方法。与KPCA算法类似, KFDA算法利用核学习思想, 首先通过一个非线性映射将原空间中的数据映射到一个高维的特征空间中, 通过引入核函数就可以在这个高维空间中进行线性Fisher判决分析, 这样, 在高维空间中所获得的样本特征相对于原空间来说就是非线性的特征, 这些非线性特征更有利于分类。

### 5) 核独立成分分析法

ICA的出发点非常简单, 它假设各成分是统计独立的且是线性的。但由于光照、表情等因素的影响, 它并不能处理人脸图像的非线性变化。核独立成分分析法(Kernel ICA)是一种基于非线性函数空间的ICA方法<sup>[51]</sup>, 它并不是现有ICA

方法的核化,而是一种新的 ICA 方法。

KICA 方法将利用满足 Mercer 条件的核函数,将特征空间映射到一个核空间,由核函数代替向量间的内积运算,实现非线性变换。KICA 有两种实现方法,核广义方差(Kernel Generalized Variance Analysis, KGVA)和核典型相关性分析(Kernel Canonical Correlation Analysis, KCCA)。核广义方差和核典型相关性分析的计算都可以归结为求特征值问题,不同之处是 KCCA 得到 F 相关系数的最小特征值,而 KGVA 考虑到所有特征值(行列式)。

在盲源分离方面, KICA 的鲁棒性优于 ICA<sup>[51]</sup>。同时,有实验表明,在人脸识别方面, KICA 也要优于 ICA<sup>[52]</sup>。但其计算复杂, KGV 和 KCCA 的不完全 Cholesky 分解的时间复杂度都达到了  $O(mM^2N)$ , 另外,在求解分解矩阵的最优化的过程中,为了防止陷入局部最优的启发性方法也大大增加了计算时间。

## 1.4 人脸识别研究的难点和意义

### 1.4.1 人脸识别研究的难点

人脸识别是一项极具挑战性的任务,其困难主要体现在以下两个方面:

第一是人脸相似性。所有的人脸都具有相似的结构,在纹理上也比较接近,因而构成了一种在空间上极接近的类别。人脸识别系统只能利用不同人脸之间的细微差别来实现正确的识别任务。

第二是人脸的非刚性性。同一个人在光照、姿态、表情、人脸大小等不同条件下所获得的图像有很大不同,更不用说发型、年龄、化妆以及装饰物的变化了。因此,欲建立一种具有各种条件不变性的描述模型是极其困难的。

然而,人脸识别的复杂性和挑战性并没有影响人们对这一课题的研究热情,这是与人脸识别潜在的应用价值和重要的理论意义是分不开的。

### 1.4.2 人脸识别研究的意义

人脸识别技术结合了认知科学、图象处理、计算机图形学、机器视觉和模式识别等多个研究领域,研究的成果有着广阔的应用前景,意义主要体现在以下几个方面:

首先,随着科学技术的不断发展,人们对安全防范的要求也越来越高,同时,国家相关法律法规也要求以人防、物防、技防相结合的方式加强安全防范工作。在这样的背景下,人脸识别技术也在国家安全、军事安全和公安、司法、民政、

金融、民航、海关、边境、口岸、保险及民用等领域实际应用行到了应用。如：公安布控监控、监狱监控、司法认证、民航安检、口岸出入控制、海关身份验证、银行密押、智能身份证、智能门禁、智能视频监控、智能出入控制、司机驾照验证、各类银行卡、金融卡、信用卡、储蓄卡的持卡人的身份验证，社会保险身份验证等。

其次，人脸识别研究的一个主要内容是人脸特征提取的有效性，即判断并丢弃人脸图像中的不重要的冗余信息甚至噪声，通过尽可能少的却是最具代表性的有用信息来表征并区分不同的人脸，这对于图像压缩及信息传输也具有重要应用价值。以主成分分析算法（PCA）为例，虽然单一的主成分特征用作人脸识别效果有限，但主成分分析算法以其优良的压缩性能以及简单易行的优点，被广泛用于人脸图像向量的最初降维处理，行之有效；另外，本文运用于特征融合的离散余弦变换算法（DCT），正是就是当今成熟的图像压缩技术JPEG标准的理论基础。

再次，人脸识别技术的分支三维人脸识别技术作为一种典型的对于三维自然物体的建模识别技术，自80年代末期开始，目前已经取得了一定的进展。三维人脸建模技术利用深度图像自身的几何特征，利用深度图像处理技术，分析面貌曲面的曲率等几何特征，对面貌曲面进行凹凸区域的分割、正侧面轮廓边缘的提取。三维人脸模型旨在突破二维人脸图像的固有困难：（1）人脸塑性变形（如表情等）的不确定性；（2）人脸模式的多样性（如胡须、发型、眼镜、化妆等）；（3）图像获取过程中的不确定性（如光照的强度、光源方向等）。它的研究思想和实现方法可为其它三维动态自然物体的识别提供重要理论和参考、借鉴。

最后，随着具有人脸识别功能的数码照相机、数码摄像机、手机、笔记本电脑等高科技产品逐渐进入人们的日常生活，人脸识别技术这项尖端科技的优越性已经被大众所认可，社会对此项技术的进一步需求促使越来越多的科研工作者满怀热忱致力于该方面的研究。北京交通大学信息与科学技术研究所阮秋琦教授领导的机器人智能技术课题组就为人脸识别技术的理论研究与实现提供了一个良好的平台，大大推动了人脸识别以及相关人工智能技术的发展。

## 1.5 本论文的内容安排及工作

### 1.5.1 本论文的内容安排

人脸识别是一个非常值得研究的领域。并且，要实现在复杂条件下对人脸有高准确率和快速度的目标，还需要进行长足的研究与改进。作为自动人脸识别系统中的核心人脸识别算法，对全部工作起着至关重要的作用。特别是在复杂外界

条件下,如光照、背景、表情等,都对识别率有着重要影响。在目前的流行算法中,几乎没有完美的在既能保证高识别率又能快速度的识别算法。对此,本论文在以下几个方面对课题进行了研究与创新:

第一,在研究并实现已有的经典统计特征提取算法的基础之上,将 ICA、LDA、DCT 算法进行融合,形成更为有效的人脸特征用于人脸识别。

第二,研究了支持向量机用于人脸识别的技术,提出的 cosh 核函数可以通过调节核函数的参数来使核函数适应人脸特征。

第三,研究参数选择对支持向量机分类效果的影响,将微粒群算法运用于核函数的选择,提高支持向量机参数选择效率。

第四,参与北京交通大学信息所人脸识别系统的设计,将 ICA 特征提取算法和支持向量机运用于实时人脸技术。

### 1.5.2 本论文主要工作

全书共分五章,安排如下:

第一章综述,讨论了人脸识别的研究背景,按照研究内容、技术方法叙述了人脸识别技术的发展历史,并说明了本文的研究内容和结构安排。

第二章人脸特征提取及融合,详细介绍了目前人脸识别研究的主要算法包括 PCA、ICA、LDA、DCT,并将 ICA、LDA、DCT 算法进行融合。

第三章支持向量机分类器,阐述了机器学习基本问题、统计学理论、支持向量机以及支持向量机参数选择,其中提出的基于 cosh 核函数用于人脸识别,并提出基于离散微粒群算法的支持向量机参数选择方法。

第四章人脸识别系统,介绍了系统实现环境,系统流程,并进行了详细演示。证明了本文提出的 cosh 核函数用于实时人脸识别系统的可行性。

第五章全文总结,总结本论文的工作,讨论算法的创新性,对算法和实验的不足之处做总结,对今后的工作提出了建议和想法。

## 2 人脸特征提取及融合

特征抽取是模式识别研究的基本问题之一。对于图像识别而言,抽取有效的图像特征是完成图像识别的首要任务。目前用于图像识别的特征可以分为如下几种:

直观性特征。如图像的边缘、轮廓、纹理或区域等。在人脸识别中用得最多的直观特征是几何特征,即人脸的五官尺寸及相对位置。这类特征虽然具有维数低的优点,但对于同一个人因表情及光照等条件变化往往会导致其尺寸及相对位置发生较大的变化,因此识别效果并不令人满意。

灰度的统计特征。如直方图特征,将图像看作一种二维随机过程,可以引入统计上的各阶距作为特征来描述和分析图像,他们能够在保留主要分类信息的基础上大大降低特征的维数。

变换系数特征。对图像作各种数学变换,可以将变换的系数作为图像的一种特征,例如Fourier变换、Hough变换、Hadamard变换等在图像特征抽取方面均有广泛应用。这类方法的优点是计算比较方便,但其去相关能力不强。系数特征的维数仍较大。

代数特征。它反映图像的一种内在属性。将图像作为矩阵看待,可对其进行各种代数变换,或进行各种矩阵分解。由于矩阵的特征向量反映了矩阵的一种代数属性,并具有不变性,因此可作为图像特征。在人脸识别中,用来抽取图像代数特征的典型方法是矩阵的奇异值分解(SVD)。

### 2.1 PCA 算法

PCA 算法(Principle Component Analysis)是一种主成分分析的算法<sup>[31]</sup>。这种方法将包含人脸图像区域看作一种随机向量,因此可以采用 KL 变换得到正交变换基,对应其中较大的特征值的基底具有与人脸相似的形状。PCA 算法利用这些基底的线性组合可以描述、表达人脸和逼近人脸,因此可以进行人脸的识别和重建。识别过程就是把待识别人脸映射到由特征脸张成的子空间中,与库中人脸的子空间位置进行比较。人脸的重建就是根据待识别人脸在子空间的位置,还原到人脸空间中。

主成分分析法(Principal Component Analysis, 简记为 PCA)最早由 Karl Pearson (1901)所提出,其后由 Hotelling(1933)将其扩展。其是考虑将资料中原有的  $p$  个变数做霍特林(Hotelling)变换或卡胡南-列夫(K-L 变换)得到  $k$  个新变数。此  $k$  个新变



数就是  $k$  个主成份。换种说法, 就是利用样本资料阵, 从原来  $m$  个参量中寻找少数几个既能综合反映原来  $m$  个参量的信息, 且彼此间独立的综合指标的方法, 其作用在于降低参量空间的维数, 提供多维参量的方便表示, 也为其它的多元分析方法提供较少参量的样本, 而又不损失信息。

简单地讲, 求主成分的过程就是求所研究的多变量样本阵  $X$  的协方差阵  $C$  的特征单位向量的过程。

对于某一问题同时可以考虑好几个因素时, 我们并不对这些因素个别处理, 而是将它们总和起来处理, 这就是 PCA。

Turk 等人在 20 世纪 90 年代初提出了经典的特征脸法, 利用主元分析(PCA)进行人脸特征提取, 并取得了较好的效果。

PCA 实质上是 K-L 展开的网络递推实现, K-L 变换是图像压缩中的一种最优正交变换, 其生成矩阵一般为训练样本的总体散布矩阵。特征脸方法就是将包含人脸的图像区域看作是一种随机向量, 因此可以采用 K-L 变换获得其正交 K-L 基底。对应其中较大特征值的基底具有与人脸相似的形状, 因此又称之为特征脸。利用这些基底的线性组合可以描述、表达和逼近人脸图像, 因此可以进行人脸识别和合成。识别过程就是将人脸图像映射到由特征脸张成的子空间上, 比较其与已知人脸在特征脸空间中的位置。

具体算法如下:

第一步, 计算特征脸

设人脸图像  $I(x, y)$  为二维  $N \times N$  灰度图像, 用  $N^2$  维向量  $\Gamma$  表示。人脸图像训练集为

$\{\Gamma_i | i=1, \dots, M\}$ , 其中  $M$  为训练集中图像总数。这  $M$  幅图像的平均值向量为:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i, \quad (2.1)$$

每个人脸  $\Gamma_i$  与平均人脸  $\Psi$  的差值向量  $\Phi_i$  为:

$$\Phi_i = \Gamma_i - \Psi; i=1, \dots, M \quad (2.2)$$

训练图像的协方差矩阵可表示为:

$$C = AA^T, \quad (2.3)$$

其中  $A = [\Phi_1, \dots, \Phi_M]$ 。  $L = A^T A$ 。

特征脸由协方差矩阵  $C$  的正交特征向量组成。对于  $N \times N$  人脸图像, 协方差矩阵  $C$  的大小为  $N^2 \times N^2$ , 对它求解特征值和特征向量是很困难的, 一种取而代之的

方法是解  $M \times M$  个较小的矩阵。首先计算  $M \times M$  矩阵  $L$  的特征向量

$$v_l (l=1, \dots, M): L = A^T A.$$

矩阵  $C$  的特征向量  $u_l (l=1, \dots, M)$  由差值图像  $\Phi_l (l=1, \dots, M)$  与  $v_l (l=1, \dots, M)$  线性组合得到:

$$U = [u_1, \dots, u_M] = [\Phi_1, \dots, \Phi_M] [v_1, \dots, v_M] = AV. \quad (2.4)$$

实际上,  $m(m < M)$  个特征脸足够用于人脸识别。因此, 仅取  $L$  的前  $m$  个最大特征值的特征向量计算特征脸。  $m$  由门限值  $\theta_\lambda$  确定:

$$J = \min_r \{r | \sum_{i=1}^r \lambda_i / \sum_{j=1}^M \lambda_j > \theta_\lambda\} \quad (2.5)$$

## 第二步 基于特征脸的人脸识别

基于特征脸的人脸识别过程由两个阶段组成: 训练阶段和识别阶段。在训练阶段, 每个已知人脸  $\Gamma_k$  映射到由特征脸张成的子空间上, 得到  $m$  维向量  $\Omega_k$ :

$$\Omega_k = U^T (\Gamma_k - \Psi); k=1, \dots, N_c. \quad (2.6)$$

其中  $N_c$  为已知人数。距离门限值  $\theta_c$  定义如下:

$$\theta_c = \frac{1}{2} \max_{j,k} \|\Omega_j - \Omega_k\|; j, k=1, \dots, N_c. \quad (2.7)$$

这里, 主分量构成的投影矩阵  $U$  具有如下特点:

(a) 最小化所有样本的重建误差:

$$E = \min \sum_{i=1}^k \|\Gamma_i - \Psi - UU^T (\Gamma_i - \Psi)\|; \quad (2.8)$$

(b) 最大化样本集在低维投影的方差, 比如, 对于最大特征值的特征向量:

$$U_1 = \arg \max \left| \sum_i \|\Gamma_i U^T\|^2 \right|; \quad (2.9)$$

(c)  $\Gamma_i$  的低维投影  $\Omega_i$  的协方差  $U$  矩阵  $Q = \sum_i \Omega_i \Omega_i^T$  是一个对角阵, 因此主分量分析消除了原始特征的相关性。

