

华南理工大学

硕士学位论文

一种基于视频的实时人脸识别系统的研究与实现

姓名：张鞠鹏

申请学位级别：硕士

专业：通信与信息系统

指导教师：金连文

20060501

## 摘要

由于计算机人脸识别技术具有广阔的应用前景，对计算机人脸识别方法和技术的研究已经成为国内外模式识别领域内研究的热点。本文以人脸识别为主线，研究了人脸识别中人脸检测和定位、光照和几何归一化、特征提取和特征选择等问题，设计和实现了一个基于视频的实时人脸识别系统。本系统使用普通的 USB 摄像头采集视频图像，采用基于类 Haar 特征的层叠式分类器训练和检测视频中的人脸，通过光照和几何归一化等预处理后，利用 Gabor 滤波器提取表征人脸的特征，再经过降维，训练出一系列模板，最后使用距离分类器通过设计好的策略进行分类识别，辨别出视频图像中人物的身份。

本文首先概要介绍了人脸识别的研究背景和常用的人脸识别方法，包括基于几何特征与模板匹配的人脸识别方法，基于特征脸的识别方法，基于人脸图像低频带信息的识别方法和基于神经网络与支持向量机的识别方法，并指出它们的优缺点及使用范围；然后介绍了几种目前应用最广泛的人脸检测方法，包括基于肤色特征的方法，基于模板匹配的方法，基于特征脸的方法和基于神经网络的方法，并通过 OpenCV 计算机视觉库设计了一个基于类 Haar 特征的层叠式分类器，实现了实时的人脸检测；针对不同光照条件下的人脸识别效果比较差的问题，详细介绍了几种不同的光照归一化方法，并对这些方法做了大量的实验，通过实验对比总结不同归一化方法对人脸识别的影响；同时为了解决 Gabor 特征维数较大的问题，通过对几种典型的特征选择算法的分析和实验对比，选取了 PCA+LDA 作为系统的特征降维的解决方案；最后设计和编程实现了一个基于视频的实时人脸识别系统，该系统可以完成图像捕捉、人脸检测、人脸定位、光照和几何归一化、Gabor 特征提取、降维和人脸识别的全过程。

本文所设计和实现的人脸识别系统可以用于视频监控、安全门禁、身份验证等安全领域，也可以用于视频游戏、智能玩具等娱乐领域，具有广泛的应用前景。

**关键词** 人脸识别；人脸检测；实时系统；光照归一化；特征选择

## Abstract

Face recognition methods have been the hotspot of the research in pattern recognition field all over the world since face recognition by computer can be widely applied on many aspects in the society in the future. Face recognition is the base line in this paper. Following this base line, face detection and location, illumination and position normalization, feature extraction and selection are studied and a real-time video-based face recognition system is designed and applied. In this system, the video images are captured by USB camera at first. Then a boosting classifier with Haar-like feature is trained and used to detect faces from the video. After the processing of illumination and position normalization, facial feature is extracted by Gabor filter and PCA+LDA. At last, the face from the video is recognized by a distance classifier with trained templates using a pre-determined strategy.

This thesis firstly introduces the research background and the technology of face recognition briefly. This part includes geometry feature, template matching method, eigen-faces method, facial information of low frequency method, neural network and support vector machine. Then several face detection methods are presented including the method based on skin color, the method based on template matching method, the method based on eigen-faces and the method based on neural network. A boosting classifier is designed to detect faces using OpenCV library. This classifier is based on Haar-like feature. To solve the problem of the bad recognition performance in different illumination conditions, several kinds of illumination normalization methods are introduced in detail and a great deal of experiments are done to compare these methods. Meanwhile, to solve the high dimension problem of Gabor feature, we compare several typical algorithms of feature selection by experiments, and PCA+LDA is used as the solution of reducing feature dimension in our system. Finally, a real-time face recognition system is designed and implemented. This system can achieve a whole process of face recognition.

The face recognition system designed and implemented in this thesis has broad perspective, which can be used in the area such as video surveillance, safety entrance, and identity verification etc. It can also be used for entertainment such as video game, intelligent toy and so on.

**Key Word** Face recognition; Face detection; Real-time system; Illumination normalization; Feature selection

# 华南理工大学

## 学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：张鞠鹏

日期：2006年6月2日

## 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华南理工大学。学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅（除在保密期内的保密论文外）；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。本人电子文档的内容和纸质论文的内容相一致。

本学位论文属于：

☐ 保密，在\_\_\_\_\_年解密后适用本授权书。

☒ 不保密。

学位论文全文电子版提交后：

☒ 同意在校园网上发布，供校内师生和与学校有共享协议的单位浏览。

（请在以上相应方框内打“√”）

本人签名：张鞠鹏

导师签名：廖

日期：2006.6.2

日期：2006.6.2

## 第一章 绪论

### 1.1 人脸识别简介

#### 1.1.1 生物特征识别技术

生物识别技术是利用人体生物特征进行身份认证的一种技术。人体的很多生物特征不但是唯一的，而且是可测量的，这使得计算机进行生物特征识别成为可能。与传统的身份认证技术相比，基于生物识别的身份认证技术具有以下特点：不易遗忘或丢失；防伪性能好，不易伪造或被盗；“随身携带”，随时随地可用。

人体生物的生物特征包括生理特征和行为特征。人体的生物特征主要包括人脸、指纹、掌纹、掌形、虹膜、视网膜、静脉、DNA、颅骨等，这些特征是与生俱来的，是先天形成的；而行为特征包括声纹、签名、步态、耳形、按键节奏、身体气味等，这些特征是由后天的生活环境和生活习惯决定的。

#### 1.1.2 人脸识别

人脸识别是计算机分析包含人脸图像，并从中提取出有效的信息，从而识别人物的身份的技术。人脸识别作为一种人体生物特征识别技术，它涉及人工智能、模式识别、计算机视觉、图像处理和分析、图像编码、计算机图形学等众多学科领域<sup>[1]</sup>。

与其它较成熟的人体生物特征识别方法（如指纹、虹膜、DNA检测等）相比，人脸识别系统有以下几个优点：

(1) 无侵犯性，人脸图像的获取不需要被检测人发生身体接触，可以在不惊动被检测人的情况下进行；

(2) 低成本、易安装，人脸识别系统只需要采用普通的摄像头、数码摄像机或手机上的嵌入式摄像头等被广泛使用的摄像设备即可，对用户来说也没有特别的安装要求；

(3) 无人工参与，整个人脸识别过程不需要用户或被检测人的主动参与，计算机可以根据用户预先的设置自动进行。

由于具有以上优点，近几年来，人脸识别技术引起了越来越多科研人员的关注<sup>[2]</sup>。但是，人脸识别也是一项极具挑战性的工作，其困难主

要体现在：

(1) 受环境（特别是光照和背景）的影响<sup>[3]</sup>，同一个人在不同的采样条件下捕捉到的图像有很大的差别；

(2) 受发型、年龄的影响；

(3) 受化妆、饰物的变化的影响；

(4) 受人脸姿态的影响，不同侧面的人脸的特征有很大的差异；

(5) 人脸的相似性，不同人的脸部具有相似的结构和纹理特征，识别系统只能通过极其微小的差别来进行识别，这对特征的选取提出了很高的要求。

## 1.2 研究背景和意义

最早的人脸识别技术的研究可以追溯到20世纪50年代，当时的研究人员主要涉及的是社会心理学领域；到了60年代，开始有一些工程文献陆续发表出来；但是，真正的自动人脸识别的研究是从20世纪70年代的Kanade<sup>[4]</sup>和Kelly<sup>[5]</sup>开始的，当时采用的技术基本上都是典型的模式识别技术，例如利用脸部重要特征点之间的距离进行分类识别。

随着计算机技术的发展，从80年代到90年代初期，人脸识别技术得到了很大的发展并进入了实际应用领域。在这一阶段，基于人脸外貌的统计识别方法得到了很大的发展，其中Eigenfaces<sup>[6]</sup>和Fisherfaces<sup>[7]</sup>在大规模的人脸数据库上进行的实验得到了相当不错的结果。同时，基于人脸特征的识别方法<sup>[8]</sup>也逐渐发展起来，此类方法对光线和视角的变化、人脸的定位都不太敏感，有利于识别率的提高，但是其所采用的特征提取方法还不够成熟和可靠。

自90年代后期以来，一些商业性的人脸识别系统逐渐进入市场；近几年来，人脸识别作为计算机安全技术在全球范围内迅速发展起来，特别是美国遭受911恐怖袭击以后，人脸识别技术更引起了广泛的关注。在这一阶段，更多的研究集中在基于视频的人脸识别上面。

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。表1-1列出了人脸识别的一些典型的应用。

表 1-1 人脸识别的典型应用

Table 1-1 The typical application of face recognition

| 应用领域 | 典型应用实例                    |
|------|---------------------------|
| 保安监控 | 视频监控、小区入口门禁、嫌疑犯跟踪和调查等     |
| 智能卡  | 银行卡、身份证、驾驶证、护照等           |
| 信息安全 | 个人设备登录、互联网接入控制、文件加密、家长限制等 |
| 娱乐   | 视频游戏、虚拟现实、驾驶训练等           |

### 1.3 人脸识别的常用方法

#### 1.3.1 基于几何特征与模板匹配的正面人脸识别方法

将人脸用一个几何特征矢量表示，根据模式识别中层次聚类的思想设计分类器达到识别目的。这就要求选取的几何特征矢量具有一定的独特性，能够反映不同人脸之间的差别，同时又具有一定的弹性，以消除时间跨度、光照等影响。几何特征矢量是以人脸器官的形状和几何关系为基础的特征矢量，其分量通常包括人脸指定两点间的欧氏距离、曲率、角度等。

Poggio和Brunelli<sup>[9]</sup>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。人脸器官的关键点分别对应于不同的图像灰度积分投影方式产生出的波峰波谷，根据人脸结构的先验知识，可以得出人脸各器官之间的几何位置关系。Poggio和Brunelli的系统获得90%以上的识别率，他们还将Baron<sup>[10]</sup>的模板匹配法加以改进，设计了一种自动模板匹配识别系统。Yuille等人<sup>[11,12]</sup>使用了另一种几何特征提取机制，即利用可变形模板(deformable templates)进行特征提取。Huang和Chen则在文献<sup>[13]</sup>中进一步提出了用可变形模板模型与活动轮廓模型(active contour models)进行人脸几何特征细节提取的算法，将模型的参数直接作为人脸特征。可变形模板是参数化的简单图元模型，根据先验知识定义的能量函数及相应的图像约束条件指导其收敛过程，主要用于提取眼睛和嘴巴轮廓<sup>[11,14]</sup>；而活动轮廓则是能量最小化的曲线<sup>[15,16,17]</sup>，由能量函



数指导其收敛到局部边界，主要用于提取眉毛和下巴轮廓<sup>[18,19]</sup>。用可变形模板与活动轮廓提取出的人脸器官轮廓，对人脸具有较为精确的描述能力，为几何特征矢量的生成打下了坚实的基础。另外，Chow和Li采用了自适应霍夫变换(Adaptive Hough Transform)与动态模板相结合的方法<sup>[20]</sup>。同Huang和Chen的方法类似，动态模板的作用是提取眼睛与嘴巴轮廓，而霍夫变换则主要用于确定眼球、下巴的轮廓。近3年来，这类方法得到了不断的发展，如山世光、高文等人<sup>[21]</sup>提出了一种基于面部纹理分布和可变形模板的由粗到细的面部特征提取策略；Kim等人<sup>[22]</sup>提出了一种使用可变形模板提取3D人脸特征及实时归一化方法；Pardas等人<sup>[23]</sup>提出了基于活动轮廓跟踪和定位人脸特征系统；Nikolaidis等人<sup>[24]</sup>则将可变形模板、活动轮廓与自适应霍夫变换组合在一起提出了一种提取人脸特征及确定凝视方向的方法，等等。

使用可变形模板、活动轮廓的方法，使得人脸器官模型参数化，对人脸具有较好的描述能力，但也存在两个问题，一是所使用的能量函数中各种代价的加权系数的选择是经验性的，难以推广；二是能量函数优化过程很耗时，不适于实时应用。

### 1.3.2 基于特征脸的正面人脸识别方法

这类识别法仍然是将人脸用代数特征矢量表示：即人脸图像在由“特征脸”(Eigenfaces)张成的降维子空间上的投影。Sirovich和Kirby<sup>[25]</sup>首先将KL变换用于人脸图像的最优表示。Turk和Pentland<sup>[10]</sup>进一步提出了“特征脸”方法，该方法以训练样本集的总体散布矩阵为产生矩阵，经KL变换后得到相应的一组特征矢量，因其对应的图像很像人脸，被称为“特征脸”。这样，就产生了一个由“特征脸”矢量张成的子空间，每一幅人脸图像向其投影都可以获得一组坐标系数，这组坐标系数表明了人脸在子空间中的位置，实验表明其具有较强的稳定性，可以作为人脸识别的依据。在传统的特征脸方法基础上，研究者们发现选择特征值大的特征矢量并非一定能得到最好的特征分类性能，因此提出了多种特征子空间方法，如Swets等人的线性判别分析法<sup>[26]</sup>，Peng等人的双子空间法<sup>[27]</sup>，Belhumeur等人的Fisherface方法<sup>[28]</sup>。“特征脸”分析法又称为主分量分析法(Principal Component Analysis, PCA)<sup>[29]</sup>。尽管特征脸算法是目前人脸识别技术的主流方法，但它对人脸图像的归一化要求较高，其性能易受到光照和姿态的影响，因此近几年来学者们提出了各种改进版本或扩展算法，如，Yang等人<sup>[29]</sup>提出了基于高阶统计的PCA方法；Pujol等

人<sup>[30]</sup>利用人脸样本拓扑关系提出了一种拓扑PCA方法；Takeshi等人<sup>[31]</sup>则将特征脸分解到两个正交的特征空间中以实现不同的光照条件下鲁棒性识别；张辉等人<sup>[32]</sup>采用对称主元分析神经网络，用去冗余和权值正交相结合的方法对人脸进行特征提取和识别；彭辉等人<sup>[33]</sup>也提出了特征脸的进一步改进方法，采用类间散布矩阵作为产生矩阵，以降低运算量。Brennan等人<sup>[34]</sup>的算法则组合了多精度、PCA方法。

### 1.3.3 基于人脸图像低频带信息的正面人脸识别方法

早在二十世纪七十年代的研究<sup>[35,36]</sup>就表明：脸像的低频带信息在人脸识别中起到最主要的作用，1986年，Sergent在文献<sup>[37]</sup>中进一步阐明：脸像的低频带和高频带起着不同的作用。低频成份的作用主要对应于脸像的全局(整体)描述，而高频成份的作用主要对应于脸像的局部细节的描述，人脸是非刚性的，富含各种脸部表情，脸部表情变化体现在脸部的局部器官上。Nastar等人在文献<sup>[38,39]</sup>中研究了面容的变化及其变形频谱之间的关系，发现脸部表情变化和小的遮挡物引起脸像灰度分布的局部变化，主要影响脸像频谱中的高频成份；而人脸在图像中的姿态和比例大小引起脸像灰度分布的全局变化，主要影响脸像频谱中的低频成份；只有不同的人脸像才会引起脸像频谱中的所有频率成份的改变。Bow在文献<sup>[40]</sup>中证明了如将脸像频谱划分成两半，使用低频的一半频谱就可很好地重构原脸像。由此可见，去除脸像频谱中的高频成份可减小脸部表情变化所产生的影响，并且脸像频谱中的低频成份足以胜任识别任务。小波变换已被证明是图像分析和特征提取最有效的工具之一，图像可通过小波变换很好地表征在空间与频率域中。小波变换中的多分辨率分析，通过对脸像的多层小波分解得到的低频子带图像在文献<sup>[41,42]</sup>中已证明是非常适合用于人脸识别的特征量。小波变换提取人脸特征，具有以下特点<sup>[41,42]</sup>。

(1) 人脸图像经 $k$ 层小波分解后的低频子带图像像素点数比原图像降低了 $4k$ 倍；

(2) 小波分解后适当大小的低频子带人脸图像对表情变化不敏感；

(3) 小波分解后适当大小的低频子带图像尽管失去了高频成份但却保持了充分地区分不同人脸的能力。

另外，正交小波变换有快速算法，这一点也非常有利于实际应用。

### 1.3.4 基于神经网络与支持向量机的正面人脸识别方法

基于神经网络的正面人脸识别方法研究一直方兴未艾。Cottrell和Fleming<sup>[43]</sup>采用了非线性前馈神经网络作为分类器,利用BP算法对人脸灰度图像训练后用于识别。他们采用的前馈式神经网络具有40个输入单元,80个隐层单元,40个输出单元,神经元的响应函数为Sigmoid函数。该网络实际上实现了对多幅人脸图像的压缩编码(编码信息同样储存在权值中),对人脸识别与生理分类均具有较高的准确率。Intrator等人<sup>[44]</sup>提出了一种混合型神经网络对人脸进行识别,其中的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。Lawrence等人<sup>[45]</sup>采用卷积神经网络对人脸进行识别,由于卷积神经网络中集成了图像中相邻像素之间的相关知识,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,因此得到了较好的识别结果。其它研究还有:Dai等人<sup>[46]</sup>用Hopfield神经网络进行低分辨率人脸联想与识别;Li和Liu用径向基函数(RBF)神经网络对用小波分解和特征脸法提取的人脸特征进行分类。神经网络在人脸识别上的应用与前述的几类方法相比有一定的优势,因为对人脸识别的许多规律或规则进行显性描述是很困难的,而神经网络则可以通过学习的过程获得对这些规律或规则的隐性表达。

然而,到了90年代,神经网络学习的研究也遇到一些困难,比如如何确定网络结构的问题、过学习与欠学习问题、局部最小点问题,等等。于是,在这种情况下,试图从更本质上研究机器学习问题的统计学习理论逐步得到重视。V. Vapnik及其研究小组于1992年至1995年,在统计学习理论的基础上发展出一种新的模式识别方法,即支持向量机(Support Vector Machine,简称SVM)<sup>[47,48]</sup>,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,如泛化(generalization)性能<sup>[48]</sup>好,无需先验知识等,并能够推广应用到函数拟合等其它机器学习问题中,因此近年来SVM的研究成为神经网络研究之后机器学习领域新的研究热点。人脸识别问题本质上属于小样本、非线性模式识别问题,因此,近几年出现不少利用SVM作为人脸特征分类器的文献。如, Lee<sup>[50]</sup>提出了用遗传算法提取人脸特征,然后再用支持向量机对人脸特征进行分类的方法。Guo等人<sup>[51]</sup>则先用特征脸法提取人脸特征,然后再用支持向量机对人脸特征进行分类的方法。Jonsson等人<sup>[52]</sup>还研究了在不同的人脸表征情况下支持向量机的分类性能,验证了支持向量机具有从脸像训练样本中提取可分辨信息的能力。这些文献还将SVM与传统的分类器分类性能作比较,来验证利用SVM对人脸特征进行分类的优势。尽管支持向量机有

许多特有的优势，但仍有很多问题需要进一步研究解决。支持向量机作为分类器实质上是将模式分类问题归结为一个二次规划问题，而这类问题的求解算法，也就是支持向量机的学习算法多趋于复杂。当问题规模较大时，通常需要耗费相当可观的计算时间和存贮空间，学习算法速度过慢，有时甚至导致算法无法执行。因此研究快速学习算法是支持向量机在实用中急待解决的问题之一。