在识别阶段, 首先把待识别的图像  $\Gamma$  映射到特征脸空间, 得到向量  $\Omega$ :

$$\Omega = U^T \Gamma - \Psi, \quad (2.10)$$

$\Omega$  与每个人脸集的距离定义为:

$$\epsilon_k^2 = \|\Omega - \Omega_k\|^2; k=1, \dots, N_c. \quad (2.11)$$

为了区分人脸和非人脸, 还需计算原始图像  $\Gamma$  与由特征脸空间重建的图像  $\Gamma_k$  之

间的距离  $\varepsilon$ :

$$\varepsilon_k^2 = \|\Gamma - \Gamma_k\|^2, \text{ 其中: } \Gamma_F = U\Omega + \Psi. \quad (2.12)$$

人脸分类规则如下:

- 若  $\varepsilon \geq \theta_c$ , 则输入图像不是人脸图像;
- 若  $\varepsilon < \theta_c$  且  $\forall k, \varepsilon_k \geq \theta_c$ , 则输入图像包含未知人脸;
- 若  $\varepsilon < \theta_c$  且  $\forall k, \varepsilon_k < \theta_c$ , 则输入图像为库中第  $k$  个人的人脸。

本文并没有直接把主成分作为一种特征运用于人脸识别, 而是利用主成分分析法为 ICA 和 LDA 算法提供一种简单有效的降维方法。提取人脸图像向量的主要成分作为进一步 ICA 和 LDA 算法的输入向量, 摒弃噪声特征, 提高运算的速度和识别率。

## 2.2 ICA 与 Fast ICA 算法

ICA<sup>[37]</sup>方法假设目标信号由各个互相独立的源信号组成。ICA 算法基于数据间独立性最大的度量准则。假设有  $M$  个观察  $x_1, x_2, \dots, x_M$ , 其中每个观察各为若

干个独立源信号  $s_1, s_2, \dots, s_N$  的某种线性组合  $x_i = \sum_{j=1}^N a_{ij}s_j, i=1, 2, \dots, M$ 。令

$X = [x_1, x_2, \dots, x_M]^T$ ,  $S = [s_1, s_2, \dots, s_N]^T$  则有如下矩阵形式:

$$X = AS \quad (2.13)$$

其中  $A \in R^{M \times N}$ , 称为混合矩阵。令

$$S = WX \quad (2.14)$$

其中  $W$  称为分离矩阵。若估计出  $W$ , 则独立源信号可据(5)式求出。ICA算法的核心问题就是估计分离矩阵ICA方法的实现需要选择一个准则度量信号间的独立性。

目前有多个度量信号间独立性的准则, 其中负熵是使用得较多的一个独立性度量准则。考虑最简单的情况, 负熵可使用如下的近似形式:

$$J(y) \approx [E(G(y)) - E(G(v))]^2 \quad (2.15)$$

其中, 对比函数  $G$  可为任意非二次函数, 变量具有零均值和单位协方差, 变量  $v$  服从高斯分布, 且也对应零均值和单位协方差, 符号  $E(\cdot)$  代表数学期望。在实际应用中, 文献分析了以下几种形式的对比函数:

$$G_1(u) = \frac{1}{a_1} \log \cosh(a_1 u) \quad (2.16)$$

$$G_2(u) = -\frac{1}{a_2} \exp(-a_2 u^2 / 2) \quad (2.17)$$

$$G_3(u) = \frac{1}{4} u^4 \quad (2.18)$$

分析表明, 参数  $a_1, a_2$  的取值范围为  $1 \leq a_1 \leq 2, a_2 \approx 1$  对比函数  $G_1, G_2$  的近似效果较好。另外, 对比函数  $G_1$  的定义是一个一般性的定义: 当数据为高度超高斯分布 (super-Gaussian) 时,  $G_2$  是一个较好的选择; 而当估计呈亚高斯分布 (sub-Gaussian) 的独立分量时, 选用  $G_3$ , 较合适。

快速ICA (Fast ICA) <sup>[35][38]</sup> 算法由于具有较优的收敛速度而得到广泛的应用。假设变量  $X$  为零均值矢量,  $x_1, x_2, \dots, x_M$  为  $M$  个样本。令  $X = [x_1, x_2, \dots, x_M]$  则其协方差矩阵可表示为  $C = \frac{1}{N} X X^T$  求出  $C$  的特征向量  $\phi_1, \phi_2, \dots, \phi_n$ , 令矩阵

$\Phi = [\phi_1, \phi_2, \dots, \phi_n]$  则可用下式对  $x$  进行白化变换:

$$t = \Delta^{-1/2} \Phi^T x \quad (2.19)$$

依据牛顿法, FastICA 计算权重矢量  $w$  时采用如下公式进行迭代计算

$$w = E\{tg(w^T t)\} - E\{g'(w^T t)\}w \quad (2.20)$$

且每步迭代后均对  $w$  进行单位化。  $g$  代表对比函数  $G$  的导数函数。FastICA 算法可按如下步骤实现:

- (1) 若数据不对应零均值, 则做中心化, 即将数据与均值向量相减。
- (2) 按(2.19)式对数据进行白化变换
- (3) 选定待求的独立分量的个数  $m$ , 令  $j=1$ 。
- (4) 任意选择一个初始单位矢量  $w_j$
- (5) 令  $w_j = E\{tg(w_j^T t)\} - E\{g'(w_j^T t)\}w_j$  (2.21)
- (6) 若  $j > 1$ , 做如下正交化

$$w_j = w_j - \sum_{k=1}^{j-1} (w_j^T w_k) w_k \quad (2.22)$$

否则, 转下一步

$$(7) \text{ 单位化 } w_j: w_j = w_j / \|w_j\| \quad (2.23)$$

(8) 若  $w_j$ , 不收敛, 则转到步骤(5)。

(9) 令  $j = j + 1$ , 若  $j < m$  则转到步骤(4)

令  $V = [w_1, \dots, w_m]^T$ , 则按下式计算即得分离矩阵

$$W = V \Delta^{-1/2} \Phi \quad (2.24)$$

## 2.3 LDA 算法

在用统计方法解决人脸识别问题时，碰到的问题就是维数问题。通常将一幅包含  $n \times m$  像素的图像看作是  $n \times m$  维空间里的一个点。但是， $n \times m$  维这样的高维空间对于快速而且鲁棒的识别方法是不太可行的。在低维空间里解析上或者计算上行得通的方法，在高维空间里往往行不通。因此，降低维数就成为处理实际问题的关键。线性判别分析，或者称为 Fisher 准则函数就是为了发现这样的投影方向，使得样本类间离散度和样本类内离散度的比值最大。换言之，就是在这样的投影方向，同一个类的样本聚集在一起，而不同类的样本相对比较分散。

度与一个  $K(K > 2)$  类问题，Fisher 准则函数定义的样本类间离散度并不是最优的。这是因为该定义过分强度了那些具有较大类间距离的类别，从而使得最后的投影方向较好的愤慨这些类，但是却造成了其他类别的较大重叠。

LDA 的目的是从高位特征空间里提取出最具有判别能力的低维特征，这些特征能帮助将同一个类别的所有样本聚集在一起，不同类别的样本尽可能的分开，即选择使样本类间离散度和样本类内离散度的比值最大的特征。

设有一个含有  $N$  个样本的集合  $F$ ， $F$  中每一样本  $X$  为  $n$  维向量，设模式类别有  $c$  个： $\omega_1, \omega_2, \dots, \omega_c$  每类有样本  $n_i$  个，它们的总体散度矩阵  $S_t$  (即总体协方差矩阵)、类内散度矩阵  $S_b$  和类间散度矩阵  $S_w$  分别定义如下<sup>[54]</sup>：

$$S_b = \sum_{i=1}^c P(\omega_i)(m_i - m_0)(m_i - m_0)^T \quad (2.25)$$

$$S_w = \sum_{i=1}^c P(\omega_i)E\{(X - m_i)(X - m_i)^T / \omega_i\} \quad (2.26)$$

$$S_t = S_b + S_w = E\{(X - m_0)(X - m_0)^T\} = \sum_{i=1}^N (X - m_0)(X - m_0)^T \quad (2.27)$$

其中， $P(\omega_i) = n_i / N$  为第  $i$  类训练样本的先验概率， $m_i = E\{X / \omega_i\}$  为第  $i$  类训练样本

的均值， $m_0 = E\{X\} = \sum_{i=1}^n P(\omega_i)m_i$  为全体训练样本的均值。

Fisher 鉴别准则函数定义为 (4) 式，另一等价的 Fisher 鉴别准则函数由 (5) 给出：

$$J_f(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T S_w \varphi} \quad (2.28)$$

$$J(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T S_t \varphi} \quad (2.29)$$

其中,  $\varphi$  为任一  $n$  维非零列矢量。

我们希望投影后在低维空间里不同类别的样本尽可能分得开些, 同时希望每个类别内部样本尽量密集。也就是说, 样本类间离散度越大越好, 而样本类内离散度越小越好。因此, 如果  $S_w$  是非奇异矩阵, 最优的投影方向  $W_{opt}$  就是使得样本类间离散度矩阵和样本类内离散度矩阵的行列式比值最大的那些正交特征向量。因此, Fisher 准则函数定义为

$$W_{opt} = \arg \max \frac{|W^T S_b W|}{|W^T S_w W|} = [w_1, w_2, \dots, w_n] \quad (2.30)$$

根据线性代数理论, 我们知道  $W_{opt}$  就是满足如下等式的解

$$S_b W_i = \lambda S_w W_i \quad (i = 1, 2, \dots, m) \quad (2.31)$$

也就是对应于矩阵  $S_w^{-1} S_b$  较大的特征值  $\lambda_i$  的特征向量。该矩阵最多只有  $C-1$  个非零特征值,  $C$  是类别数目。

## 2.4 DCT 算法

给定长度为  $N$  的输入序列  $u(n)$  它的 DCT 变换可由下式得到<sup>[55][56]</sup>

$$v(k) = \alpha(k) \sum_{n=0}^{N-1} u(n) \cos\left(\frac{(2n+1)\pi k}{2N}\right) \quad 0 \leq k \leq N-1 \quad (2.33)$$

式中:  $\alpha(0) = \sqrt{\frac{1}{N}}, \alpha(k) = \sqrt{\frac{2}{N}}, 1 \leq k \leq N-1$

式中可以把  $u(n)$  序列视为一个向量, 把 DCT 看作一个变换矩阵, 得到输出向量  $v(k)$ , 即

$$v = Cu \quad (2.34)$$

式中:  $C$  为 DCT 变换矩阵

$$C(k,n) = \begin{cases} \frac{1}{\sqrt{N}} & \text{其中 } k=0, 0 \leq n \leq N-1, \\ \sqrt{\frac{2}{N}} \cos\left(\frac{(2n+1)\pi k}{2N}\right), & \\ 1 \leq k \leq N-1, 0 \leq n \leq N-1 \end{cases} \quad (2.35)$$

式中,  $k, n$  为变换矩阵的行和列。

得到的  $v$  即为所需要的特征向量。还可以利用DCT的逆变换, 将原图像进行复原

$$u = C^{-1}v \quad (2.36)$$

$C^{-1}$  为  $C$  的逆矩阵。

有以上DCT的理论基础, 可以看到DCT在以下几点优于特征脸方法: 1) 每一幅图像矩阵看成一个向量, DCT可以对每个向量进行单独处理, 提取出特征向量, 放入特征库里面, 用于后面的识别; 2) 当新的图像需加入库中, 经过预处理之后, 可以对该图像进行DCT变换, 无需考虑其他的图像。这样即使库中人脸图像很多时, 计算量也不大。而特征脸方法当有新的图像加入时, 必须重新计算新图像与原库中所有图像的协方差矩阵, 以求出新的特征空间。

## 2.5 信息融合

### 2.5.1 信息融合概念

通过研究可以发现, 对于复杂模式识别问题, 如手复杂背景下的人脸识别, 可以说目前还没有一种简单的方法可以达到较高的识别率, 每一种方法都有各自的优点、缺陷和不同的适用范围, 不同的特征和分类方法之间具有一定的互补性。因此, 研究如何将不同的方法有机地结合起来以充分发挥各自的优势, 克服其缺陷, 从而构成信息融合型的识别系统, 就成为当前模式识别研究的一个主要方向。

信息融合是针对一个系统中使用多种传感器这一特定问题而展开的一种信息处理的新研究方向, 因此, 信息融合又可称作多传感器融合(MSF)。信息融合比较确切的定义可概括为利用计算机技术对按时序获得的若干传感器的观测信息在一定准则下加以自动分析、综合以完成所需的决策和估计任务而进行的信息处理过程。

信息融合的基本目标是通过信息组合而不是出现在输入信息中的任何个别

元素，推导出更多的信息，这是最佳协同作用的结果，即利用多个传感器共同或联合操作的优势提高传感器系统的有效性。现在，信息融合技术已经成功地应用于众多的研究领域，这些领域主要包括：机器人和智能仪器系统、智能制造系统、战场任务与无人驾驶飞机、航天应用目标检测与跟踪图像分析与理解惯性导航自动目标识别多源图像复合等。

IEEE 国际遥感数据融合技术委员会将遥感数据融合分为数据、特征和决策等三级融合<sup>[57][58]</sup>。

2.5.2 数据层融合

数据层融合<sup>[59]</sup>是直接在采集到的原始数据层上进行的融合。在各种传感器的原始信息在未经预处理之前就进行数据的综合和分析。这是最低层次的融合，如成像传感器中通过对包含若干特征的模糊图像进行图像处理和模式识别来确认目标属性的过程就属于数据层融合。数据层融合强调图像之间的配准精度。通常用于复合图像分析和理解等方面。

数据级融合的优点是能保持尽可能多的现场数据。提供其他融合层次所不能提供的细微信息。缺点是没有有效的方法进行特性的一致性检验，具有一定的盲目性。信息融合原则上不赞成在数据层上直接进行，主要局限在于：

- (1) 所要处理的传感器数量大，处理代价高、时间长、实时性差。
- (2) 约数据通信量较大，抗干扰能力较差。
- (3) 数据级融合是在信息的最低层进行的，由于传感器原始信息的不确定性、不完全性和不稳定性，要求在数据级融合时有较高的纠错能力。
- (4) 各传感器信息之间具有的精确度要达到一个像素的校准精度，因此各

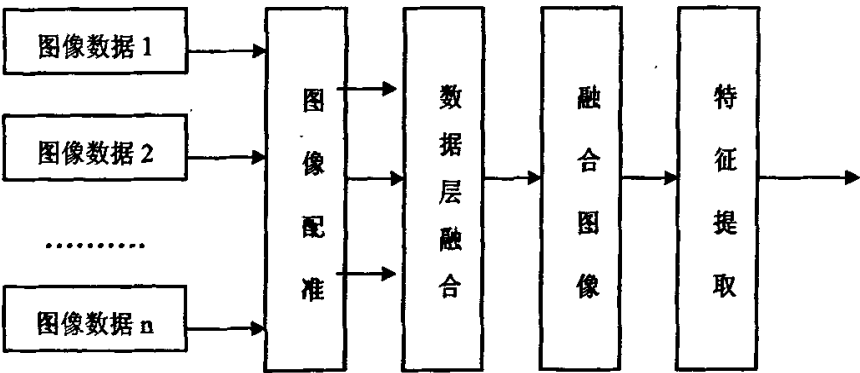


图 2.1 数据层融合  
Figure 2.1 Data Level Merge



传感器信息必须来自同质传感器。

从发展过程来看，基于像素的融合方法包括简单的融合法、金字塔融合法和小波融合法；从作用域来看，可以分为空间域融合法和空间变换域融合法两大类。

2.5.3 特征层融合

特征层融合<sup>[62]</sup>是利用从各个传感器的原始信息中提取的特征信息进行综合分析和处理的中间层次过程。通常所提取的特征信息应是原始信息的充分表示量或统计量，据此对多传感器信息进行分类、汇集和综合。特征层融合可以有效地改善识别性能，多特征提取可以提供比单特征提取更多的针对识别目标的特征信息，增大特征空间维数，然而其技术却具有一定的复杂性。

特征级融合划分为两大类：目标状态数据融合和目标特性融合。

(1) 特征级目标状态数据融合

特征级目标状态数据融合主要用于多传感器目标跟踪领域。融合系统首先对传感器数据进行预处理以完成数据校准，然后实现参数相关的状态向量估计。

(2) 特征级目标特性融合

特征级目标特性融合就是特征层联合识别，该融合方法是模式识别的相应技术，只是在融合前必须先对特征进行相关处理，把特征向量分成有意义的组合。

特征级融合要求从原始影响数据提取特征，可提取的特征主要包括四类①几何特征；②结构特征；③统计特征；④光谱特征。本文是对统计特征进行融合。

属于这类技术的方法有 Bayes 和 Dempster-Shafer 推理、带权平均法、神经网络法、聚类分析法、表决法和小波变换融合法等。

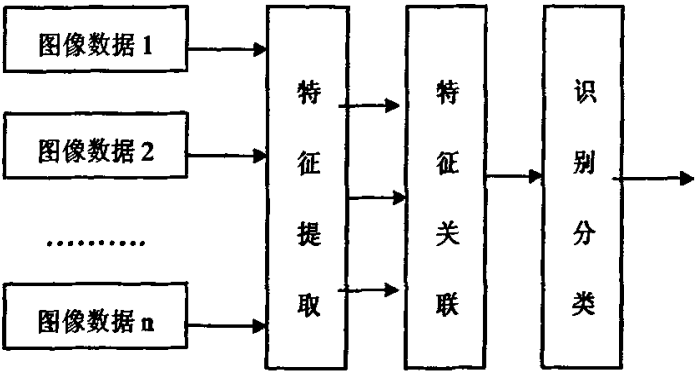


图 2.2 特征层融合

Figure 2.2 Feature Level Merge

## 2.5.4 决策层融合

决策层融合<sup>[60][61]</sup>是在信息表示的最高层次上进行的融合处理。不同类型的传感器观测同一个目标,每个传感器各自完成判断识别,得出对所观察目标的初步结论,然后通过相关处理、决策级融合判决,最终获得联合推断结果,从而直接为决策提供依据。因此决策级融合是直接针对具体决策目标,充分利用特征层融合所得出的各类目标特征信息,并给出简明而直观的结果。

决策级融合主要优点是:

- (1) 融合中心处理代价低,具有很高的灵活性。
- (2) 通信量小,抗干扰能力强。
- (3) 当一个或几个传感器出现错误时,通过适当的融合,系统还能获得正确结果,所以具有容错性。
- (4) 对传感器的依赖性小,传感器可以是同质的,也可以是异质的。
- (5) 系统对信息传输带宽要求较低。
- (6) 能有效地反映环境或目标各个侧面的不同类型信息。

决策级融合的缺点是要对原传感器信息进行预处理以获得各自的判定结果,所以预处理代价高。

基于决策的融合技术分为两类:

- (1) 基于知识的决策技术,包括采用专家系统、神经网络和模糊逻辑等的决策融合;
- (2) 基于识别的决策技术。

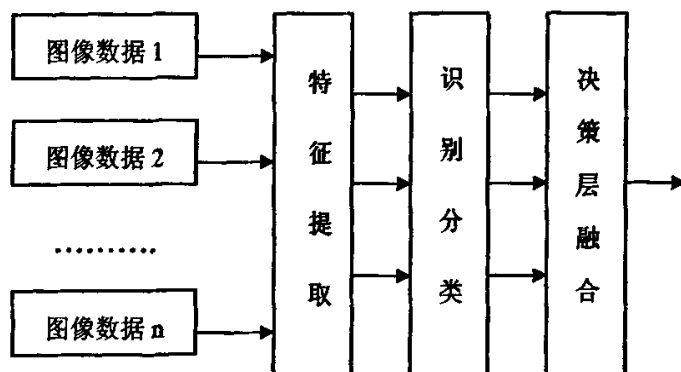


图 2.3 决策层融合

Figure 2.3 Decision Level Merge

2.5.5 人脸图像特征层融合

目标的识别分类有很多种方法,但获取的特征信息主要从不同的信号处理技术手段得到,如特征空间投影、经典谱分析、小波分析、高阶统计量分析等,这些技术手段只是根据人脸图像的部分信息进行相应的识别分类,各种单一特征难免具有局限性,所以识别分类的效果很难令人满意。而运用信息融合技术,通过对各传感器采集得到的各种各样的、大量的信息进行多层次的、综合处理,从而最大限度地获得人脸图像的状态、特征等完整信息,进而产生有意义的、新的融合信息,则可提高人脸图像的识别分类效果,为人脸识别技术的智能化发展提供研究基础。本文研究的目的在于采用多特征融合的信息融合进行识别分类,从而提高人脸识别分类的概率。

由试验结果可知,利用本章以上提到的单一特征所构建的人脸识别系统的识别率均不理想,如表 2.1 所示,以 ORL 人脸库为例,共 40 个人,每人 10 幅图像,前 5 幅用于训练,后 5 幅用作检测。不同的特征提取方法 (PCA, ICA, LDA) 与不同的分类器 (最小距离分类器, 马氏距离分类器, 余弦分类器, 不同核函数的支持向量机分类器) 结合, 识别率在 90%上下浮动, 难以达到实用要求。而作为 PCA 改进方法的 KPCA 算法, 虽然采用了目前流行的核方法, 但以计算复杂度为代价得到的识别率提升却十分有限。

表 2.1 分类器与特征提取方法组合  
Table 2.1 Classifiers Combined with Features

	最小距离	马氏距离	cos	SVM(rbf)	SVM(poly)	SVM(liner)
PCA	83.50%	<u>88.00%</u>	83.00%	84.50%	66.50%	53.00%
KPCA	<u>90.00%</u>	87.50%	83.50%	89.50%	89.00%	89.00%
ICA	86.50%	89.00%	71.50%	<u>90.50%</u>	87.00%	90.00%
LDA	89.50%	89.00%	83.50%	89.00%	91.50%	<u>92.00%</u>

为了改善特征质量,笔者曾尝试了一些特征处理的方法,如将人脸图像切割后提取眼部、嘴部局部特征;或是对人脸图像加载蒙板突出局部特征的方法,但由于这些局部特征方法破坏了人脸图像的整体特征,使得特征而且器官定位不准确也使得识别效果不甚理想,甚至是画蛇添足。通过之前的理论描述本文认为,这三种特征的识别率相差不多但每种特征的侧重点却完全不同: DCT 特征相对其它两种特征而言更加关注人脸的轮廓特征,实验得到的最佳维数为 20; ICA 特征是通过将人脸图像向量投影到一个特征子空间,从而可以选择性的保留更多的细节部分信息,其最佳维数为 135; LDA 特征以追求可分性最优为基础一向都

被广泛使用，本文中维数为 50。因此，本文尝试性的提出了基于 LDA+DCT+ICA 的特征层融合技术，对不同的方法提取的原始特征向量经过归一化处理之后线性组合，形成具有一定的互补性的融合特征向量，综合利用了单一特征的有效鉴别信息，并且弥补了单一特征的局限性，提高了特征的质量从而提高了识别率。

如表 2.2 所示，以 ORL 人脸库为例，共 40 个人，每人 10 幅图像，前 5 幅用于训练，方后 5 幅用作检测。其中每一种特征提取以下的实验结果清楚的表明了本文特征融合得优势，由于特征融合的影响，表 2.2 中的识别率普遍高于表 2.1 中相应的识别率，两种特征融合的识别率至少为 92.00%；三种特征融合的识别率至少为 93.50%；而最优识别率由单一特征条件下的 92.00%提高到 97.50%，提高了 6%。（由下划线标出）。

本文认为，人脸图像特征提取在人脸识别技术中具有决定性意义，其重要程度甚至要高于分类器的设计。在特征提取阶段，所得到的特征中有效信息是否突出，是否可以从多个角度全面反映原有信息的特点，特征中的冗余信息甚至噪声被丢弃得是否彻底，以及特征数据的大小即特征向量维数的多少都决定了特征提取的成功与否。良好的特征是高识别率的保障，否则，不论分类器多么优秀识别率都无法得到重大的改善，这一点从表 2.1 和表 2.2 数据比较即可明显的获得。

信息融合技术是一项复杂研究，本文从特征融合得的角度入手，将 LDA+DCT+ICA 的特征线性组合作为人脸特征用于识别，对于特征质量的改善以及识别率的提升已经比较明显，可见特征融合以及信息融合技术确实是一种行之有效的特征处理方法，应当得到足够的重视与进一步的研究。

表 2.2 特征融合识别率  
Table 2.2 Recognition based on Feature Merge

	最小距离	马氏距离	cos	SVM(rbf)	SVM(poly)	SVM(liner)
LDA+DCT	93.00%	94.00%	92.00%	96.00%	93.00%	<u>96.50%</u>
LDA+DCT+ICA	93.50%	96.00%	95.00%	<u>97.50%</u>	97.00%	97.00%

2.6 本章小结

本章从理论上分析了几种经典的基于统计的人脸特征提取技术，并提出了本文的一个创新点：基于 ICA、LDA、DCT 的人脸特征融合。

基于统计特征的方法尽管还存在很多不足，但总的来说在人脸识别中是有效可行的，而且作为统计模式识别的方法还具有实现方便，直观等的优点。所以本文将算法研究的重点放在基于统计特征的算法上，进行创新和改进。实验表明，

ICA、LDA、DCT 的人脸特征融合可以很好的表征人脸特征，作为进一步的人脸分类与识别奠定了良好的基础，并提出了加强特征融合研究的必要性。

### 3 支持向量机分类器

#### 3.1 概述

统计模式识别问题可以看作是一个更广义问题的特例,就是基于数据的机器学习问题。基于数据的机器学习问题是现代智能技术中十分重要的一个方面,主要研究如何从一些观察数据(样本)出发得出目前尚不能通过原理分析得到的规律,利用这些规律去分析客观对象,对未来数据或无法观测的数据进行预测。现实世界中存在大量我们尚无法准确认识但却可以进行观测的事物,因此这种机器学习在从现代科学、技术到社会、经济等各领域都有着十分重要的应用。当我们把要研究的规律抽象成分类关系时,这种机器学习问题就是模式识别。

迄今为止,关于机器学习还没有一种被共同接受的理论框架,关于其实现方法大致可以分为三种:第一种是经典的(参数)统计估计方法。包括模式识别、神经网络等在内,现有机器学习方法共同的重要理论基础之一是统计学。参数方法正是基于传统统计学的,在这种方法中,参数的相关形式是已知的,训练样本用来估计参数的值。这种方法有很大的局限性,首先,它需要已知样本分布形式,这需要花费很大代价,还有,传统统计学研究的是样本数目趋于无穷大时的渐近理论,现有学习方法也多是基于此假设。但在实际问题中,样本数往往是有限的,因此一些理论上很优秀的学习方法实际中表现却可能不尽人意。第二种方法是经验非线性方法,如人工神经网络(ANN)。这种方法利用已知样本建立非线性模型,克服了传统参数估计方法的困难。但是,这种方法缺乏一种统一的数学理论。

与传统统计学相比,统计学习理论(Statistical Learning Theory 或 SLT)是一种专门研究小样本情况下机器学习规律的理论。该理论针对小样本统计问题建立了一套新的理论体系,在这种体系下的统计推理规则不仅考虑了对渐近性能的要求,而且追求在现有有限信息的条件下得到最优结果。V. Vapnik 与其领导的贝尔实验室的小组从六、七十年代开始致力于此方面研究,到九十年代中期,随着其理论的不断发展和成熟,也由于神经网络等学习方法在理论上缺乏实质性进展,统计学习理论开始受到越来越广泛的重视。由于统计学习理论是建立在一套较坚实的理论基础之上的,为解决有限样本学习问题提供了一个统一的框架。它能将很多现有方法纳入其中,有望帮助解决许多原来难以解决的问题(比如神经网络结构选择问题、局部极小点问题等);同时,在这一理论基础上发展了一种新的通用学习方法——支持向量机(Support Vector Machines 或 SVM),它已初步表现出很多优于已有方法的性能。一些学者认为,SLT 和 SVM 正在成为继神经网络研究之后新的研究热点,并将有力地推动机器学习理论和技术的发展。我国早在

八十年代末就有学者注意到统计学习理论的基础成果。

支持向量机方法是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的,根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度, Accuracy)和学习能力(即无错误地识别任意样本的能力)之间寻求最佳折衷,以期获得最好的推广能力(Generalization Ability)。支持向量机方法的几个主要优点有:

1)它是专门针对有限样本情况的,其目标是得到现有信息下的最优解而不仅仅是样本数趋于无穷大时的最优值。

2)算法最终将转化成为一个二次型寻优问题,从理论上说,得到的将是全局最优解解决了在神经网络方法中无法避免的局部极值问题。

3)算法将实际问题通过非线性变换转换到高维的特征空间(Feature Space),在高维空间中构造线性判别函数来实现原空间中的非线性判别函数,其特殊性能保证机器有较好的推广能力,同时它巧妙地解决了维数问题,其算法复杂度与样本维数无关。

统计学习理论从七十年代末诞生,到九十年代之前都处在初级研究和理论准备阶段,近几年才逐渐得到重视,其本身也趋向完善,并产生了支持向量机这一将这种理论付诸实现的有效的机器学习方法。目前, SVM 算法在模式识别、回归估计、概率密度函数估计等方面都有应用。例如,在模式识别方面,对于手写数字识别、语音识别、文章分类、基因识别等问题, SVM 算法在精度上已经超过传统的学习算法或与之不相上下。