### 1.3.5 其它方法

90年代中期以来，研究人员逐渐认识到，一种有效的人脸识别算法应该充分地挖掘人脸不同方面的特征信息，即有效地利用人脸的形状拓扑结构特征、局部灰度特征和全局灰度分布特征等多种特征。因此出现了融合上述特征信息的一些新算法，如Lyons等人<sup>[53]</sup>采用的基于小波和弹性图匹配（elastic graph matching）的人脸识别算法就是一种人脸的形状拓扑和局部灰度特征的算法。这类算法较好地利用了人脸的结构信息和局部灰度分布信息，而且还具有自动精确定位面部特征的功能，因而具有良好的识别效果。其缺点是时间复杂度高，实现复杂。此外，Hafed等人<sup>[54]</sup>提出了利用DCT对归一化人脸图像提取人脸特征方法。Lai等人<sup>[42]</sup>的人脸识别算法，利用小波分解与Fourier变换提出了人脸脸谱（Spectroface）表征法，脸谱特征对人脸在图像平面内的位置、旋转、尺寸变化具有一定的不变性，对表情变化也有一定的鲁棒性。隐马尔可夫模型（HMM）<sup>[55]</sup>、奇异值分解（SVD）<sup>[56]</sup>、独立元分析（ICA）<sup>[49]</sup>等方法也在人脸识别中得到了应用。

## 1.4 本文的主要工作及内容安排

本文的主要工作是研究了人脸检测和识别的基本理论和关键技术，重点讨论了在基于视频的彩色图像和灰度图像中的人脸检测、光照归一化、特征提取和特征选择、人脸识别等问题，并提出了一个基于视频的实时人脸识别系统的设计和实现方案。

基于视频的实时人脸识别系统的主要目的是建立一个实时系统，计算机通过摄像头摄取图像，然后从视频中检测出人脸的位置，并对捕获的人脸图像进行光照、大小、位置的预处理，接着提取人脸的特征并选出有用的特征，最后对所选取的特征进行训练和识别，判断出该特征所属的人脸类别，识别人的身份。

本文将用如下五章论述所做的工作。第一章为绪论部分，对生物特征识别技术涉及的内容进行简要的介绍，简述本文的研究背景和意义，对人脸识别方法进行归纳和总结。第二章是人脸检测与定位，将对比几种人脸检测的方法，介绍基于类 Haar 特征的层叠式分类器，并叙述利用 Intel OpenCV 库实现人脸检测系统的方法。接下来的第三章将介绍图像预处理和归一化过程，重点对比了几种不同的光照归一化方法及其实验结果。第四章介绍特征提取和特征选择的方法，并列出了几种特征降维方法，并对比它们的实验结果。第五章将详细介绍基于视频的实时人脸识别系统的设计和实现。最后，对整个研究工作做了总结和回顾，并提出了一些对于后继工作的难点和展望，发作为论文和结束语。

## 第二章 人脸检测算法介绍及其实现

### 2.1 人脸检测概述

人脸检测是指在输入图像中确定是否有人脸存在，如果有人脸存在，则确定其数目和每个人脸的位置和大小<sup>[1]</sup>。人脸检测作为人脸信息处理中的一项关键技术，是人脸识别系统中的第一个环节，能否正确地从一幅图像或视频流中检测出人脸的位置，对系统后续的特征提取和识别有着重大的影响。

同时，人脸检测的研究也具有重要的学术价值。人脸是一类具有相当复杂的细节变化的自然结构目标，此类目标检测问题的挑战性在于：

(1) 由人脸和摄像头的位置之间的差别导致的人脸图像的不同，如获得了人脸正面、侧面、45度角的图像；(2) 人脸上可能存在眼镜、胡须等附属物，包括它们不同的形状、大小和颜色的影响；(3) 人的外貌、表情和肤色的影响；(4) 摄影的条件，如光照的强度、光源的分布、摄像机的特性等的影响。因此，如果能够找到解决这些问题的方法，成功构造出人脸检测与跟踪系统，将为解决其它类似的复杂模式检测问题提供重要启示<sup>[57]</sup>。

人脸检测所包含的问题十分广泛，从不同的角度可以有多种分类方法（表 2-1）。

从一幅图像中检测人脸的方法可以分为以下几类：基于肤色特征的方法、基于模板匹配的方法、基于特征脸的方法、基于神经网络的方法。

### 2.2 基于肤色特征的方法

肤色是人脸的重要信息，不依赖于面部的细节特征，对于旋转，表情等变化情况都能适用，具有相对的稳定性并且和大多数背景物体的颜色相区别。因此肤色特征在人脸检测中是最常用的一种特征。肤色特征主要由肤色模型描述。

在使用肤色模型进行描述之前，我们必须了解计算机表示颜色的方法、颜色空间的概念、不同颜色空间的用法及不同颜色空间之间相互的转换，因此，我们在下面将简单介绍颜色空间的概念等。

表 2-1 人脸检测问题的分类  
Table 2-1 Category of Face Detection

| 分类依据    |      | 类别   |   |
|---------|------|--|---|
| 图像类型    | 图像来源 | 静止图像（包括如数字化的照片、数码相机拍摄的图片有情，目前考虑的主要问题是算法的适应性的鲁棒性，算法速度在其次） | 动态图像（即视频序列，包括工作台前的人脸序列、报案监控录像、影视资料等，往往与人脸的跟踪问题交织在一起，对算法的速度有很高的要求） |
|         | 颜色信息 | 彩色   | 灰度  |
| 图像前景    | 镜头类型 | 头肩部图像  | 半身/全身图像   |
|         | 人脸姿态 | 正面（包括端正及平面内旋转）   | 侧面（包括俯仰、侧影及旋转）  |
|         | 人脸数目 | 单人（又可以称为人脸定位，是人脸检测问题在已知人脸树木情况下的特例）                       | 未知（需要判定图像中是否存在人脸，人脸的树木以及各个人脸的尺度和位置，即是完全的检测问题）                     |
| 图像背景复杂度 |      | 简单背景（指无背景或背景的特征被严格约束，在该条件下只利用人脸的轮廓、颜色、运动等少量特征，就能够进行准确检测） | 复杂背景（指背景的类型和特征不受约束，某些区域可能在色彩、纹理等特征上与人脸类似，必须利用较多的人脸特征才能做到准确检测）     |
| 应用领域    |      | 人脸信息处理（验证、识别、表情分析等）系统，视频会议或远程教育系统，视频监视与跟踪，基于内容的图像与视频检索等等 |   |

### 2.2.1 颜色空间

光在频域分布中的不同位置正好体现了各种颜色，颜色是人对光谱中可见区域段光的感知结果。根据三基色原理，自然界常见的各种颜色光都可以用相互独立的三种基本颜色混合得到，这三种基本颜色就称作

三基色，由此就构成了三维彩色空间。例如最常见的由红、绿、蓝三基色构成的 RGB 彩色空间。彩色空间有许多中，下面我们介绍常用的几种彩色空间及其转换。

### 2.2.1.1 RGB 彩色空间

RGB 是最常用的彩色空间。国际照明委员会 (CIE) 采用红 (R)、绿 (G)、蓝 (B) 三种单色作为彩色空间的三个分量，这就是 RGB 彩色空间。如图 2-1 所示，三维空间中的三个轴分别与红、绿、蓝三基色对应，

原点对应于黑点，离原点最远的顶点对应于白点，而其他颜色则落在三维空间中有红、绿、蓝三基色构成的彩色立方体中。其他彩色空间一般以 RGB 彩色空间为基础，RGB 彩色空间的线性或非线性函数。

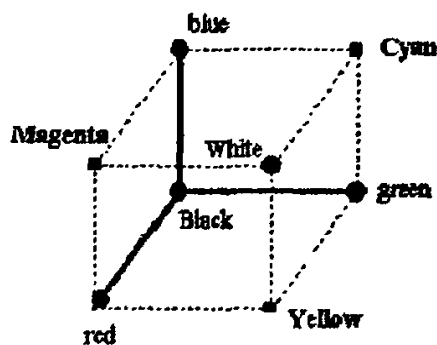


图 2-1 RGB 彩色空间

Figure 2-1 RGB Color Space

### 2.2.1.2 YUV 彩色空间

还有一种值得一提的，是 YUV 颜色模型了。在计算机里 YUV 颜色模型是仅次于 RGB 模型的使用最广泛的颜色模型。事实上 YUV 只是一类颜色模型 (YCrCb) 的总称，YCrCb 它是个很大的家族，具有相当多的存储格式，一般来说凡是基于 YCrCb 颜色模型的都可以称的上是 YUV。YCrCb 是一种由 RGB 转化来的亮度-色度模型，其中 Y 代表了光源的亮度，色度信息组合在 Cr、Cb 中，其中 Cr 代表了光源中的红色分量，Cb 代表了光源中的蓝色分量，因此它的变换公式如式 (2-1) 示：

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.1687 & -0.3313 & 0.5 \\ 0.5 & -0.4187 & -0.0813 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

$$R = Y + 1.402 * (C_r - 0.5)$$

$$G = Y - 0.34414 * (C_b - 0.5) - 0.71414 * (C_r - 0.5)$$

$$B = Y + 1.772 * (C_b - 0.5)$$

(2-1)

后面加上 0.5 使得 Cr, Cb 总为正数，便于处理。

如前所述，人眼对于亮度的敏感程度大于对于色度的敏感程度，所以完全可以让相邻的像素使用同一个色度值，而人眼的感觉不会起太大



的变化，通过损失色度信息来达到节省存储空间的目的，着就是 YUV 的基本思想。基于此，我们可以定义出许多 YUV 的格式，例如相邻两个像素使用一个色度值的 YUYV，JPEG/MPEG 中相临四个像素使用一个色度值的 YUV12 等等。

### 2.2.1.3 HSI 彩色空间

这一彩色空间中颜色的参数分别是：色彩（Hue），纯度（Saturation），明度（Intensity）。HSI 模型的三维表示从 RGB 立方体演化而来。设想从 RGB 沿立方体对角线的白色顶点向黑色顶点观察，就可以看到立方体的六边形外形。六边形边界表示色彩，水平轴表示纯度，明度沿垂直轴测量。HSV 对用户来说是一种直观的颜色模型。我们可以从一种纯色彩开始，即指定色彩角  $H$ ，并让  $V=S=1$ ，然后我们可以通过向其中加入

黑色和白色来得到我们需要的颜色。增加黑色可以减小  $V$  而  $S$  不变，同样增加白色可以减小  $S$  而  $V$  不变。例如，要得到深蓝色， $V=0.4$   $S=1$   $H=240$  度。要得到淡蓝色， $V=1$   $S=0.4$   $H=240$  度。

一般说来，人眼最大能区分 128 种不同的色彩，130 种色饱和度，23 种明暗度。如果我们用 16Bit 表示 HSV 的话，可以用 7 位存放  $H$ ，4 位存放  $S$ ，5 位存放  $V$ ，即 745 或者 655 就可以满足我们的需要了。

由于 HSI 是一种比较直观的颜色模型，许多情况下在该彩色空间中进行图像分析和处理比在 RGB 彩色空间中更加方便和有效。所以在许多图像编辑工具中应用比较广泛，如 Photoshop（在 Photoshop 中叫 HSB）等等。

在肤色检测中 HSI 彩色空间是比较常用的，下面是 RGB 和 HSI 彩色空间转换的公式[Smith 1978]：

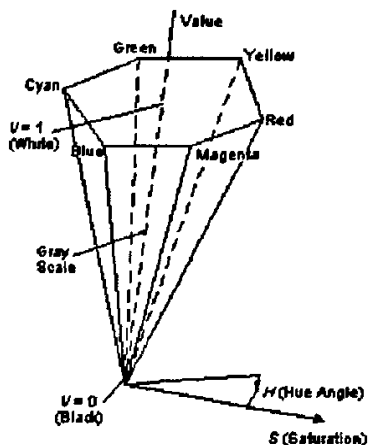


图 2-2 HSI 彩色空间

Figure 2-3 HSI Color Space

$$\begin{aligned}
 I &= \frac{1}{3}(R+G+B) \\
 S &= 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \\
 H &= \arccos \left\{ \frac{[(R-G)+(R-B)]/2}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\}
 \end{aligned} \tag{2-2}$$

## 2.2.2 肤色模型

使用何种形式的肤色模型与颜色空间的选择密切相关。可以从两个方面考察某种颜色空间<sup>[58]</sup>：(1) 在该颜色空间中能否用给定的模型描述“肤色”区域的分布；(2) 颜色空间中“肤色”与“非肤色”区域的重叠有多少。人脸检测常用的颜色空间如 2.2.1 节所述。

常用的肤色模型有高斯模型、混合高斯模型和直方图模型。Terrillon 等<sup>[58]</sup>考察了各种颜色空间及其归一化模型，比较了高斯模型和混合高斯模型在不同颜色空间中的性能，发现除了少数情况外，一般需要使用混合高斯模型才能较好地描述肤色区域的分布。Terrillon 等同时指出，最终限制检测性能的因素是不同颜色空间中“肤色”与“非肤色”区域的重叠程度。

除上述 3 种肤色模型外，还有直接利用几何参数描述肤色区域分布范围的模型<sup>[59]</sup>、三维投影模型<sup>[60]</sup>、基于神经网络的肤色模型<sup>[61]</sup>等。此外也有同时考虑“肤色”与“非肤色”像素分布的基于贝叶斯方法的模型<sup>[62]</sup>。

## 2.3 基于模板匹配的方法

模板匹配方法是指存储几种标准的人脸模式，用来分别描述整个人脸和面部特征；计算输入图像和存储的模式间的相互关系并用于检测。

### 2.3.1 预定义模板匹配的方法

一个标准的人脸模板首先被预定义和参数化，然后给定一幅输入图像，分别计算出脸部轮廓、眼睛、鼻子和嘴巴与模板之间的相关度，如果相关度在一定阈值之内，则可判定输入图像为人脸，反之，则可判定为非人脸。

Sakai 等<sup>[1]</sup>人使用眼睛、鼻子、嘴和人脸轮廓等子模板建模，检测照片中的正面人脸。每一个子模板按照线分割定义，基于最大梯度变化提

取输入图像的线，然后与子模板匹配，计算子图像和轮廓模板之间的相互关系去检测人脸的候选区域，完成用其他子模板在候选区域的匹配。

Craw 等<sup>[1]</sup>人提出了一种基于正面人脸的形状模板（也就是人脸的外形）的定位方法。用 Sobel 滤波器提取边缘，将边缘组织在一起，根据几个约束条件去搜索人脸模板。在头轮廓定位以后，用相同的过程以不同的尺度重复定位眼睛、眉毛和嘴唇等特征。

Govindaraju 等<sup>[63]</sup>人提出两个阶段的人脸检测方法。人脸模型根据边缘定义的特征构成，这些特征描述了正面人脸的左边、发线和右边的曲线。人脸必须是垂直、无遮挡和正面的。

### 2.3.2 改进的模板匹配的方法

预定义模板匹配的方法的优点是算法简单，容易实现。但是，它不能有效地适应人脸大小、位置和形状的变化。为了解决此类问题，多分辨率、多刻度、多模板和可变形模板等方法被相继提出。

Miao 等<sup>[64]</sup>人提出了用于人脸检测的层次模板匹配方法。在第一阶段，为了处理旋转图像，输入图像从 $-20^{\circ}$ ~ $20^{\circ}$ 旋转，每次旋转 $5^{\circ}$ 。多分辨率图像层次形成和边缘提取使用 Laplacian 操作符，人脸模板通过六个人脸成分产生的边缘组成：两个眼眉、两只眼睛、一个鼻子和一张嘴。最后，应用启发式确定人脸的存在。实验结果表明在图像含有单个人脸要比图像中含有多个人脸的结果好。

Yuille 等<sup>[65]</sup>人使用可变形模板建立一个人脸特征相匹配的先验弹性模型。这种方法用参数化的模板还描述人脸特征，用一个能量函数来定义输入图像的边界、峰值和谷值与模板的参数的对应关系，选择使能量函数最小的参数作为最匹配的弹性模型。虽然实验结果表明该方法在跟踪非刚性特征时具有良好的性能，但是该方法的不足是可变形模板必须在接近感兴趣的区域初始化。

## 2.4 基于特征脸的方法

模板匹配中的模板是由专家预定义的，与模板匹配中的方法不同，基于特征脸的方法中的“模板”是从图像中的样本学习的。通常，基于特征脸的方法依靠统计分析和机器学习技术找到相应的人脸和非人脸图像的特征。学习的特征由分布模型或判别函数形成，用于人脸检测，同时，由于计算效率和检测有效性的原因通常需要降维。对于降维方法，

我们将在第四章中详细讨论。

基于特征脸的方法是根据图像的统计特征进行正交变换，以消除原有样本之间的相关性，变换得到依次递减的特征值所对应的特征向量，即特征脸。

设  $N$  个人脸样本组成集合  $X: \{x_1, x_2, \dots, x_N\}$ ,  $\forall x_i \in R^n$ ，定义线性算子  $W \in R^{n \times m}$  使得：

$$y_i = W^T x_i, \quad i=1,2,\dots,N \quad (y_i \in R^m, m < n) \quad (2-3)$$

因为  $m < n$ ，所以样本集合  $X$  被压缩为  $Y$ ，定义  $X$  的整体离散度矩阵：

$$S_T = \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T, \quad \mu = R^n \quad (2-4)$$

为样本期望。则  $Y$  的离散度矩阵为  $W^T S_T W$ ， $W$  的选取应使  $Y$  的离散度矩阵模最大，即：

$$W_{opt} = \arg \max_W |W^T S_T W| = (w_1, w_2, \dots, w_m) \quad (2-5)$$

所以，取  $w_i$  是  $S_T$  的特征向量。如果将  $S_T$  的特征向量所对应的特征值从大到小排列，选取最大的  $m$  个特征向量组成  $W$ ，这种方法称为主分量分析 (Principle Component Analysis, PCA)。经过  $W$  变换后，每个样本表现为各个特征向量的加权线性组合。

Sung 和 Poggio<sup>[66]</sup>提出了基于事例学习的方法实现人脸识别，他们先对  $19 \times 19$  像素的“人脸”和“非人脸”样本利用主分量分解的方法得到特征脸，然后采用  $k$ -均值聚类方法在特征空间中建立 6 个“人脸簇”，同时建立包围“人脸簇”的 6 个“非人脸簇”，强化“人脸”与“非人脸”的边界，对每个检测窗口计算它与这 12 个簇的马氏距离，结果输入一个多层感应器进行判断。

特征脸技术主要是利用样本集合整体散射来评估转换算子  $W$ ，但是考虑到整体离散度包含了样本集合中类间散射和类内散射，应该尽量降低类内离散度，增大类间离散度，即不同类样本尽可能远，同类样本尽可能近。所以有学者<sup>[7]</sup>提出了特征脸的改进方法 Fisher 脸 (Fisherfaces)。

隐 Markov 模型 (Hidden Markov Model, HMM) 也被人们用于人脸检测<sup>[67]</sup>。在 HMM 用于模式识别问题时，需要确定许多隐藏的状态形成一个模式，然后，训练 HMM 从样本中学习传统状态间的概率，其每个样本被表示为观察序列。训练 HMM 的目的是通过 HMM 模型中的参数调整最大化观察的训练数据的概率。训练 HMM 模型以后，观察的输出概率确定了其所属的类别。

近年来，人们又提出了支持向量机 (Support Vector Machines, SVM)

和其他的核函数方法<sup>[68]</sup>。这些方法将模式投影到较高维数空间，然后在投影的人脸和非人脸模式之间形成一个决策面。SVM 分类器是线性分类器，所选择的分类超平面使未知测试样本的分类错误最小。最优超平面是通过权值结合训练向量的小子集（称为支持向量）来定义的，然而，其时间和内存的计算量是很大的。

利用特征空间变换的识别方法虽然在统计意义上可达到最优，但是应该看到这个最优是基于有限样本空间的，不能代表整个事例空间，而且由于样本图像通常按行列顺序展开形成一维向量，丢失了人脸的很多拓扑信息，用于识别的特征向量不仅包含了人脸的信息而且包含了亮度信息，所以主分量受光照的影响较大。目前，很多研究都针对自己的系统联合使用或改进了多种方法来消除光照对识别的影响，我们在第三章中将会详细介绍各种方法。