支持向量机是建立在统计学习理论的基础上一种通用学习方法,因此先对统计学习理论做一个简单介绍,给出 SVM 的理论基础。

## 3.2 机器学习的基本问题

### 3.2.1 机器学习理论

机器学习<sup>[63][64]</sup>(Machine Learning)是研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。它是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域,它主要使用归纳、综合而不是演绎。

机器学习是关于理解与研究学习的内在机制、建立能够通过学习自动提高自身水平的计算机程序的理论方法的学科。近年来机器学习理论在诸多应用领域得到成功的应用与发展,已成为计算机科学的基础及热点之一。采用机器学习方法的计算机程序被成功用于机器人下棋程序、语音识别、信用卡欺诈监测、自主车

辆驾驶、智能机器人等应用领域,除此之外机器学习的理论方法还被用于大数据集的数据挖掘这一领域。实际上,在任何有经验可以积累的地方,机器学习方法均可发挥作用。

学习能力是智能行为的一个非常重要的特征,但至今对学习的机理尚不清楚。人们曾对机器学习给出各种定义。H.A.Simon 认为,学习是系统所作的适应性变化,使得系统在下次完成同样或类似的任务时更为有效。R.s.Michalski 认为,学习是构造或修改对于所经历事物的表示。从事专家系统研制的人们则认为学习是知识的获取。这些观点各有侧重,第一种观点强调学习的外部行为效果,第二种则强调学习的内部过程,而第三种主要是从知识工程的实用性角度出发的。

机器学习在人工智能的研究中具有十分重要的地位。一个不具有学习能力的智能系统难以称得上是一个真正的智能系统,但是以往的智能系统都普遍缺少学习的能力。例如,它们遇到错误时不能自我校正;不会通过经验改善自身的性能;不会自动获取和发现所需要的知识。它们的推理仅限于演绎而缺少归纳,因此至多只能证明已存在事实、定理,而不能发现新的定理、定律和规则等。随着人工智能的深入发展,这些局限性表现得愈加突出。正是在这种情形下,机器学习逐渐成为人工智能研究的核心之一。它的应用已遍及人工智能的各个分支,如专家系统、自动推理、自然语言理解、模式识别、计算机视觉、智能机器人等领域。其中尤其典型的是专家系统中的知识获取瓶颈问题,人们一直在努力试图采用机器学习的方法加以克服。

机器学习的研究是根据生理学、认知科学等对人类学习机理的了解,建立人类学习过程的计算模型或认识模型,发展各种学习理论和学习方法,研究通用的学习算法并进行理论上的分析,建立面向任务的具有特定应用的学习系统。这些研究目标相互影响相互促进。

机器学习已经有了十分广泛的应用例如搜索引擎、医学诊断、检测信用卡欺诈、证券市场分析、DNA 序列测序、语音和手写识别、战略游戏和机器人运用。

自从 1980 年在卡内基-梅隆大学召开第一届机器学术研讨会以来,机器学习的研究工作发展很快,已成为中心课题之一。

目前,机器学习领域的研究工作主要围绕以下三个方面进行:

- (1) 面向任务的研究 研究和分析改进一组预定任务的执行性能的学习系统。
- (2) 认知模型 研究人类学习过程并进行计算机模拟。
- (3) 理论分析 从理论上探索各种可能的学习方法和独立于应用领域的算法。



机器学习是继专家系统之后人工智能应用的又一重要研究领域,也是人工智能和神经计算的核心研究课题之一。现有的计算机系统和人工智能系统没有什么学习能力,至多也只有非常有限的学习能力,因而不能满足科技和生产提出的新要求。本章将首先介绍机器学习的定义、意义和简史,然后讨论机器学习的主要策略和基本结构,最后逐一研究各种机器学习的方法与技术,包括机械学习、基于解释的学习、基于事例的学习、基于概念的学习、类比学习和基于训练神经网络的学习等。对机器学习的讨论和机器学习研究的进展,必将促使人工智能和整个科学技术的进一步发展。

### 3.2.2 机器学习的定义和研究意义

学习是人类具有的一种重要智能行为,但究竟什么是学习,长期以来却众说纷纭。社会学家、逻辑学家和心理学家都各有其不同的看法。按照人工智能大师西蒙的观点,学习就是系统在不断重复的工作中对本身能力的增强或者改进,使得系统在下一次执行同样任务或类似任务时,会比现在做得更好或效率更高。西蒙对学习给出的定义本身,就说明了学习的重要作用。

1959 年美国的塞缪尔(Samuel)设计了一个下棋程序,这个程序具有学习能力,它可以在不断的对奕中改善自己的棋艺。4年后,这个程序战胜了设计者本人。又过了3年,这个程序战胜了美国一个保持8年之久的常胜不败的冠军。这个程序向人们展示了机器学习的能力,提出了许多令人深思的社会问题与哲学问题。

机器的能力是否能超过人的,很多持否定意见的人的一个主要论据是:机器是人造的,其性能和动作完全是由设计者规定的,因此无论如何其能力也不会超过设计者本人。这种意见对不具备学习能力的机器来说的确是对的,可是对具备学习能力的机器就值得考虑了,因为这种机器的能力在应用中不断地提高,过一段时间之后,设计者本人也不知它的能力到了何种水平。

至今,还没有统一的“机器学习”定义,而且也很难给出一个公认的和准确的定义。为了便于进行讨论和估计学科的进展,有必要对机器学习给出定义,即使这种定义是不完全的和充分的。顾名思义,机器学习是研究如何使用机器来模拟人类学习活动的一门学科。稍为严格的提法是:机器学习是一门研究机器获取新知识和新技能,并识别现有知识的学问。这里所说的“机器”,指的就是计算机;现在是电子计算机,以后还可能是中子计算机、光子计算机或神经计算机等等。

### 3.2.3 机器学习的发展史

机器学习是人工智能研究较为年轻的分支，它的发展过程大体上可分为 4 个时期。

第一阶段是在 50 年代中叶到 60 年代中叶，属于热烈时期。

第二阶段是在 60 年代中叶至 70 年代中叶，被称为机器学习的冷静时期。

第三阶段是从 70 年代中叶至 80 年代中叶，称为复兴时期。

机器学习的最新阶段始于 1986 年。机器学习进入新阶段的重要表现在下列诸方面：

(1) 机器学习已成为新的边缘学科并在高校形成一门课程。它综合应用心理学、生物学和神经生理学以及数学、自动化和计算机科学形成机器学习理论基础。

(2) 结合各种学习方法，取长补短的多种形式的集成学习系统研究正在兴起。特别是连接学习符号学习的耦合可以更好地解决连续性信号处理中知识与技能的获取与求精问题而受到重视。

(3) 机器学习与人工智能各种基础问题的统一性观点正在形成。例如学习与问题求解结合进行、知识表达便于学习的观点产生了通用智能系统 SOAR 的组块学习。类比学习与问题求解结合的基于案例方法已成为经验学习的重要方向。

(4) 各种学习方法的应用范围不断扩大，一部分已形成商品。归纳学习的知识获取工具已在诊断分类型专家系统中广泛使用。连接学习在声图文识别中占优势。分析学习已用于设计综合型专家系统。遗传算法强化学习在工程控制中有较好的应用前景。与符号系统耦合的神经网络连接学习将在企业的智能管理与智能机器人运动规划中发挥作用。

(5) 与机器学习有关的学术活动空前活跃。国际上除每年一次的机器学习研讨会外，还有计算机学习理论会议以及遗传算法会议。

### 3.2.4 机器学习系统的基本结构

图 3.1 表示学习系统的基本结构。环境向系统的学习部分提供某些信息，学习部分利用这些信息修改知识库，以增进系统执行部分完成任务的效能，执行部分根据知识库完成任务，同时把获得的信息反馈给学习部分。在具体的应用中，环境，知识库和执行部分决定了具体的工作内容，学习部分所需要解决的问题完全由上述 3 部分确定。下面我们分别叙述这 3 部分对设计学习系统的影响。

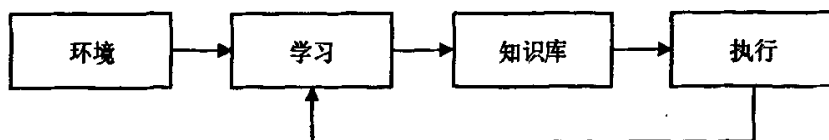


图 3.1 学习系统的基本结构

Figure 3.1 Structure of Machine Learning

影响学习系统设计的最重要的因素是环境向系统提供的信息。或者更具体地说是信息的质量。知识库里存放的是指导执行部分动作的一般原则，但环境向学习系统提供的信息却是各种各样的。如果信息的质量比较高，与一般原则的差别比较小，则学习部分比较容易处理。如果向学习系统提供的是杂乱无章的指导执行具体动作的具体信息，则学习系统需要在获得足够数据之后，删除不必要的细节，进行总结推广，形成指导动作的一般原则，放入知识库，这样学习部分的任务就比较繁重，设计起来也较为困难。

因为学习系统获得的信息往往是不完全的，所以学习系统所进行的推理并不完全是可靠的，它总结出来的规则可能正确，也可能不正确。这要通过执行效果加以检验。正确的规则能使系统的效能提高，应予保留；不正确的规则应予修改或从数据库中删除。

知识库是影响学习系统设计的第二个因素。知识的表示有多种形式，比如特征向量、一阶逻辑语句、产生式规则、语义网络和框架等等。这些表示方式各有其特点，在选择表示方式时要兼顾以下 4 个方面：

(1)表达能力强。(2)易于推理。(3)容易修改知识库。(4)知识表示易于扩展。

对于知识库最后需要说明的一个问题是学习系统不能在全然没有任何知识的情况下凭空获取知识，每一个学习系统都要求具有某些知识理解环境提供的信息，分析比较，做出假设，检验并修改这些假设。因此，更确切地说，学习系统是对现有知识的扩展和改进。

执行部分是整个学习系统的核心，因为执行部分的动作就是学习部分力求改进的动作。同执行部分有关的问题有 3 个：复杂性、反馈和透明性。

### 3.3 统计学理论

#### 3.3.1 VC 维

VC 维的定义如下：对于一个指示函数集，如果存在  $h$  个样本能够被函数集中的函数按所有可能的  $2$  的  $h$  次幂种形式分开，则称函数集能够把  $h$  个样本都打散， $h$  的最大值就是函数集的 VC 维。VC 维是统计学习理论 (SLT) 中的一个重要概念，它是函数集学习性能的重要指标。目前尚没有通用的关于任意函数集 VC 维计算的理论，只知道一些特殊的函数集的 VC 维。比如，在  $n$  维空间中线性分类器和线性实函数的 VC 维是  $n+1$ ，而  $f(x, \alpha) = \sin(\alpha x)$  的 VC 维则为无穷大。对于给定的学习函数集，如何（用理论或实验的方法）计算其 VC 维是当前统计学习理论中有待研究的一个问题<sup>[63]</sup>。

#### 3.3.2 推广性的界

SLT (Statistical Learning Theory) 中关于经验风险和实际风险之间的关系结论即为推广性的界。推广性的界是建立在 VC 维基础上的，关于二值分类问题，结论是：对于指示函数集中的所有函数（包括使经验风险最小的函数），经验风险  $R_{emp}(w)$  与实际风险  $R(W)$  之间以至少  $1-n$  的概率满足下式

$$R(W) \leq R_{emp}(w) + \sqrt{\frac{h(\ln(\frac{2n}{h}) + 1) - \ln(\frac{\eta}{4})}{n}} \quad (3.1)$$

式中， $h$  为函数集的 VC 维， $n$  为样本数。

这一结论从理论上说明了学习机器的实际风险由两部分组成：一是经验风险（训练误差），另一部分称为置信范围，它与学习机器的 VC 维及训练样本数有关。其关系可以简单地表示为式：

$$R(W) \leq R_{emp}(w) + \phi\left(\frac{n}{\eta}\right) \quad (3.2)$$

它表明在有限训练样本下，学习机器 VC 维越高（复杂性越高）则置信范围越大，导致真实风险与经验风险之间可能差别就越大。这就是为什么会出现过学习现象地原因。机器学习过程不但要使经验风险最小，还要使 VC 维尽量小，以缩小置信范围，才能取得较小地实际风险，即对未来样本有较好地推广性。

需要指出，推广性的界是对于最坏情况的结论，在很多情况下是较松的，尤其当 VC 维较高时更是如此。当  $h/n > 0.37$  时这个界肯定是松弛的，当 VC 维无

穷大时这个界就不再成立。而且,这种界只在对同一类学习函数进行比较时有效,可以指导我们从函数集中选择最优的函数,在不同函数集之间比较却不一定成立。Vapnik 指出,寻找更好地反映学习机器能力的参数和得到更紧的界是学习理论今后的研究方向之一。

### 3.3.3 机构化风险最小化原则

ERM 原则在样本有限时是不合理的。实际上需要同时最小化经验风险和置信范围。其实,在传统方法中,学习模型和算法的过程就是调整置信范围的过程,如果模型比较适合现有的训练样本,则可以取得较好的效果。但是因为缺乏理论指导,这种选择只能依赖先验知识和经验,造成了如神经网络等方法对使用者“技巧”的过分依赖。

统计学习理论提出了一种新的策略,即把函数集构造为一个函数子集序列,使各个子集按照 VC 维的大小(即自信范围的大小)排列;在每个子集中寻找最小经验风险,在子集间折中考虑经验风险和置信范围,取得实际风险的最小值。这种思想称作结构风险最小化(Structural Risk Minimization, 结构风险最小化)即 SRM 原则。统计学习理论还给出了合理的函数子集结构应满足的条件及在 SRM 原则下实际风险收敛的性质。

实现 SRM 原则有 2 种思路:

- 1) 在每个子集中求最小经验风险,然后选择使最小经验风险和置信范围和最小的子集。显然这种方法比较费时,当子集数目很大甚至无穷时是不可行的。
- 2) 设计函数集的某种结构,使每个函数集都能够取得最小的经验风险(如使经验误差为 0),然后只需选择适当的子集使置信范围最小,函数就是最优函数。

## 3.4 支持向量机

支持向量机<sup>[46][63]</sup> (SVM, Support Vector Machines)方法是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的,根据有限的样本信息在模型的复杂性,即对特定训练样本的学习精度(Accuracy)和学习能力(即无错误地识别任意样本的能力)之间寻求最佳折衷,以期获得最好的推广能力(Generalization Ability)。在 SVM 方法中,只要定义不同的内积函数(或者称为核函数),就可以实现许多不同的学习算法。