## 2.5 基于神经网络的方法

人工神经网络方法是把模型的统计特征隐含于神经网络的结构和参数中，对于人脸这类复杂的、难以显式描述的模型，基于人工神经网络的方法具有独特的优势。

CMU 的 Rowley 等<sup>[69]</sup>建立了一种三层向前网：输入层对应  $20 \times 20$  像素的图像区域；隐层分为 26 个单元，其中 4 个单元处理区域为  $10 \times 10$  像素，16 个单元为  $5 \times 5$ ，6 个单元为  $20 \times 5$ ，如图 2-3 所示。隐层的输出与输出层相连进行综合判断，输出区间  $[-1, 1]$  的值表示这个区域是否为人脸。为了解决人脸姿态的问题，Rowley 分别设置了三个姿态检测器，对输入图像先进行预处理，尽量消除光照的影响，然后送入姿态检测器，估计人脸的姿态，经校准和预处理后，送入三个检测器（正面人脸、半侧面和侧面），最后对检测器的结果进行综合判断。

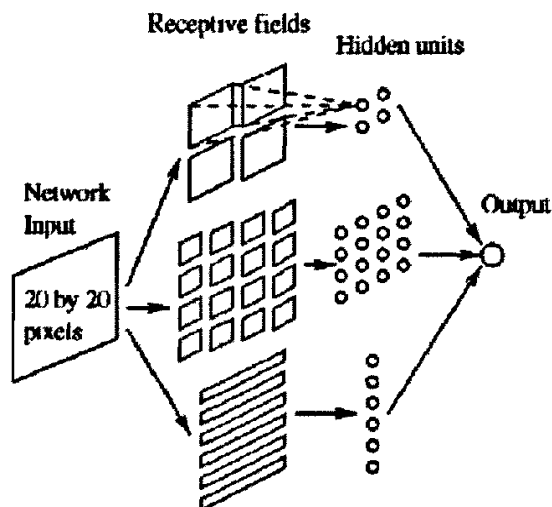


图 2-3 三层结构神经网络

Figure 2-3 Three layers structure of Neutral Network

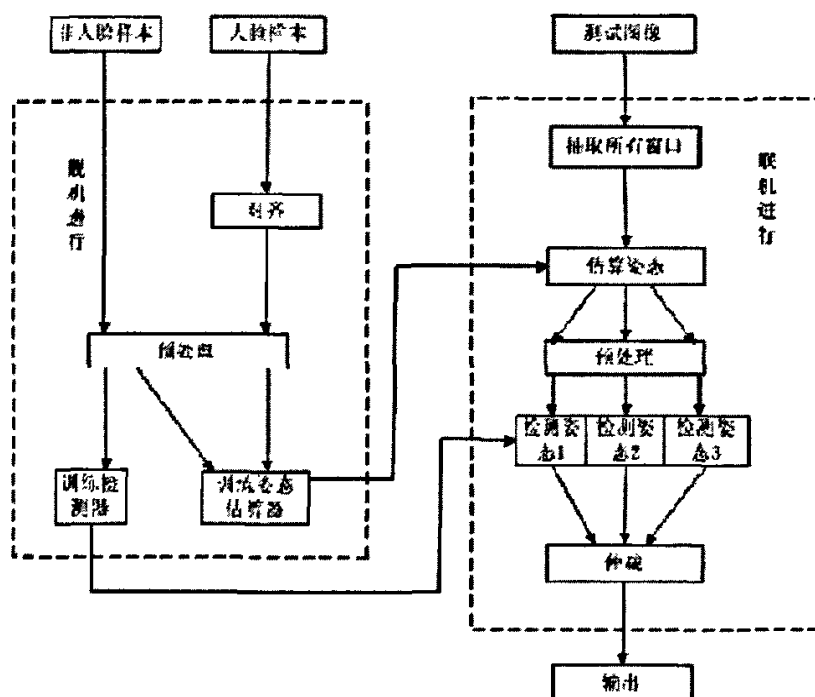


图 2-4 Rowley 等人提出的人脸检测系统

Figure 2-4 Face detection system issued by Rowley etc.

人工神经网络的方法便于建模，鲁棒性较好，但是运算速度较慢。Rowley 统计过单纯用人工神经网络的方法，在 170Mhz R10000 SGI 工作站上处理  $320 \times 240$  像素的图像大约需要 140s。

## 2.6 基于类 Haar 特征的分层式分类器

### 2.6.1 类 Haar 特征

类 Haar 特征是由 Papageorgiou 等<sup>[70]</sup>人在利用 Haar 小波变换从人脸中提取特征时，提出的局部 Haar 特征。类 Haar 特征通常由 2 至 4 个矩形组成，分别检测边界、细线或对角线特征。

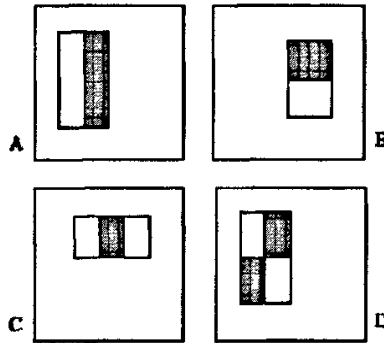


图 2—5 典型的类 Haar 特征

Figure 2—5 Typical Haar-like Feature

定义：

$$feature_k = \sum_{i \in \{1, \dots, N\}} \omega_i RctSum(r_i) \quad (2-6)$$

$\omega_i \in R$  为矩形  $r_i$  的权， $RctSum(r_i)$  所围图像的灰度积分， $N$  是组成  $feature_k$  的矩形个数。图 2—5 为 4 个典型的类 Haar 特征。其中 A 和 B 的特征值是两个具有相同大小和形状的水平或垂直排列的矩形区域的像素值之差，用于检测边界特征；C 的特征值是左右两个矩形区域的像素值之和减去中间矩形区域的像素值，用于检测细线特征；D 的特征值是由两个对角线上的矩形区域的像素和相减得到，用于检测对角线特征。

类 Haar 特征的优点是计算方便，而且适于多分辨率分析。类 Haar 特征可以通过图像的快速积分算法快速计算。

## 2.6.2 积分图像及其快速算法

积分图像<sup>[71]</sup>是指在点  $(x, y)$  的左上区域的所有像素的像素值之和，即

$$ii(x, y) = \sum_{x' < x, y' < y} i(x', y') \quad (2-7)$$

其中， $ii(x, y)$  为积分图像， $i(x, y)$  为原始图像。

$$\begin{aligned} s(x, y) &= s(x, y-1) + i(x, y) \\ ii(x, y) &= ii(x-1, y) + s(x, y) \end{aligned} \quad (2-8)$$

采用公式 2—8 的递归算法可以计算出原始图像的积分图像，其中  $s(x, y)$  为每行像素值的累积和， $s(x, -1) = 0$ ， $ii(-1, y) = 0$ 。

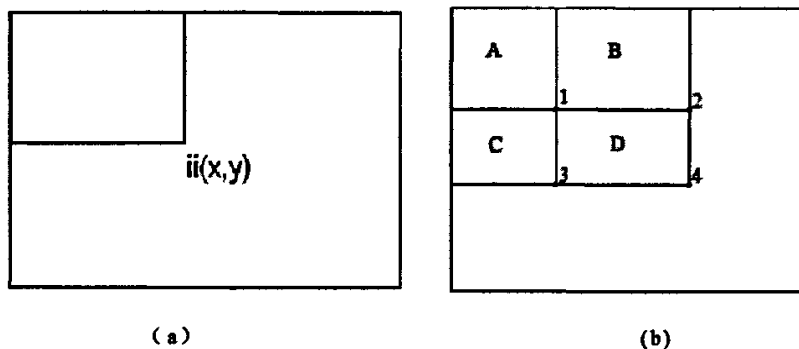


图 2—6 积分图像算法

Figure 2—6 Algorithm of Integral Image

使用积分图像的概念，任何一个原始图像上矩形区域的像素值之和均可由积分图像中相同的矩形区域的四个角的值来表示。如图 2—6 (b) 所示，积分图像中点 1 的值代表了原始图像中区域 A 的像素值之和，积分图像中点 2 的值代表了原始图像中区域 A+B 的像素值之和，积分图像中点 3 的值代表了原始图像中区域 A+C 的像素值之和，积分图像中点 4 的值代表了原始图像中区域 A+B+C+D 的像素值之和。因此，两个矩形区域之间像素和的差值可以由 8 个角点的积分值计算。对于由两个矩形组成的类 Haar 特征（如图 2—5 (A) (B)），只需通过 6 个点的值计算，三个矩形区域组成的需 8 个点，四个矩形区域也只需 9 个点。



### 2.6.3 AdaBoost 算法

Boosting 算法是 20 世纪 90 年代才提出来的, 利用该算法可以将弱学习算法提升为强学习算法。利用某学习算法来识别一组概念, 如果识别正确率很高, 那么这组概念是强可学习的; 如果识别正确率仅比随机猜测略好, 那么这组概念是弱可学习的。Kearns 和 Valiant<sup>[72]</sup>提出了弱学习算法与强学习算法的等价性问题, 即是否可以将弱学习算法提升成强学习算法。如果两者等价, 那么在学习概念时, 只要找到一个比随机猜测略好的弱学习算法, 就可以将其提升为强学习算法, 而不必直接去找通常情况下很难获得的强学习算法。

AdaBoost 算法<sup>[88]</sup>是 Freund 等人在 Boosting 算法的基础上提出的, 其主要思想是给定一弱学习算法和一训练集  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。这里  $x_i$  为第  $i$  个训练样本的输入,  $y_i$  为分类问题的类别标志。算法开始先给每一个训练样本赋以相等的权值  $1/n$ , 然后用该学习算法对训练集训练  $T$  轮, 每次训练后, 对训练失败的训练样本赋以较大的权值, 也就是让学习算法在后续的学习中集中对比较难的训练样本进行学习, 从而得到一个预测函数序列  $h_1, h_2, \dots, h_T$ , 其中  $h_j$  也有一定的权值, 预测效果好的预测函数的权值较大, 反之较小。最终的预测函数  $H$  采用有权值的投票方式对新样本进行判别。

AdaBoost 算法的具体描述为:

- 给定一组样本图像  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , 其中  $y_i=0, 1$  分别表示正确和错误的分类。
- 分别对正确和错误的分类初始化权重  $\omega_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ , 其中  $m$  和  $l$  分别代表正确和错误的分类的数目。
- 对于  $t=1, \dots, T$ :

1. 归一化权重  $w_{t,j} \leftarrow \frac{w_{t,j}}{\sum_{j=1}^n w_{t,j}}$ , 即  $w_{t,j}$  表示样本的概率分布;

2. 对于每一个特征  $j$ , 训练一个只使用单一特征的分类器  $h_j$ , 其分类错误率为  $\varepsilon_j = \sum_i \omega_i |h_j(x_i) - y_i|$ ;

3. 选择错误率  $\varepsilon_j$  最低的分类器  $h_j$ ;

4. 更新权重:  $\omega_{t+1,i} = \omega_{t,i} \beta_i^{1-\varepsilon_i}$ , 其中,  $\beta_i = \frac{\varepsilon_i}{1-\varepsilon_i}$ , 当  $x_i$  被正确分类时

$\varepsilon_i = 0$ , 否则  $\varepsilon_i = 1$ 。

- 最终的强分类器可以表示为：

$$h(x) = \begin{cases} 1 & \sum_{i=1}^T \alpha_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^T \alpha_i \\ 0 & \text{otherwise} \end{cases}$$

其中  $\alpha_i = \log \frac{1}{\beta_i}$ 。

#### 2.6.4 层叠式分类器

层叠式分类器是一个退化的决策树结构，树的每一个结点为一个训练好的分类器，该分类器可以剔除部分非目标模式，并检测出所有目标模式<sup>[71]</sup>（如人脸等），而且该分类器的正确目标识别率（hit rate，将目标模式归类为正确目标）非常高（接近 100%），误检率（false alarm rate，将非目标模式归类为正确目标）则在一个比较高的水平（如 30~40%）。如图 2—7 所示为一个有  $N$  个节点的层叠式分类器，假设每个分类器（结点）可以识别出 50% 的非目标模式且只有 0.2% 的目标模式被错误分类，即  $f=0.5$ ,  $h=0.998$ 。在  $N=13$  时，我们可以得到其误检率为  $0.5^{13} \approx 1.2e-04$ ，识别正确率为  $0.998^{13} \approx 0.97$ 。也就是说，经过了 13 个结之后，正确的目标模式有 97% 被保留了下来，而错误的目标模式只有不到 0.01% 被保留下来。

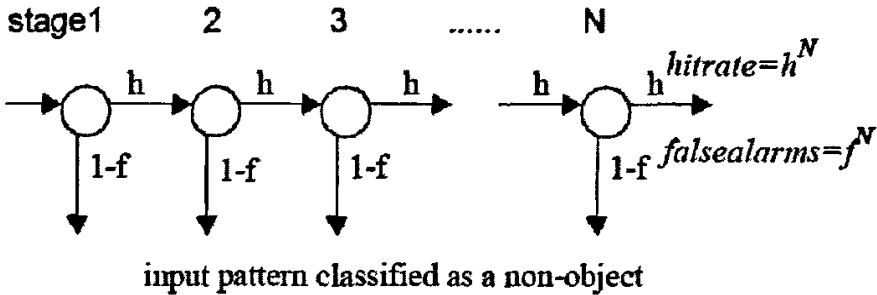


图 2—7 层叠式分类器

Figure 2—7 Boost Classifier

利用 AdaBoost 算法，我们只需要从所有特征中选出其中一个，训练出一组比随机猜测略好一点的分类器，按照权重组合成一个层叠式分类器。在 AdaBoost 的每一次迭代中，最好的分类器不断地被加入进来，随着迭代次数的增加，分类器数量也逐渐增加，直到达到所需要的误检率

为止。

## 2.7 人脸检测系统的实现

### 2.7.1 人脸特征的选取

根据 2.6.3 和 2.6.4 的方法，构造一个基于 AdaBoost 算法的层叠式分类器，其正确目标识别率为 95%，误检率为 1/14084，从 200 个类 Haar 特征中选取适合用于人脸检测的特征。Viola 等的实验结果表明<sup>[71]</sup>，在这些特征中，最适合于人脸检测的是“人的眼睛所在的矩形区域比眼睛下面的面颊区域的灰度值小”，这个特征主要对用于检测的特征窗口的大小比较敏感，而对人脸的大小及其在图像中所处的位置不太敏感；而排在第二位的特征是“两只眼睛的灰度值比两眼中间的鼻梁的灰度值要小”。

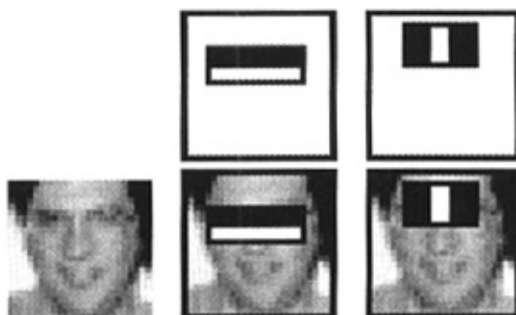


图 2—8 AdaBoost 算法选出的排在前两位的特征

Figure 2—8 The first 2 selection of AdaBoost algorithm

### 2.7.2 分类器的训练

分类器的训练需要从两方面考虑，使用更多的特征可以得到更高的正确检测率和更低的误检率，但是这就意味着需要更多的时间进行运算。因此，我们必须在分类器的阶数、每个弱分类所使用的特征数目和分类器的阈值这三方面选取一个最优的组合。

在实际训练时，我们采用一个简单的方法来训练一个高效的分类器。从单一特征开始，不断地在每一阶的弱分类器中加入特征，直到该阶弱分类器的正确目标识别率和误检率达到预先设定的要求；然后逐渐加大

分类器的阶数，直到总体的正确目标识别率和误检率达到设计的要求为止。

## 2.7.3 分类器的实现

### 2.7.3.1 检测快速算法

在人脸检测过程中，检测子窗口的尺寸是固定的（实验中是  $20 \times 20$  像素），但是图像中的人脸的尺寸却是变化的。为了尽可能多的检测到各种尺寸的人脸，我们采用图像金字塔的扫描方式。将图像在一定范围内，依次进行尺寸的缩放；将缩放之后的图像作为人脸检测器的输入。

实验中，我们在  $320 \times 240$  的原始图像上，以 1.25 因子依次进行图像缩小，共进行 8 次。经过 8 次递减后的图像尺寸为  $54 \times 40$ 。这一过程就相当于，将检测子窗体进行了从  $20 \times 20$  到  $120 \times 120$  上 8 个梯度的放大。也就是说，检测器可以在  $320 \times 240$  的图像上，满足  $20 \times 20$  到  $120 \times 120$  尺寸大小的人脸的检测。整个过程的扫描计算时间比原来只检测原始图像增加了将近 6 倍。

### 2.7.3.2 重叠检测的合并

一个人脸可能在相近的位置和尺寸上被多次检测到，这叫做重复检测。必须对重复检测的结果进行合并。在一个指定的邻域内，对检测到的人脸的数量计数，如果超过某一个阈值，那么就认为这是同一个人脸被多次检测到。具体的说就是将检测到人脸子窗口的中心点（即子窗口中心点坐标）记录下来。如果在一个邻域内这种中心的数量较多；也就是说一个人脸在相近的位置和尺度内，被多次验证；那么就说明在这个区域内最有可能存在一个真正的人脸。而单一的、孤立的中心点通常代表错误的检测。保留那些在相近的邻域内有较多数量中心点的检测，而去除那些单一的中心点。

因为在一个小邻域内出现的多个中心点同属于一个脸，所以需要将这些中心点进行归并，即将多种尺度图像的检测结果合并到一起，以最终确定单一人脸所在的位置。此过程可以排除对同一个脸的多重检测。合并的方法可以采用对多个中心点求其型心。

## 2.7.4 实验结果

在实验中，我们使用了 Intel OpenCV 计算机视觉库提供的函数进行训练和识别（关于 OpenCV 函数我们将在 2.8 节介绍），并在 AR<sup>[79]</sup>和 CMU VASC 两个人脸数据库上进行检测，实验结果如下：

表 2-2 人脸检测实验结果

Table 2-2 Experiment result of face detection

|              | AR (简单背景)    | CMU (复杂背景)   |
|--------------|--------------|--------------|
| 人脸总数         | 195          | 169          |
| 正确检测数 (检测率)  | 184 (94.36%) | 120 (71.01%) |
| 未检测出人脸 (漏检率) | 11 (5.64%)   | 49 (28.99%)  |
| 误检测人脸        | 9            | 4            |

从上表可以看出,对于背景简单和人脸大小大于  $20 \times 20$  的 AR 数据库,我们的系统可以达到较高的检测率,对于未检测出的人脸都是脸部带有遮掩物(眼镜或围巾)的图像。对于背景复杂且人脸大小不一的 CMU VASC 数据库,我们的系统的检测率虽然有所下降,但是未检测出的人脸绝大多数属于人脸大小小于  $10 \times 10$  的图像。在 CPU P4 3G/RAM 512MB 的计算机上,人脸速度可以达到 10 帧/秒,基本上达到了实时处理的要求。

## 2.8 本章小结

本章主要介绍了人脸检测中常用的方法,在 2.2 节中介绍了基于肤色特征的方法,该方法具有较低的运算复杂度,在实时系统中有较广泛的应用,但它会受到种族、光照变化和摄像白平衡的影响,影响准确度。在 2.3 节介绍了基于模板匹配的方法,其中利用固定模板检测的方法比较直观,但是固定模板对位姿、表情、尺度变化敏感,而可变形模板对非刚性模式具有较好的适应性,但可变形模板的选择和参数的确定非常困难。在 2.4 节介绍了基于特征脸的方法,该方法提取特征的运算复杂度较大,通常提取的原始特征具有很高的维数,在进行处理前要先进行降维,因而在实时系统中的应用也受到了很大的限制。在 2.5 节介绍了基于神经网络的方法,该方法便于建模,鲁棒性较好,但是运算速度较慢。在 2.6 节介绍了基于类 Haar 特征的 AdaBoost 算法,这也是本系统中人脸检测部分采用的方法,实验表明,此方法能够实时有效地检测和定位人脸。在 2.7 节介绍了人脸检测系统的实现和实验结果,实验表明,此人脸检测系统可以达到理想的准确率和实时处理的要求。