### 3.4.1 最优分类超平面

支持向量机是从数据分类问题的研究中发展而来的,在数据分类问题中,如果采用通常的神经网络方法,可以简单地描述为:系统随机产生一个超平面并移动它,直到数据集中属于不同类的点正好位于超平面的不同侧面。这种处理机制决定了采用神经网络进行数据分类最终获得的分类超平面将相当靠近训练集中的点,在绝大多数情况下,它并不是最优解。而 SVM 考虑寻找一个满足分类要求的超平面,并且使训练集中的点距离分类面尽可能的远,也就是寻找一个分类面使它两侧的空白区域(margin)最大<sup>[66]</sup>。

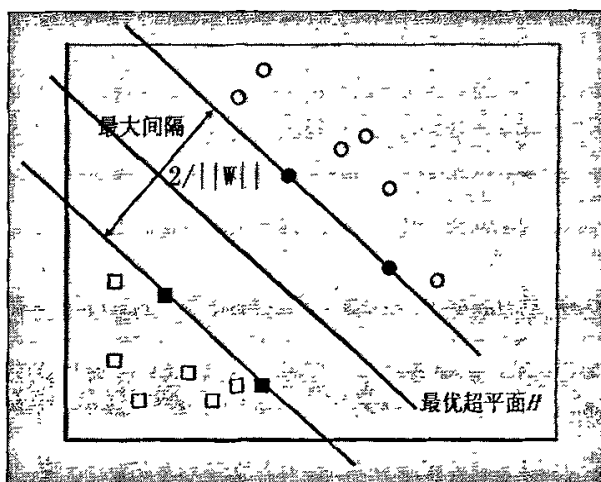


图3.1 最优分类超平面

Figure 3.1 Optimal Hyperplane

如图 3.1 所示,方点和圆点代表两类样本, $H$  为把两类样本正确地分开的分类线。 $H_1, H_2$  分别为通过各类样本中离分类线最近的点且平行于分类线的直线,也就是要求分类线不但能将两类样本正确地分开,而且要使  $H_1$  和  $H_2$  之间的间隔最大。分类线方程为

$$x \cdot w + b = 0 \quad (3.3)$$

对它归一化,使对线性可分样本集  $(x_i + y_i)$ ,  $i = 1, \dots, n, x \in R^d$ ,  $y \in \{1, -1\}$ , 满足

$$y_i[(w \cdot x_i) + b] - 1 \geq 0, i = 1, \dots, n. \quad (3.4)$$

这样分类间隔就等于  $2/\|w\|$ , 因此使分类间隔最大就等价于最小化  $\|w\|$ 。因此满足(3.4)并且使  $\|w\|$  最小的分类面就是最优分类面,  $H_1$  和  $H_2$  上的训练样本就是支持向量(图 3.1 中的实心的圆点和方点), 满足 (3.4) 的等号。

实际上使分类间隔最大就是对推广能力的控制,这是 SVM 的核心思想之一。统计学习理论指出,在  $N$  维空间中,设样本分布满足  $\|x\| \leq R$ ,  $\|w\| \leq A$ , 则指示

函数集  $f(x, w, b) = \text{sgn}\{(w \cdot x) + b\}$  的 VC 维满足下面的界

$$h \leq \min([R^2 A^2], N) + 1 \quad (3.5)$$

从式(3.5)可以看出最小化  $\|w\|$  就是使 VC 维的上界最小, 从而实现结构风险最小化准则对函数复杂性的选择。

利用 Lagrange 优化方法, 根据 Wolfe 的对偶理论可以把上述分类问题转化为它的对偶问题, 最大化泛函:

$$Q(\alpha) = \sum_{j=1}^n \alpha_j - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \quad (3.6)$$

$$\text{s. t. } \sum_{i=1}^n y_i \alpha_i = 0, \quad (3.7)$$

$$0 \leq \alpha_i, i = 1, \dots, n \quad (3.8)$$

$\alpha_i$  为第  $i$  个样本相对应的 Lagrange 乘子, 这是一个有不等式约束的二次规划问题, 因此存在唯一解。解上述问题后得到的最优分类函数是

$$f(x) = \text{sgn}\{(w \cdot x) + b\} = \text{sgn}\left\{\sum_{i=1}^n \alpha_i y_i (x_i \cdot x) + b\right\} \quad (3.9)$$

其中,  $\alpha$  是 (3.6), (3.7), (3.8) 的解,  $b$  是分类阈值, 可以用任一个支持向量满足。

值得注意的是, 对于  $N$  维空间中的线性函数, 其 VC 维为  $N+1$ , 当维数较高时 VC 维相应较大, 由 (3.5) 可知, 由于受到  $\|w\| \leq A$  的约束, 其 VC 维有可能大大减小, 也就是说在十分高维的空间中也可以得到较小 VC 维的函数集, 使有效地对付维数灾难成为可能。

### 3.4.2 最优广义分类面

最优分类超平面是在线性可分的前提下讨论的, 在线性不可分的情况下, 就是某些训练样本不能满足式 (3.4) 的条件, 因此可以在条件中增加一个非负的松弛变量  $\xi$ ,  $i = 1, \dots, l$ , 放宽条件 (3.4), 从而变为:

$$y_i(w \cdot x_i + b) \geq 1 - \xi, \quad i = 1, \dots, l \quad (3.10)$$

$$\xi \geq 0, \quad i = 1, \dots, l \quad (3.11)$$

显然, 当分割出现错误时,  $\xi_i$  大于零, 于是  $\sum_i \xi_i$  是训练错误数量的一个上界,

在约束条件 (3.10) 下对  $\sum_i \xi_i$  求极小, 就得到线形不可分情况下的最优分类面, 成为广义最优分类面。引入错误惩罚分量, 目标函数变为:

$$\Phi(w, \xi) = \frac{1}{2}(w, w) + C\left(\sum_{i=1}^l \xi_i\right) \quad (3.12)$$

其中  $C$  是一个可调参数, 它实际上起控制对错分样本惩罚的程度作用, 现在在错分样本与算法复杂度之间折衷,  $C$  越大, 对错误的惩罚越重。类似线性可分时的技巧处理, 可知为了找到广义最优分类面的系数:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (3.13)$$

必须解决一个二次规划问题:

在约束条件

$$0 \leq \alpha_i \leq C, i=1, \dots, l \quad (3.14)$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (3.15)$$

下, 求

$$W(\alpha) = \sum_{j=1}^n \alpha_j - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \quad (3.16)$$

的最大值。

我们需要初始 Lagrange 函数的 Kohn-Tucker 条件, 下面先给出初始的 Lagrange 函数:

$$L(w, b, a) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^l \alpha_i \{y_i(w \cdot x + b) - 1 + \xi_i\} - \sum_{i=1}^l \mu_i \xi_i \quad (3.17)$$

由 Kohn-Tucker 定理可知, 最优解满足

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \quad (3.18)$$

$$\alpha_i (y_i (w \cdot x) + b - 1 + \xi_i) = 0, \forall i \quad (3.19)$$

$$\alpha_i, \mu_i, \xi_i \geq 0, \forall i \quad (3.20)$$

$$\mu_i \xi_i = 0, \forall i \quad (3.21)$$

由 (3.18), (3.19) 可知, 如果  $\alpha_i < C$ , 则  $\xi_i$  为零。于是只选择  $i$  满足  $0 < \alpha_i < C$ , 即可通过 (3.19) 求出  $b$ 。式 (3.18) 到 (3.21) 又称为 KKT 条件, 由于 KKT 条件是最优解应满足的充要条件, 所以目前提出的一些算法几乎都是以是否违反 KKT 条件作为迭代策略的准则。

### 3.4.3 支持向量机

前面讨论的最优和广义线性分类函数, 其最终分类判别函数 (3.9) 中只包含待分类样本与训练样本中的支持向量的内积运算 ( $x \cdot x$ ), 同样, 它的求解过程式 (3.6) (3.7) 中也只涉及训练样本之间的内积运算 ( $x \cdot x$ ), 可见, 要解决一个特征空间中的最优分类问题, 我们只需要知道这个空间中的内积运算即可。

按照广义线性判别函数的思路, 要解决非线性问题, 可以设法将它通过非线性



性变换转化为另一个空间中的线性问题,在这个变换空间求最优或广义最优分类面。考虑到最优分类面算法的性质,在这个变换空间中,我们只需要进行内积运算,甚至没有必要知道采用的非线性变换的形式,而只需要它的内积运算就可以了。只要变换空间中的内积可以用原空间中的变量直接计算得到,则即使变换空间的维数增加很多,在其中求解最优分类面的问题并没有增加多少计算复杂度。实际上,我们只要定义变换后的内积运算,而不必真的进行这种变换。统计学习理论指出,根据 Hilbert-Schmidt 原理,只要一种运算满足 Mercer 条件,它就可以作为这里的内积使用。

定理 3.1 (Mercer 条件)<sup>[46]</sup>对于任意的对称函数  $K(x, x')$ , 它是某个特征空间中的内积运算的充分必要条件是, 对于任意的  $\varphi(x) \neq 0$  且  $\int \varphi^2(x) dx < \infty$  有

$$\iint K(x, x') \varphi(x) \varphi(x') dx dx' > 0 \quad (3.22)$$

这一条件不难满足。

只要一种核函数  $K(x_i, x_j)$  满足 Mercer 条件, 它就对应某一特征空间中的内积, 即在最优超平面中采用适当的核函数  $K(x_i, x_j)$  就可以实现某一非线性变换后的线性分类, 而计算复杂程度却不会增加。此时, 目标函数 (3.4.13) 变成

$$\begin{aligned} \max W(\alpha) &= \sum_{j=1}^n \alpha_j - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\ \text{s.t. } &\sum_{i=1}^l \alpha_i y_i = 0 \\ &0 \leq \alpha_i \leq C, i = 1, \dots, l \\ &w = \sum_{i=1}^l \alpha_i y_i \Phi(x_i) \end{aligned} \quad (3.23)$$

相应的判别函数 (3.9) 变为:

$$y = \text{sgn} \left[ \sum_{x_j \in SV} \alpha_j y_j K(x, x_j) + b \right] \quad (3.24)$$

推导过程与前面完全相同, 只是把那里的  $x, x_i$  分别换成  $\Phi(x), \Phi(x_i)$ 。

不难看出为了求分离曲面, 不必知道  $\Phi(x)$  的确切表达式, 只要知道如何由输入  $x, x_i$  计算内积  $(\Phi(x) \square \Phi(x_i))$  就够了, 即

$$(\Phi(x) \square \Phi(x_i)) = K(x, x_i) \quad (3.25)$$

$K(x, x_i)$  称为核函数。

这个事实对构造支持向量机有重要意义, 当特征向量的维数巨大时(实际情况常常如此有时甚至是无穷维), 若直接计算内积其复杂程度是可想而知的。因此行不通。上述事实指出: 高维特征空间中内积运算, 可转化为低维输入空间(相

对而言维数要低得多)上一个简单的函数计算。

定义3.1 称判别函数

$$y = \text{sgn} \left[ \sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \right] \quad (3.26)$$

为支持向量机(Support Vector Machines), 简称为: SVM.

概括的讲, SVM就是首先通过用内积空间定义的非线性变换把输入空间变换到一个高维特征空间, 在这个空间中求(广义)最优超平面。支持向量机分类函数形式上类似于一个神经网络, 其输出是若干中间层节点的线性组合, 而每一个中间层节点对应于输入样本与一个支持向量的内积。(如图3.2所示)

在支持向量机中, 函数  $K$  称为内积的卷积核函数, 它可以看作在样本之间定义的一种距离。由于最终的判别函数中实际只包含与支持向量的内积求和, 内积在这里最为描述相似度地工具。因此识别时的计算复杂度取决于支持向量的个数, 而不是特征空间的维数。

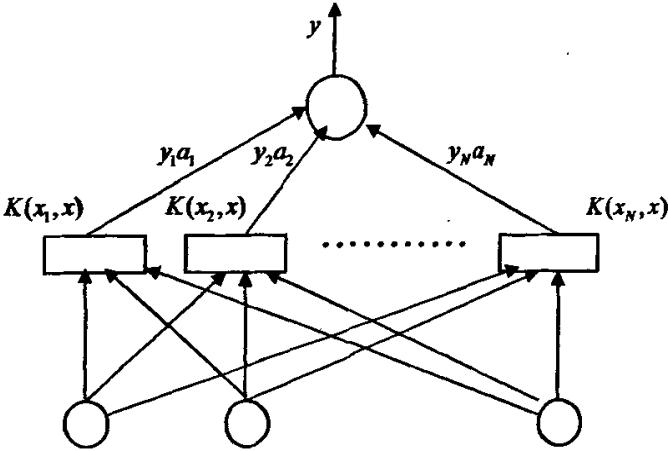


图 3.2 SVM 示意图

Figure 3.2 SVM Structure

### 3.4.4 核函数研究及改进

选择满足Mercer条件的不同内积核函数,, 我们可以构造实现输入空间中不同类型的非线性决策面的学习机。所以如何构造、选择核函数是一个重要的问题。目前研究最多的核函数主要有三类核函数<sup>[67]</sup>:

(1) 多项式形式核函数:

$$K(x, y) = ((x, y) + 1)^d \quad (3.27)$$

此时得到的支持向量机就是一个 $d$ 阶多项式分类器。特别当 $d=1$ 时，此时就是线性核函数，也成为线性支持向量机。

(2) 径向基核函数：

$$K(x,y)=\exp(-\frac{\|x-y\|^2}{2\sigma}) \quad (3.28)$$

得到的支持向量机是一种径向基函数分类器。它与传统的神经网络种的RBF网络的基本区别是：这里的每一个基函数的中心都对应于一个支持向量，它们以及输出权值都是由算法自动确定的。

(3) Sigmoid 核函数：

$$K(x,y)=\tanh(v(x,y)+c) \quad (3.29)$$

其中 $v$ 和 $c$ 是常数；则支持向量机实现的就是一个两层的感知器神经网络，不过这里网络的权值、隐层结点数目都是由算法自动确定的。

图3.3、图3.4分别是两种核函数的示意图，图中 $x$ 坐标、 $y$ 坐标代表核函数 $K(x,y)$ 中的 $x$ 、 $y$ 值， $z$ 坐标代表核函数的值。可以很直观的看出，随着 $x$ 、 $y$ 值越来越相似， $K(x,y)$ 的值就会逐渐增大；当 $x$ 、 $y$ 相似度最高，即 $x=y$ 时， $K(x,y)$ 达