人脸检测是模式识别中极富挑战性的一类问题,尽管目前存在许多的人脸检测算法,但是每种算法都是针对不同的应用环境而得到的,还没有一种算法适用各种不同的环境。虽然人们在人脸检测研究领域取

得了显著的进展，但是一个稳定性和鲁棒性强的人脸检测系统需要在不同的环境下如光照条件、人脸方向、姿势和部分遮挡、面部表情以及外部饰物等干扰，也具有很好的有效性和稳定性。

人脸检测作为人脸识别系统的一部分，有非常广泛的应用，如可用于保安系统、案件侦破系统以及涉及到人机交互领域，另外还可广泛用于人的跟踪、监控等系统，人脸检测的结果直接影响到整个系统的性能，因此，人脸检测的准确性和稳定性是决定后续工作的关键。目前，人们日益意识到人脸检测工作的重要性，相关的研究也日益增多。

## 第三章 图像的预处理及归一化

人脸图像受到很多因素的影响，比如：光照条件、姿态、背景、面部表情以及附属物等，这些因素的变化，都会导致人脸图像的明显不同，目前还没有有效的识别算法能够很好地解决这些因素的影响，很多识别算法都是对光照条件、姿态等进行约束化简，比如对检测窗口先用模板进行处理，去掉人脸与背景交界信息<sup>[66]</sup>，仅考虑了正面人脸的识别<sup>[71]</sup>。

在上述的因素中，主要的影响来自于光照条件和姿态的变化。为了消除它们对识别效果的影响，通常的做法是扩大样本空间，收集各种光照和姿态下的样本，识别判断时考虑测试图像与各种条件下样本的差异，然后进行综合分类。另一种做法是利用先验知识，结合多种识别方法，对于光照影响，利用直方图均衡化处理图像，改进面部的亮度和对比度，本文的光线归一化处理主要采用此类方法。也可以通过建立人脸 3D 模型，估计光照的影响。对于姿态影响，可以利用弹性图匹配的方法，跟踪面部关键特征点的变化，估计姿态参数，本文采用的方法是跟踪眼睛的位置，通过眼睛定位将人脸图像归一化到一个固定的姿态。总体来说，光照和姿态变化仍是人脸识别所面临的挑战，特别是当两种因素混在一起时。

### 3.1 光照归一化

在目前众多人脸和表情识别方法中，光照和视角是影响识别效果的主要因素，同一个人的人脸图像在不同光照条件下的特征差异往往比不同的人在同一光照条件下的特征差异还要大，因此，把图像的光照影响去掉，即光照归一化问题也引起了许多研究者的关注<sup>[7,73,74]</sup>。

目前的光照归一化方法主要有：基于特征子空间的方法<sup>[7,28,75]</sup>，基于直方图修正技术的方法<sup>[76]</sup>，基于人脸光照模型的方法<sup>[73,77]</sup>等。

本节将详细介绍了几种光照归一化的方法，并对这几种方法在人脸表情识别中的应用进行比较和分析。关于基于特征子空间的方法，将在第四章中结合特征提取和特征选择作详细介绍。

#### 3.1.1 基于直方图修正技术的光照归一化方法

##### 3.1.1.1 传统的直方图均衡化算法（HE）

图像的灰度直方图表示灰度图像中具有每种灰度的像素的个数，反映了图像中每种灰度级出现的频率，是图像的基本统计特征之一。直方图均衡是利用直方图的统计数据，进行直方图的修改，以通过某种对应关系改变图像中各点灰度值，使得变换后图像的直方图相对原始图像的直方图比较均匀、平坦，灰度层次清晰，从而达到图像增强、减少光照对识别的影响的目的。

若以变量  $r$  代表图像中象素灰度级，并对  $r$  做归一化处理： $0 \leq r \leq 1$ ，其中  $r=0$  代表黑， $r=1$  代表白。对一幅给定的图像来说，每个象素取得  $[0, 1]$  区间内的灰度级是随机的。 $r$  是一个随机变量，可以用概率密度函数  $\text{Pr}(r)$  来表示原始图像的灰度分布，从图像灰度级的分布可以看出一幅图像的灰度分布特征。若图中大多数象素的灰度值取在接近于 0 的区域，则整幅图像较暗，反之，则整幅图像较亮。

理想的直方图均衡处理的目的是使处理后的图像的直方图是平直的，即各灰度级具有相同的出现频数，由此各灰度级具有均匀的概率分布，图像看起来就更清晰。因此需要找到一种变换关系  $S=T(r)$ ，使变换后新的灰度直方图较变换前直方图平直。传统的直方图均衡采用  $r$  的累积概率分布函数作为转换函数，即

$$S = T(r) = \int_0^r \text{Pr}(w) dw \quad (3-1)$$

式(1)中， $w$  是积分变量， $\int_0^r \text{Pr}(w) dw$  是  $r$  的累积概率分布函数。用  $r_k$  表示离散灰度值，用  $\text{Pr}(r_k)$  代表  $\text{Pr}(r)$ ，有下式成立：

$$\text{Pr}(r_k) = \frac{n_k}{n} \quad 0 \leq r_k \leq 1, k=0,1,2,\dots,L-1 \quad (3-2)$$

式中， $n_k$  为图像中出现这种灰度的象素数， $n$  为图像中象素总数， $L$  表示图像灰度值共分为  $L$  级。

转换函数的离散形式为：

$$S_k = T(r_k) = \sum_{j=0}^k \frac{n_j}{n} = \sum_{j=0}^k \text{Pr}(r_j) \quad 0 \leq r_k \leq 1, k=0,1,2,\dots,L-1 \quad (3-3)$$

可以证明，当图像的直方图呈均匀分布时，所包含的信息量为最大。

### 3.1.1.2 一种改进的直方图均衡化算法 (NHP)

由式(3-3)可知，对于一幅  $256 \times 100$  的拥有 256 个灰度级的图像，如果原始图像中灰度为  $k$  的象素个数少于 100，那么在转换后的图像中就会出现与临近的灰度的合并，不可能有自己的灰度值。这就造成了变换后图像的灰度范围一般很难达到图像格式所允许的最大灰度变化范围，



同时，由于灰度的“吞噬”也容易造成图像信息的丢失。特别是在原始图像直方图分布极不均匀时，如在很亮或很暗的环境下采集到的人脸图像，变换后图像层次感更差，信息丢失更严重。

Jiang Duan 和 GuoPing Qiu 提出了一种改进的直方图修改算法<sup>[76]</sup>，这种方法是建立在图像的 RGB 色彩空间上，采用高分辨率的灰度值表示方法，即用浮点数来表示原始图像像素的灰度值，从而有效地减少了灰度被吞并所造成的影响。

### 3.1.1.2.1 算法原理

采用下式作为 RGB 色彩空间灰度转换公式，其中 Lu 为浮点数类型，在我们的实验中，Lu 取小数点后两位：

$$Lu(x, y) = 0.299 \times R(x, y) + 0.587 \times G(x, y) + 0.114 \times B(x, y) \quad (3-4)$$

构造一个高分辨率的直方图：

$$H[k] = \Pr[Lu(x, y) = k] \quad (3-5)$$

基于此直方图，我们采用一种分级分割算法，将灰度级投影到动态范围为 [0, 255] 的 256 个灰度级中。

### 3.1.1.2.2 分级分割算法

首先，在 [0, 255] 中找到一个满足式 (3-6) 的  $\beta_0$  值， $0 \leq \beta_0 \leq 255$ ，

$$\sum_{k=0}^{\beta_0} H[k] = \sum_{k=\beta_0}^{255} H[k] \quad (3-6)$$

接着，我们将灰度动态范围分为两段 [0,  $C_0$ ] 和 [ $C_0$ , 255]，其中  $C_0$  由下式确定：

$$C_0 = \frac{0 + 255}{2} + \alpha(\beta_0 - \frac{0 + 255}{2}) \quad (3-7)$$

式中， $\alpha$  为算法的控制参数， $0 \leq \alpha \leq 1$ 。

对于 [0,  $C_0$ ]，我们采用相似的方法，找到一个值  $\beta_{1,0}$ ， $0 \leq \beta_{1,0} \leq C_0$ ，满足下式：

$$\sum_{k=0}^{\beta_{1,0}} H[k] = \sum_{k=\beta_{1,0}}^{C_0} H[k] \quad (3-8)$$

接着再分为两段 [0,  $C_{1,0}$ ] 和 [ $C_{1,0}$ ,  $C_0$ ]，其中

$$C_{1,0} = \frac{0 + C_0}{2} + \alpha(\beta_{1,0} - \frac{0 + C_0}{2}) \quad (3-9)$$

同理，我们可以在 [ $C_0$ , 255] 中找到  $\beta_{1,1}$  和  $C_{1,1}$ ，将 [ $C_0$ , 255] 分为两段 [ $C_0$ ,  $C_{1,1}$ ] 和 [ $C_{1,1}$ , 255]。

$$\sum_{k=C_0}^{\beta_{1,1}} H[k] = \sum_{k=\beta_{1,1}}^{255} H[k] \quad (3-10)$$

$$C_{i,0} = \frac{C_0 + 255}{2} + \alpha (\beta_{i,1} - \frac{C_0 + 255}{2}) \quad (3-11)$$

这样，我们便把灰度的动态范围分成了 4 段，继续在各段中重复上述过程，经过 8 次的分割之后，我们就可以把灰度动态范围划分为  $2^8=256$  段。落在第一段的灰度值被映射到灰度级 0，落在第二段的灰度值被映射到灰度级 1，依此类推，所有灰度值被映射到 256 个灰度级中。