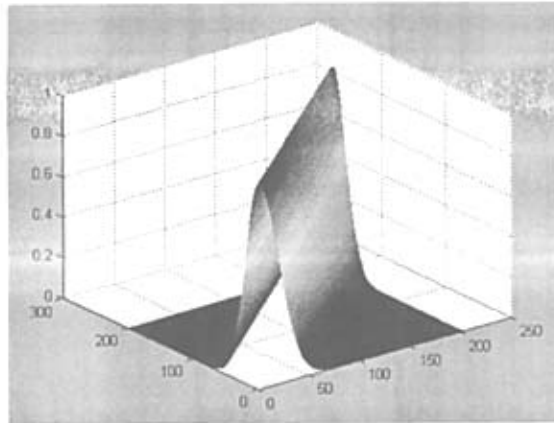


图 3.3 RBF 核函数示意图

Figure 3.3 RBF Kernel Function

到最大值。这就是通过内积来体现相似度的基本原理。

但不同的核函数所体现的相似度也有很大的差异，如图3.4，图形在中部有一个很明显的凹陷，我们可以认为线性核函数，即(3.4.23)中 $d=1$ 的情况，更加关注取边缘值的 $x$ 、 $y$ 的相似度。

惠康华，李春利提出 $W_2'$ 再生核函数<sup>[71]</sup>：

$$K(x,y)=\frac{\cosh(x+y-a-b)+\cosh(x-y-b+a)}{2\sinh(b-a)} \tag{3.30}$$

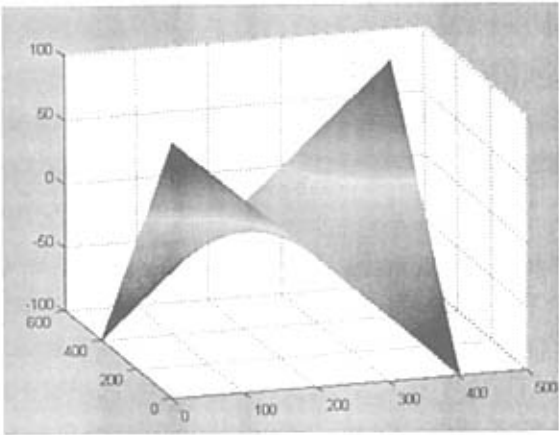


图 3.4 线形核函数简单示意图

Figure 3.4 Linear Kernel Function

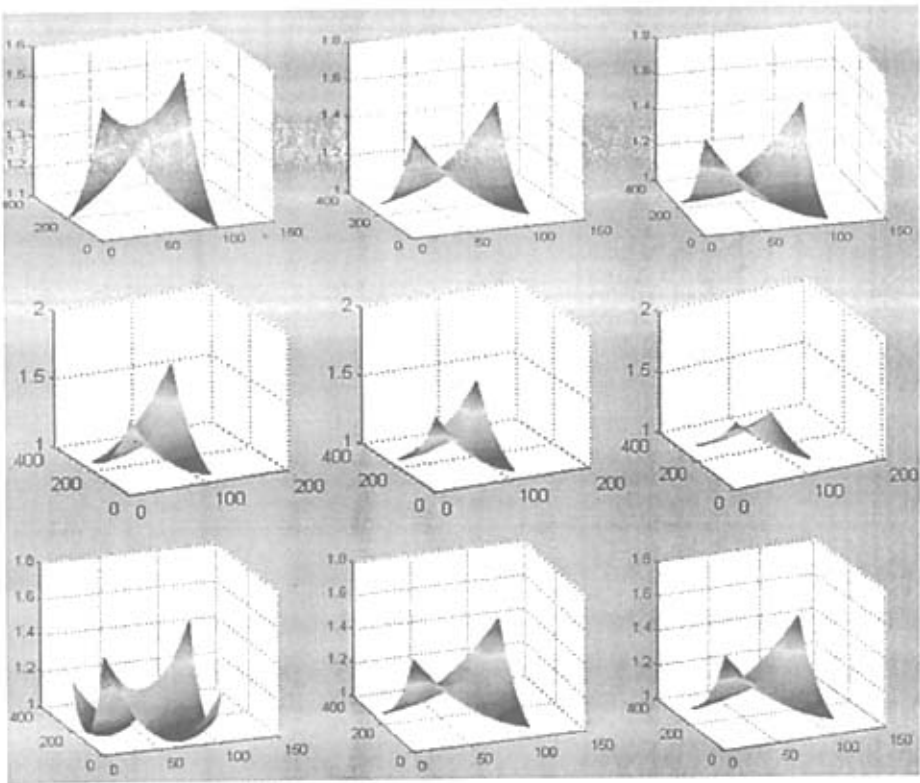


图 3.5 参数选择对核函数的影响示意图

Figure 3.5 Character's Impact on Kernel Function

其中 $a=0$ ,  $b=1$ 鉴于 $a$ ,  $b$ 确定后该核函数分母为常数, 本文进行对其改进为:

$$K(x,y)=\lambda \cosh\left(\frac{|x+y|}{w1}-a-b\right)+(1-\lambda)\cosh\left(\frac{|x-y|}{w2}-b+a\right) \quad (3.31)$$

图3.5展示了参数变化对本文中改进的cosh核函数的影响, 第一行中三幅核函数图像从左到右分别是 $\lambda=0.2, 0.5, 0.7$ ; 第二行中三幅核函数图像从左到右分别是 $w1=0.5, 1, 1.5$ ; 第二行中三幅核函数图像从左到右分别是 $w2=0.5, 2, 2.5$ 。

可见, 通过改变参数, 可对核函数进行细致的微调, 以适用于不同特性的样本提高分类正确率。

仍以ORL人脸库为例, 共40个人, 每人10幅图像, 前5幅用于训练, 后5幅用作检测。不同的特征提取方法与不同的分类器结合, 识别结果见表4.1。其中每一行是一种特征提取方法, 每一列一种分类器。实验证明, 本文中改进的cosh核函数的识别率高于常用的三种核函数, 见表3.1。

表 3.1 改进的 cosh 核函数的识别率  
Table 3.1 Recognition Ratio With the improved cosh Kernel Function

	SVM(RBF)	SVM(poly)	SVM(liner)	SVM(cosh)
DCT+DCT	93.00%	88.50%	92.50%	93.00%
Fisher+DCT	96.00%	93.00%	96.50%	96.50%
Fisher+DCT+ICA	97.50%	96.50%	97.00%	98.00%

## 3.5 支持向量机参数选择

### 3.5.1 参数选择方法

我们称学习机对未知数据进行分类的性能为推广能力, 推广能力的优劣直接影响着学习机的性能, 遗憾的是, 对学习机的推广能力的估计还没有形成简单、准确的方法。

通过对支持向量机的大量实验研究表明, 参数选择(主要是核函数的参数选择)在支持向量机中是至关重要的, 由于参数变换导致的推广能力的波动甚至大于核函数类型变换对推广能力的影响, 不恰当的参数可能导致推广能力下降为零(即识别错误率为100%)。核函数参数对支持向量机的推广能力的影响如图3.7所示。

然而, 支持向量机核函数参数的选择至今没有一个完善而具体的理论指导,

已证明的留一法对分类性能的估计是无偏的,但其计算量比较大,因此提出以下比较公认的简化计算量的定理,实现对支持向量机推广能力的衡量:

定理3.1<sup>[70]</sup>: 如果一组训练样本能够被一个最优分类面和支持向量机分开,则对于测试样本分类错误率的期望的上界是训练样本中平均的支持向量占总训练样本数的比例,即:

$$E(P(error)) \leq \frac{E[\text{支持向量个数}]}{\text{训练样本数} - 1} \quad (3.32)$$

因此只要能够选择适当的参数,构造一个支持向量数相对较少的最优分类面,便可以得到较好的分类。

优化过程就是要找到使  $E(P(error))$  最小的  $C$  和  $\lambda$ ,  $w_1$ ,  $w_2$ , 通过对径向基核函数SVM的参数  $C$  研究可知,  $C$  是在结构风险和样本误差之间做出折衷,取值与可容忍的误差相关,较大的  $C$  值允许较小的误差,较小的  $C$  值则允许较大的误差。

由图3.6可知,  $C$  取值足够大,即可满足较好的分类能力,因此,参数选取重点关注支持向量机核函数的参数选取。传统的选取方法是Grid方法<sup>[69]</sup>,其实质是一种遍历的方法,各个参数在其范围内遍历所有可选值,选取最优解。该方法随着参数和格点数的增加,计算时间大大增加。因此本文提出将离散微粒群算法用于支持向量机的参数选择。

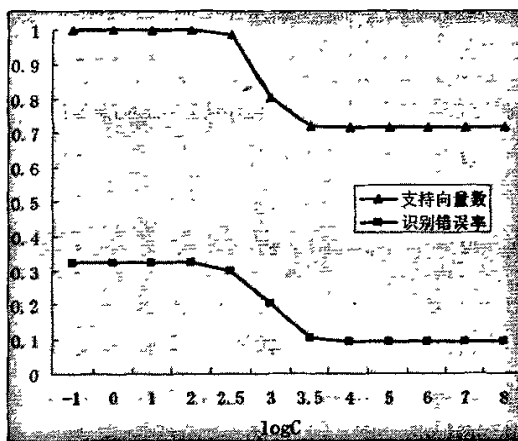


图 3.6 参数  $C$  与支持向量数与错误识别率关系

Figure 3.6 Character  $C$  and Recognition Error Based on SVM

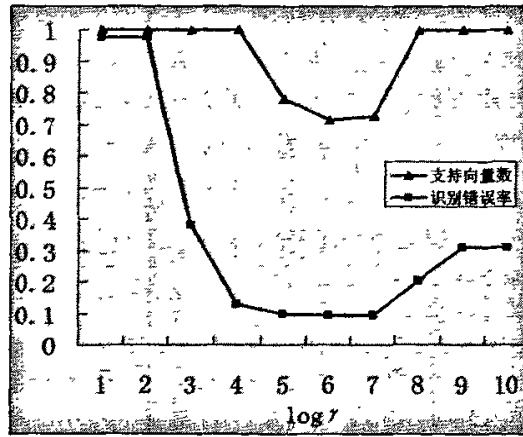


图 3.7 参数  $\gamma$  与支持向量数与错误识别率关系

Figure 3.7 Character  $\gamma$  and Recognition Error Based on SVM

### 3.5.2 微粒群算法原理及算法流程

受到鸟群觅食的启发, J.Kennedy和R.C. Eberhart于1995年提出了微粒群算法(PSO)<sup>[72]</sup>, 该算法是一种进化计算技术, 经过众多学者十余年的研究, 它已被广泛应用于解决多个领域的优化问题, 如神经网络训练、电磁学等等<sup>[73]</sup>。与传统的优化算法相比, 微粒群算法能搜索非线性多峰的复杂空间, 同时能处理参数空间变化的优化问题。与同是进化算法的遗传算法相比较, 微粒群算法能更快达到全局最优, 且基本不受问题峰数和维数的限制。

微粒群算法将群体内的每个个体看作没有体积和重量的微粒, 微粒在 $n$ 维搜索空间内以一定的速度飞行搜索。飞行速率和方向根据微粒自身的飞行经验和整个微粒群的飞行经验动态调整。微粒 $i$ 的当前位置为 $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , 当前飞行速度为 $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$ , 所经历的最好位置即 $i$ 所经历过的具有最好适应值的位置为 $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$ 。群体中所有微粒所经历的最好位置为 $P_{best}$ 。带惯性权重的微粒群算法的飞行方程为<sup>[74]</sup>:

$$v_j(t+1) = w(t)v_j(t) + c_1(t)r_{1j}(t)(p_{1j}(t) - x_j(t)) + c_2(t)r_{2j}(t)(p_{bestj}(t) - x_j(t)) \quad (3.33)$$

$$x_j(t+1) = x_j(t) + v_j(t+1) \quad (3.34)$$

其中 $j$ 表示微粒的第 $j$ 维,  $t$ 表示第 $t$ 代,  $w$ 为惯性权重, 通常取 0.8 到 1.2 之间,  $c_1, c_2$ 为通常在 0 到 2 间取值,  $w, c_1, c_2$ 都是随时间随机变化的数,  $r_1, r_2$ 为两

个相互独立的随机函数,服从 $U(0,1)$ 分布。为了减少飞行过程中微粒离开搜索空间的可能性通常设定 $v_y \in [-v_{\max}, v_{\max}]$ 。

基本 PSO 的参数主要有惯性权值和加速系数。

### (1) 惯性权值 $w$

$w$  对 PSO 能否收敛起重要作用,它使微粒保持运动惯性,使其有扩展搜索空间的趋势,有能力探索新的区域。 $w$  值大些有利于全局搜索,收敛速度快,但不易得到精确解; $w$  值小些有利于局部搜索和得到更为精确的解,但收敛速度慢且有时会陷入局部极值。合适的  $w$  值在搜索精度和搜索速度方面起协调作用。

### (2) 加速系数 $c_1$ 、 $c_2$

$c_1$  和  $c_2$  对 PSO 的收敛速度影响颇大,合适的加速系数有利于算法较快收敛和脱离局部极值。它们代表将每个微粒推向其个体历史最好位置和全局历史最好位置的统计加速项的权值。低的加速系数值允许微粒在被拉回之前可以在目标区域外徘徊,而高的加速系数值则可能导致微粒突然地冲向或越过目标区域。

在公式 (3.33) 中,若  $c_1 = c_2 = 0$ ,微粒将一直以当前的速度惯性飞行,直到到达边界;由于它只能搜索有限的区域,所以很难找到最好解。若  $c_1 = 0$ ,则微粒没有认知能力,只有社会部分,所以  $c_1$  又称为社会参数;此时收敛速度比基本 PSO 快,但对复杂问题,则比基本 PSO 容易陷入局部极值。若  $c_2 = 0$ ,则微粒之间没有社会信息共享,只有认知部分,所以  $c_2$  又称为认知参数;此时个体间没有交互,一个规模为  $m$  的群体等价于  $m$  个单个微粒的运行,得到最好解的概率非常小。

通常,  $c_1 = c_2 = 2$ 。也有实验结果显示,当  $c_1 = c_2 = 0.2$  时能取得更好的效果。最近一些研究还表明认知参数  $c_2$  选择的大些而社会参数  $c_1$  选择的小些,但  $c_1 + c_2 < 4$  时能得到更好的结果。

此外,随机数  $r_1$ 、 $r_2$  可以保证微粒群体的多样性和搜索的随机性。最大最小速度可以决定当前位置与最好位置之间区域的分辨率(或精度)。如果最大速度太高或最小速度太低,微粒可能会飞过好解;如果最大速度太小或最小速度太大,则微粒不能在局部最好区间之外进行足够的探索,导致陷入局部极值。该限制的目主要是防止计算溢出,改善搜索效率和提高搜索精度

基本 PSO 算法流程如下:

① 初始化设置微粒群的规模、惯性权值、加速系数、最大允许迭代次数或适应值误差限,各微粒的初始位置和初始速度等。

② 按目标函数评价各微粒的初始适应值。

③ 根据公式 (3.33) 计算各微粒新的速度,并对各微粒新的速度进行限幅处理。



④ 根据公式 (3.34) 计算各微粒新的位置, 并对各微粒新的位置进行限幅处理。

⑤ 按目标函数重新评价各微粒的适应值。

⑥ 对每个微粒, 比较其当前适应值和其个体历史最好适应值, 若当前适应值更优, 则令当前适应值为其个体历史最好适应值, 并保存当前位置为其个体历史最好位置。

⑦ 比较群体所有微粒的当前适应值和全局历史最好适应值, 若某微粒的当前适应值更优, 则令该微粒的当前适应值为全局历史最好适应值, 并保存该微粒的当前位置为全局历史最好位置。

⑧ 若满足停止条件(适应值误差达到设定的适应值误差限或迭代次数超过最大允许迭代欢数), 搜索停止, 输出搜索结果。否则, 返回③继续搜索。

### 3.5.3 基于离散微粒群算法的 SVM 参数选择

根据本文中参数选择的特点, 为保证离散微粒群算法在整数空间内进行搜索, 我们对算法进行如下修改: 令

$$\phi_1 = c_1(t)r_{1j}(t)(p_{ij}(t) - x_{ij}(t)) \quad (3.35)$$

$$\phi_2 = c_2(t)r_{2j}(t)(p_{bestj}(t) - x_{ij}(t)) \quad (3.36)$$

其中  $\phi_1 \in [a_1, b_1]$ ,  $\phi_2 \in [a_2, b_2]$ ,  $\phi_1$  和  $\phi_2$  为在各自区间内等概率分布的整数。区间范围确定如下:

$$a_1 = \begin{cases} 0, p_{ij}(t) > x_{ij}(t) \\ c_1(p_{ij}(t) - x_{ij}(t)), p_{ij}(t) \leq x_{ij}(t) \end{cases} \quad (3.37)$$

$$b_1 = \begin{cases} c_1(p_{ij}(t) - x_{ij}(t)), p_{ij}(t) > x_{ij}(t) \\ 0, p_{ij}(t) \leq x_{ij}(t) \end{cases} \quad (3.38)$$

$$a_2 = \begin{cases} 0, p_{bestj}(t) > x_{ij}(t) \\ c_2(p_{bestj}(t) - x_{ij}(t)), p_{bestj}(t) \leq x_{ij}(t) \end{cases} \quad (3.39)$$

$$b_2 = \begin{cases} c_2(p_{bestj}(t) - x_{ij}(t)), p_{bestj}(t) > x_{ij}(t) \\ 0, p_{bestj}(t) \leq x_{ij}(t) \end{cases} \quad (3.40)$$

则式 (3.33) 修改成:

$$v_{ij}(t+1) = \text{int}(w(t)v_{ij}(t)) + \phi_1 + \phi_2 \quad (3.41)$$

其中  $\text{int}(w(t)v_{ij}(t))$  是对  $w(t)v_{ij}(t)$  取整。将  $X_i$  初始化为整数, 式(3.3)保证了算

法在整数空间内搜索。

以本文提出的基于 cosh 的核函数为例，优化过程就是要找到使式 (3.32) 中  $E(P(error))$  最小的  $C$  和  $\lambda$ ， $w1$ ， $w2$ ，由上文可知只要  $C$  取值足够大，既可得到较好的分类能力。于是本文将利用微粒群算法选取最优参数  $\lambda$ ， $w1$  和  $w2$ 。

选取适应函数

$$d_0 = \frac{\text{支持向量个数}}{\text{训练样本数}-1} \tag{3.42}$$

需要搜索的参数  $\lambda$ ， $w1$  和  $w2$ ，从而确定微粒的搜索空间为三维  $X_i = (x_{i1}, x_{i2}, x_{i3})$ ，搜索优化过程如下：

Step1 初始化设置微粒群的规模、惯性权值、加速系数、最大允许迭代次数或适应值误差限，各微粒的初始位置和初始速度等。

Step2 由式 (3.42) 评价各微粒的初始适应值。

Step3 对于每个微粒，将其适应值与所经历过的最好位置  $P_i$  的适应值进行比较，若较小，则将其作为当前的最好位置；

Step4 对于每个微粒，将其  $P_i$  与全局所经历的最好位置  $P_{best}$  的适应值进行比较，若较小，则将其作为当前全局的最好位置；

Step5 确定惯性权重  $w$ ；

Step6 根据式 (3.33)，(3.34) 对微粒的速度和位置进行进化；

Step7 如果  $d_0$  没有达到迭代终止条件，则返回 Step2；否则结束，此时最好位置。

数据库仍为为 orl 人脸库，共 40 类，每类 10 个样本。每一类的前五个样本作为训练样本，后五个样本作为检测样本。通过 LDA，DCT 和 ICA 提取人脸图像的特征，组成人脸特征的组合特征，用于 SVM 的训练与识别。

搜索范围  $0.1 \leq \lambda \leq 0.9, 0.5 \leq w1 \leq 3, 0.5 \leq w2 \leq 3$ 。当格点数逐渐增加，即搜索精度逐渐提高，由 (3.42) 式得到的错误率上界见表 3.2。其中  $E = \text{支持向量数}/(\text{训练样本数}-1)$ 。

表 3.2 搜索精度与错误上界

Table 3.2 Search Precision and Error Upper Bound

格点数	64	125	196	343	5625
$\lambda$	0.63	0.7	0.74	0.77	0.8
$w1$	0.5	0.5	0.5	0.5	0.5
$w2$	2.16	2.38	2	2.17	2.2
E	0.7977	0.7914	0.7881	0.7863	0.7838

将获得的最优参数用于人脸识别，两种参数搜索方法比较见表 3.3。说明本方法搜索的最优参数与网格搜索法完全一致，但搜索速度大大提高，耗时比较见图 3.8。

图 3.8 中横坐标为网格点数，纵坐标为搜索时间随着网格点数的增加，网格法的耗时大幅度增加，而 DPSO 的耗时则增加很慢，说明在同等搜索精度下，DPSO 方法在搜索速度上有明显的优势。

表 3.3 格点法与 DPSO 比较  
Table 3.3 Comparison between Grid and DPSO

方法	$\lambda$	$w1$	$w2$	E	识别正确率	搜索时间(秒)
Grid	0.8	0.5	2	78.38%	98%	101250
DPSO	0.8	0.5	2	78.38%	98%	3110

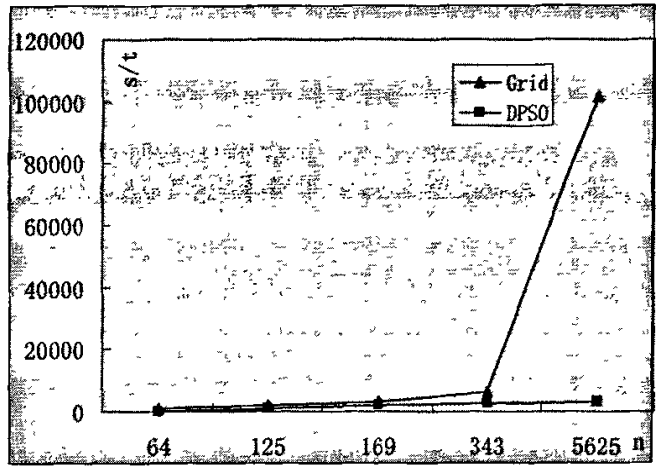


图 3.8 Grid 与 DPSO 耗时比较

Figure 3.8 Time-Consuming Comparison between Grid and DPSO

3.6 本章小结

本章从理论推倒和实验分析的角度详细介绍了支持向量机用于人脸识别技术。本章包括了本文的两个创新点：改进的基于 cosh 的核函数；基于离散微粒群算法的支持向量机参数选择方法。

经实验验证，本文提出的改进 cosh 核函数具有良好的分类能力，通过微小

调整核函数的参数，能够得到比较流行的 RBF 核函数有更好的识别率。鉴于支持向量机分类器的正确率严重依赖核函数参数的选择，而核函数的参数选择非常耗时，所以本文提出了基于离散微粒群算法的支持向量机参数选择方法，大大提高了支持向量机参数的选择速度。

## 4 人脸识别系统

构建了一个实时的人脸识别系统,将本文所研究的算法应用其中,实现实时的人脸检测、特征提取以及识别。也为进一步研究人脸识别算法提供一个平台。希望能随着系统的不断完善,对人脸识别算法的研究有越来越大的帮助,同时希望表人脸别的研究也越来越成熟。在本章中,就目前所搭建的系统的软硬件,算法流程、用法给出介绍和描述。

### 4.1 实验环境介绍

人脸图像采集: logitech USB 摄像头

计算机: Intel 奔腾 IV2.4G, 内存 256M。

开发工具: Visual Studio 2005 集成开发环境。

操作系统: Windows XP。

### 4.2 系统整体构成及功能

本系统算法实现流程图见图 4.1, 具体流程如下:

第一, 训练阶段。

1, 载入人脸图像训练集, 将训练集的保存为三位数组形式(人脸图像大小为  $60 \times 60$ , 已经过直方图均衡化, ) ;

2, 选择的特征提取方法(如 ICA 算法), 根据所选参数(如特征维数)对训练集进行处理, 得到投影矩阵;

3, 根据上一步得到的投影矩阵, 将向量化的人脸图像投影到相应的特征子空间(如 ICA 子空间), 即得到人脸图像训练集的特征矩阵;

4, 选择支持向量机作为分类器, 选定核函数(如本文改进的基于 cosh 的核函数)并设定支持向量机参数(如  $C$ ,  $w_1$ ,  $w_2$ ), 对上一步得到的特征矩阵进行归一化处理后, 输入到支持向量机进行训练, 得到支持向量机参数  $\alpha$  矩阵。至此训练完成。

第二, 识别阶段。

1, 利用连接在 PC 机上的 logitech 摄像头采集人脸图像(包括人脸检测, 双眼定标, 人脸图像分割), 并对待识人脸图像进行直方图均衡化处理, 并向量化。

2, 通过训练阶段第 2 步骤得到的投影矩阵, 将待识人脸向量投影到相应的特征子空间, 得到待识人脸特征向量。

- 3, 将待识人脸特征向量输入已训练好的支持向量机, 进行分类。
- 4, 输出识别结果。

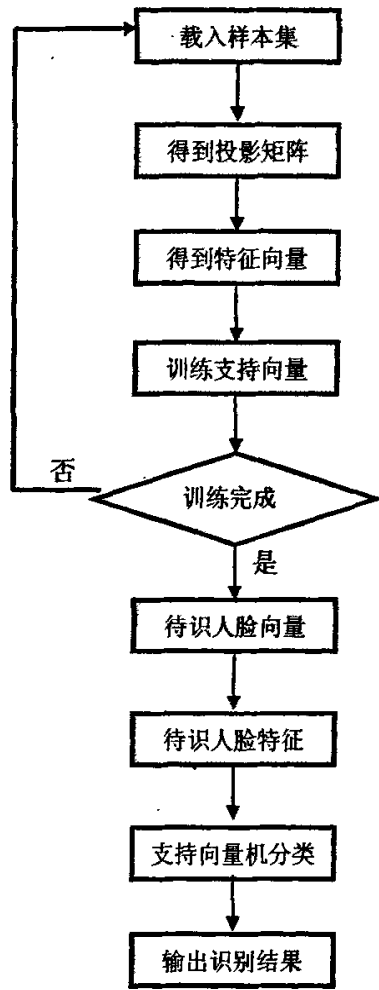


图 4.1 人脸识别系统流程图

Figure 4.1 Face Recognition Flowchart

4.3 系统介绍及演示

本系统结合了 FastICA 算法和支持向量机理论, 利用 Visual Studio 2005 集成开发环境, 实现了人脸实时识别功能。

运行程序，打开程序主界面，如图 4.2，界面下部是对相关参数的设定，右侧显示实时识别结果。

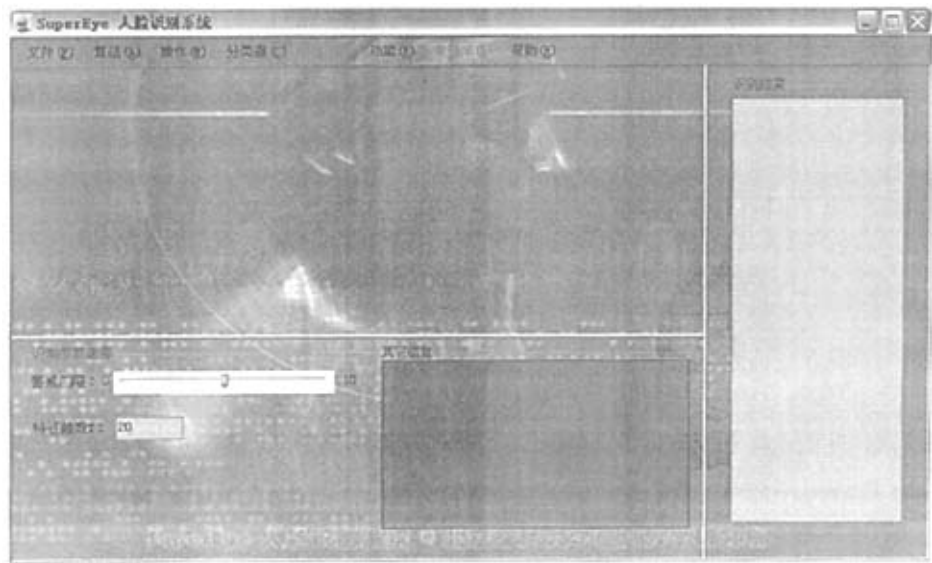


图 4.2 人脸识别系统主界面  
Figure 4.2 Face Recognition System Interface

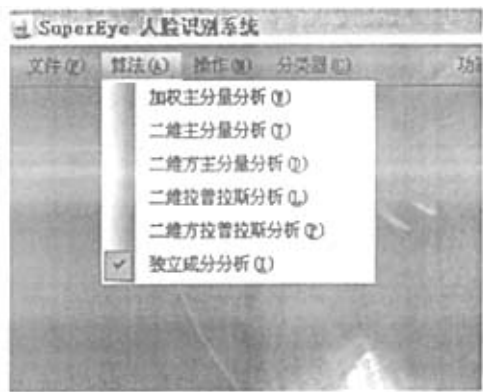


图 4.3 选择特征提取方法  
Figure 4.3 Feature Extraction Selections

通过算法菜单栏，选择特征提取算法独立分量分析（ICA），如图 4.3。  
通过操作菜单栏载入训练样本集，得到投影矩阵，得到特征矩阵，如图 4.4。  
通过分类器菜单栏选择支持向量机作为分类器。通过训练菜单栏训练支持向量机，通过界面下部的参数选择区域选择支持向量机的相关参数，如图 4.5。  
通过功能菜单栏开始实时人脸识别。界面右侧显示实时识别结果，包括训练

集中的样本图片及待识者的姓名信息，如图 4.6。



图 4.4 得到训练集特征矩阵  
Figure 4.4 Obtain Feature Matrix



图 4.5 支持向量机参数  
Figure 4.5 Characters of SVM

## 4.4 本章小结

本章详细地介绍了本论文实验的实验环境，实验方案设计，并演示了实时人脸识别系统的工作过程，通过实验证明了本文中所提到的算法的可行性，实验证明本算法的运算速度较快，可以满足实时的要求。但是算法的鲁棒性仍存在不足，识别率随着周围环境的变化（主要是光照影响）。此后的工作主要有两方面：验证本系统在大训练集的情况下的能否达到实时的要求；改善本系统对光照变化的鲁棒性。





图 4.6 实时人脸识别结果输出  
Figure 4.6 Real Time Face Recognition Output

## 5 总结与展望

### 5.1 论文总结

人脸识别是人机交互与信息处理领域中的一个重要课题,具有很高的学术价值和应用潜力。本文就人脸识别算法的研究及人脸识别系统的建立已经在前面的章节给出了较为详尽的论述,下面将主要工作成果总结如下:

第一,在研究并实现已有的经典统计特征提取算法的基础之上,将 ICA、LDA、DCT 算法进行融合,形成更为有效的人脸特征用于人脸识别。通过对 orl 人脸库的人脸识别实验证明该特征的有效性。

第二,研究了支持向量机用于人脸识别的技术,尝试多种核函数用于人脸识别。改进了  $W_2^1$  再生核,提出了基于 cosh 函数的核函数,同常用的 RBF、多项式、线形核函数相比,本文提出的 cosh 核函数可以通过调节核函数的参数来适应不同类型的特征向量。通过对 orl 人脸库的人脸识别实验证明该核函数具有更好的识别效果。

第三,研究参数选择对支持向量机分类效果的影响,将微粒群算法运用于核函数的选择,提出基于离散微粒群算法的支持向量机参数选择技术,大大提高了支持向量机参数选择的速度。

第四,参与北京交通大学信息所人脸识别系统的设计,将 ICA 特征提取算法和支持向量机运用于实时人脸技术,证明了本文提出的 cosh 核函数的有效性和可行性,为日后进一步研究奠定了基础。

### 5.2 工作展望

尽管本文的研究取得了比较理想的结果,但是由于课题时间有限,仍存在一些值得进一步的研究解决:

第一,特征融合技术虽然能提高识别率,但是会大幅度的增加特征向量的维数,相应的计算量增加导致人脸识别系统的实时性降低。另外特征向量的信息量过多会引入无用的冗余信息,反而会影响识别能力。因此今后需要在特征融合方面寻求理论论证,从而解决特征向量的复杂度和识别能力之间的问题。

第二,支持向量机是目前流行的模式识别方法,有很好应用的前景,但仍需要进一步的研究提高其实用性,比如,当样本集庞大时,支持向量机的训练速度就会变得很慢,需要一种快速优化算法用于支持向量机的训练。

第三,本文中描述的实时人脸识别系统是对现有算法的简单实现,还不能满

足大样本集、复杂背景等环境下的实时人脸识别要求，具有一般人脸识别系统的通病，即对光照、复杂背景的鲁棒性不高，在这一问题上仍需进行大量的研究工作。

总之，人脸识别领域还有很大的空间需要辛勤的科研工作者去共同努力共同探索，相信终有一天我们能克服所有问题使人脸识别技术得以广泛的应用。

## 参考文献

- [1] 李武军, 王崇骏, 张伟, 陈世福, 人脸识别研究综述, 模式识别与人工智能, 2006年2月, 第19卷, 第1期
- [2] 张翠平, 苏光大, 人脸识别技术综述, 中国图像图形学报, 2000年11月
- [3] 李刚, 高政, 人脸识别理论研究进展, 计算机与现代化, 2003年第3期
- [4] W. ZHAO Face Recognition: A Literature Survey ACM Computing Surveys, Vol. 35, No. 4, December 2003, pp. 399-458.
- [5] 牛丽平, 付仲良, 魏文利, 人脸识别技术研究, 计算机工程与设计, 2006年11期
- [6] 李健, 廖秋筠, 人脸识别的方法研究, 微计算机信息(嵌入式与SOC), 2006年, 第22卷第4-2期
- [7] SIR FRANCIS GALTON, "Personal identification and description-1[J]", Nature, 1998(21):173-177
- [8] W.Bledsoe, "Man-machine facial recognition", Panoramic Research Inc,Palo Alto,CA,1966,Rep PRL22
- [9] KellyMD, "Visual Identification of People by Computer[J]", ph.d.thesis, 1970,Stanford U
- [10] Y.Kaya and K.Kobayashi, "A basic study on human face recognition[J]", Frontiers of Patern Recognition cs Watanabe, Ed, 2000:265-269
- [11] I Craw,H Ellis,J Lishman, "Automatic Extraction of Face Features[J]", Pattern Recognition Letters,1987,5:183-187
- [12] R Bmneli, T Poggio, "Face Recognition: Features vs Templates [J]", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(10):1042-1052
- [13] E Osuna,R Freund,F Girosi, "Training Support Vector Machines: An Applicationto Face Detection[C]", IEEE Conf.Computer Vision and Pattern Recognition, 1997.130-136
- [14] R Brunelli,T Poggio, "Face Recognition: Features Versus T-emplates[J]", IEEE Trans PAMI, 1993, 15(10):1042-1052
- [15] X.Jia, M.S.Nixon, "Analyzing front view face profile for face recognition via the Walsh transform", Pattern Recognition Letters, Vol.15, 1944. pp. 551-558
- [16] Xiaobao Li,Nicholas Roeder, "Face Reangnition Extraction From Front-View Images[J]", Pattern Recognition, 1995:28(8):1168-1179
- [17] Turk M, Pentland A, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 1991,3(1):71-86
- [18] Shackleton MA,Welsh W j, "Computer Vision & Pattarn Recognition", IEEE Conf.ON 1991:573-579
- [19] L.D.Harmon, M.K.Knan, Richard Lasch and P.F.Rarnig, Machine identification of human faces, Pattern Recognition, Vol.13, No.2, pp.97-110, 1981
- [20] 洪子泉, 杨静宇, 基于奇异值特征和统计模型的人像识别算法, 计算机研究与发展, 1994, 31(3):60-65
- [21] Hong Z, "Algebraic Feature Extraction of Image for Recognition [J]", Pattern Recognition, 1991, 24(3):211-219
- [22] Cheng Y. "Human Face Recognition Method Based on the Statistical Model of Small

Sample Size [A]”, SPIE Proceedings of Intelligent Robots and Computer Vision [C]. Boston, Massachusetts, USA, 1991, 85-95

[23] Brunelli R, Poggio T, “Face Recognition: Features Versus Templates [J]”, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1993, 15 (10):1042-1052

[24] Comon P. “Independent components of analysis—a new concept [J]”, Signal Processing, 1994, 36(3): 287-314

[25] Samaria F.S., “Face Recognition Using Hidden Markov Models [D]”, Univ. of Cambridge, 1994

[26] Rowley H A, Baluja S, Kanade T, “Neural network-Based Face Detection [J]”, IEEE Trans. PAMI, 1998, 20 (1): 23-38

[27] Chellappa R, Wilson CL, Sirohey S, “Human and machine recognition of faces: a survey”, Proceedings of the. IEEE, 1995, 83(5): 705-711

[28] M Bichsel. “Human Face Recognition and the Face Image Set's Topology [J]”, CVGIP:Image Understanding, 1994, 59(2):254-261

[29] 彭辉, 张长水, 荣钢等, 基于 KL 变换的人脸识别方法, 清华大学学报, 1997 年 3 月, 37(3):67-70

[30] 郭跃飞, 姜志华, 杨静宇等, 一种新的代数特征抽取方法及人脸识别, 南京理工大学学报, 1997, 21 (5): 387-390

[31] 李丽, 易从琴, 孟传良, 利用主元分析识别人脸, 贵州工业大学学报, 2003 年, 8 月, 第 32 卷第 4 期

[32] Turk M, Pentland, A. Eigenface for Recognition [J]. J Cog Neurosci, 1991, 23(3): 71—86

[33] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans on Pattern Analysis and Machine Intelligence. 1997. 19(7): 711-720

[34] 韩争胜, 李映 张艳宁, 基于 LDA 算法的人脸识别方法的比较研究, 微电子学与计算机, 2005 年第 22 卷第 7 期

[35] Hyvariene A. Fast and robust fixed- point algorithm for independent component analysis [J]. IEEE Trans. On Neural Networks, 1999, 10(3): 626- 634.

[36] 孟继成, 杨万麟独立分量分析在模式识别中的应用, 计算机应用, 第 24 卷第 8 期 2004 年 8 月

[37] Hvarinen A. Independent component analysis: A tutorial [EB/OL]. <http://www.cis.hut.fi/projects/ica/>.

[38] Aapo Hyvärinen, Ekki Oja, Fast Fixed-Point Algorithms for Independent component analysis, Neural Computation, 9:1483-1492, 1997

[39] 徐 勇, 张重阳, 杨静宇基于主分量特征与独立分量特征的人脸识别实验计算机工程与设计 2005 年 5 月第 26 卷第 5 期

[40] Wiskott L, Fellows J M, N KruK ger, et al. "Face Recognition by Elastic Bunch Graph Matching". IEEE Trans. on Patern. Anal. Mach. Intell., 1997, 19:775-779

[41] 曾述兵, 基于弹性图匹配的正面人脸识别技术研究, 20050501

[42] John Shawe-Taylor, Nello Cristianini, 模式识别的核方法, 机械工业出版社 2006

[43] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines", Cambridge, UK: Cambridge Univ. Press, 2000.

[44] Christopher J. C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition,

Data Mining and Knowledge Discovery 2,121-167,1998

- [45] 张学工, 关于统计学习理论与支持向量机[J].自动化学报, 2000 (26): 32-42
- [46] 邓乃扬, 田英杰, 数据挖掘中的新方法 [M]:支持向量机.北京:科学出版社
- [47] Nello Cristianini, John Shawe-Taylor, 李国正, 王猛, 曾华军, 支持向量机导论[M] 北京 电子工业出版社, 2004.
- [48] 张国宣, 基于统计学习理论的支持向量机分类方法研究, 2004
- [49] 黄国宏, 邵惠鹤, 核主成份分析及其在人脸识别中的应用, 计算机工程, 2004, 7
- [50] Bernhard Scholkopf, Alexander Smolla, and Klaus-Robert Muller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, Neural Computation, 1998, 10(5):1299
- [51] Francis R Bach, Michael I Jordan, Kernel Independent Component Analysis[J]. Journal of Machine Learning Research, 2002; (3):1-48
- [52] 姚伏天, 金连甫, 戴光, 基于核独立成分分析的盲源信号分离, 计算机工程与应用, 2004/40/6
- [53] 尹克重, 龚卫国, 李伟红, 梁毅雄, 张红梅, 基于核独立成分分析的人脸识别研究, 计算机应用 2005/25/6
- [54] 边肇祺, 张学工等, 模式识别[M], 第二版, 北京:清华大学出版社, 2000
- [55] 阮秋琦, 数字图像处理[M], 电子工业出版社, 2004
- [56] 袁晓琴, 黄凤岗, 张健沛, 基于 DCT 的人脸特征提取, 应用科技 2003, 4
- [57] 多传感器信息融合技术述评, 郁文贤, 雍少为, 国防科技大学学报, 1994
- [58] 多传感器信息融合及其应用综述, 王耀南, 李树涛, 控制与决策, 2001
- [59] 基于奇异值分解和数据融合的脸像鉴别, 王蕴红, 谭铁牛, 计算机学报, 2000
- [60] 决策层信息融合的神经网络模型与算法研究, 黎湘, 郁文贤, 电子学报, 1997
- [61] MING LI, ZHI-YUN LIU, A Novel Method Of Recognition Based On The Fusion Of Classifiers, Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006
- [62] Zhong-Qiu Zhao, De-Shuang Huang and Bing-Yu Sun, Human Facial Recognition Based On Multiple Feature Domains Proceedings of the 5<sup>th</sup> World Congress on Intelligent Control and Automation, June 15-19. 2004, Hangzhou. P.R. China
- [63] 机器学习与智能决策支持系统, 杨善林, 倪志伟, 2004, 科学出版社
- [64] T Mitchell, B Buchanan, G DeJong, T Dietterich, P Rosenbloom, and A Waibel, Machine Learning Vol. 4: 417-433 (Volume publication date June 1990)
- [65] Vladimir N V.统计学习理论的本质[M], 张学工译, 北京, 清华大学出版社, 2000
- [66] 吴雄华, 改进的 SVM 模式识别方法, 2004/05/01
- [67] 孔锐, 基于核的学习方法及其在人脸识别中的应用研究, 2004/05/01
- [68] 马宏伟, 吴力文, 王宇, 一种支持向量分类机的最优参数选择方法
- [69] 董春曦, 饶鲜, 杨绍全, 徐松涛, 支持向量机参数选择方法研究系统, 工程与电子技术, 2004 年 8 月
- [70] 张燕昆, 基于核方法的人脸识别技术的研究[D] 上海 上海交通大学 2003
- [71] 惠康华, 李春利, 基于再生核  $H^1_2$  支持向量机的模式分类研究, 计算机工程, 2005 年 7 月
- [72] Eberhart R C, Kennedy J. A New Optimizer Using Particles Warm Theory [A]. Proceedings of the Sixth International Symposium on Micro Machine and Human Science [C]. New York, NY, USA: IEEE, 1995. 39-43.

[73] Shi Y, Eberhart R C, A Modified Particle Swarm Optimizer [A]. Proceedings 1998 IEEE International Conference on Evolutionary Computation [C] New York, NY, USA: IEEE, 1998.69~73.

[74] 陈国初, 俞金寿, 微粒群优化算法信息与控制, 第 34 卷, 第 3 期, 2005 年 6 月

## 作者简历

1982 年 12 月 18 日出生于山西省太原市。

2001 年 9 月考入哈尔滨工程大学电子信息工程专业，2005 年 7 月获工学学士学位。

2005 年 9 月考入北京交通大学计算机学院信息所，攻读模式识别与智能系统专业至今，攻读工学硕士学位。

发表论文一篇：

《基于离散微粒群算法的 SVM 参数选择》

《计算机技术与发展》2007. 12



# 人脸识别算法研究和实现

作者：[马驰](#)

学位授予单位：[北京交通大学](#)

本文链接：[http://d.g.wanfangdata.com.cn/Thesis\\_Y1229286.aspx](http://d.g.wanfangdata.com.cn/Thesis_Y1229286.aspx)