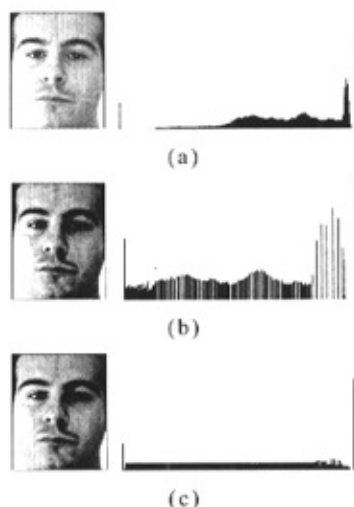


图 3—1 两种直方图均衡化算法的比较（图中左列为人脸图像，右列为左列图像的直方图）

(a)原始图像及其直方图；(b)采用传统直方图均衡化算法处理后的图像及其直方图；(c)采用改进的直方图均衡化算法处理后的图像及其直方图( $\alpha=1$ )

Figure 3—1 Comparison of two histogram equalization algorithm (left column is face image and right column is the histogram of the right one)

(a)Original Image and its Histogram; (b)Traditional Histogram Equalization and its histogram; (c)Novel Histogram Equalization and its Histogram

在本算法中，只有一个可调整的参数  $\alpha$ 。由式(3—7)、(3—9)、(3—11)可知，当  $\alpha=0$  时，原始图像的直方图完全不变地映射到新的直方图中；当  $\alpha=1$  时，映射关系相当于对图像做直方图均衡化处理；当  $0 \leq \alpha \leq 1$  时，新的灰度直方图分布的均匀程度将处于原始图像灰度分布和灰度均匀分布之间。

在实验中，我们取  $\alpha=1$ ，从图 3—1 我们可以看到，采用改进的直方图均衡算法处理后的图像的像素灰度分布更加均匀。

### 3.1.1.3 其他直方图均衡算法

#### 3.1.1.3.1 局部直方图均衡 (LHE)

直方图均衡是在整个图像的范围内对像素值的分布作变换,但是,在全局范围内取得的最佳效果在局部范围可能并不是最优的。局部直方图均衡采用了和传统全局直方图均衡同样的算法,只是改变了算法应用的图像区域,即将图像分解为若干个  $m \times n$  的窗口,对每个窗口范围内的区域作直方图均衡,使每个区域的图像都得到最佳的效果。在实验中,“我们将原始图像均分为  $m$  行  $n$  列,对  $m \times n$  个区域分别作直方图均衡,其实验结果参见第 3.1.1.4 小节。

#### 3.1.1.3.2 自适应直方图均衡 (AHE)

自适应直方图均衡算法采用计算以像素点为中心的局部相关区域的直方图,将该区域的直方图均衡的结果应用到该像素点的方法,而不是以整个图像区域的直方图为标准对像素点进行直方图修改。也就是说,对图像的每一个像素点,首先找到一个区域(如以该点为中心的  $m \times n$  的窗口),在这个区域里按第 3.1.1.1 小节介绍的直方图均衡算法计算映射函数,根据映射函数计算该像素点的新值;对整个图像的所有像素点重复以上过程,计算出均衡后的图像。

自适应直方图均衡需要对图像的每一个像素点都计算一次直方图,因此计算量是相当大的,但是,由于相邻的像素点对应的区域是相关的,我们可以利用其相关性来优化计算过程。在第 3.1.1.4 小节我们也将对比此方法的实验结果。

#### 3.1.1.3.3 局部阈值修改 (LRM)

局部阈值修改是一种线性映射的算法。这种方法是找到相关区域的像素的最大值和最小值,然后根据下式将它们扩展到我们所需要的范围。

$$Y(x,y) = \frac{C}{(\max - \min)} (X(x,y) - \min) \quad (3-12)$$

式中,  $C$  受数据存储设备或显示设备限制。

### 3.1.1.4 几种直方图修改算法的比较

图 3—2 列出了以上介绍的直方图修改算法的部分实验结果。其中(a)为原始图像;(b)为传统直方图均衡结果;(c)为采用  $20 \times 20$  窗口的自适应直方图均衡结果;(d)为采用  $51 \times 51$  窗口的局部直方图均衡结果;(e)为采用  $20 \times 20$  窗口的局部阈值修改的结果;(f)为改进的直方图均衡化算法的实验结果,参数  $\alpha = 0.8$ ;(g)为改进的直方图均衡化算法的实验结果,参数  $\alpha = 0.5$ ;(h)为改进的直方图均衡化算法的实验结果,参数  $\alpha = 0.2$ 。直方图左上角数字为图像灰度均值。

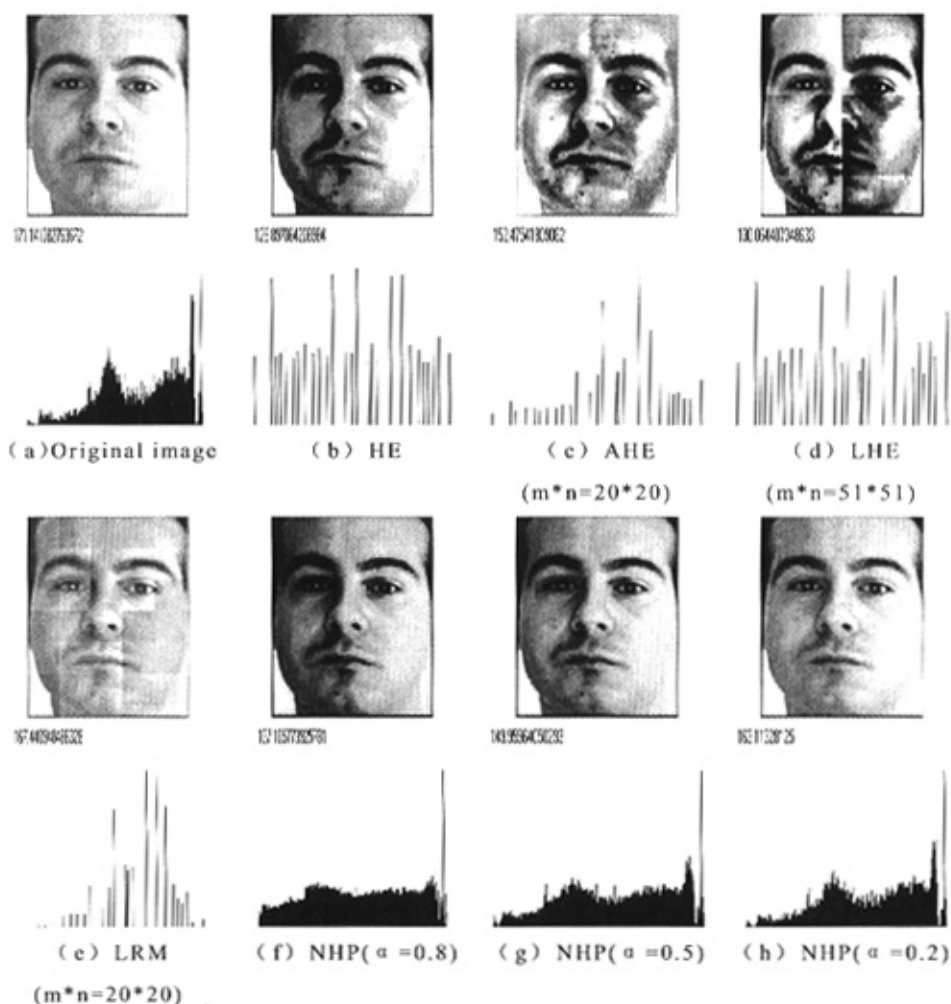


图 3—2 直方图修改算法实验结果

Figure 3—2 Experiment result of histogram modification algorithm

[注] 第 1、3 行为人脸图像，第 2、4 行为其上行图像对应的直方图

从上图可以看出，所有直方图修改算法都增强了图像的对比度。传统的直方图均衡（HE）和改进的直方图均衡化算法（NHP）是对整幅图像进行处理，局部直方图均衡（LHE）、自适应直方图均衡（AHE）和局部阈值修改（LRM）是对图像的局部区域进行处理。自适应直方图均衡处理后的图像看上去有些走样，而且其计算复杂量大，虽然通过增大窗口的大小可以减少这种走样的程度，但是其处理时间也会相应增加。局部直方图均衡和局部阈值修改处理后的图像会产生块状的效果，但是

其计算量较小，处理速度也较快。改进的直方图均衡化算法不仅保持了原始图像像素值的分布规律，而且扩展了像素值分布的范围，在归一化的同时保持了原始图像的特征，避免了结果的走样。但是，改进的直方图均衡化算法的缺点是它不能对灰度图像进行处理，而且，但把该算法扩展到 HSI 空间来处理饱和度分量时，其计算量是比较大的。

### 3.1.2 基于仿射变换光照模型的光线归一化方法

JunHua 等人提出了一种基于仿射变换的光照模型<sup>[77]</sup>，此模型假定一幅受光照影响的图像  $G'(x, y)$  是由原图像  $G(x, y)$  经过一个增益  $A(x, y)$  后和背景低通噪声分量  $B(x, y)$  叠加而成的，其模型为：

$$G'(x, y) = A(x, y) \cdot G(x, y) + B(x, y) \quad (3-13)$$

如果能够估计出低通分量  $\hat{B}(x, y)$  和增益  $\hat{A}(x, y)$ ，则由式 (3—13) 我们可以推导出

$$\hat{G}(x, y) = \frac{K}{\hat{A}(x, y)} \cdot G'(x, y) - \hat{B}(x, y) \quad (3-14)$$

在我们的实验中，首先利用一个高斯滤波器估计出加性低通分量  $B(x, y)$ ；然后我们可以将式 (3—13) 改写为：

$$G'(x, y) - B(x, y) = A(x, y) \cdot G(x, y) \quad (3-15)$$

因为乘性干扰  $A(x, y)$  一般来说被认为是在整幅图像中变化较小的，而  $G(x, y)$  在整幅图像中变化比较大，因此我们采用同态滤波器来估计  $A(x, y)$ ，利用同态滤波器保留低频分量，抑制高频分量的特点，保留  $A(x, y)$ ，抑制  $G(x, y)$ 。

设

$$f(x, y) = \ln(G'(x, y) - B(x, y)) = \ln A(x, y) + \ln G(x, y) \quad (3-16)$$

$f(x, y)$  经过高斯低通滤波后， $A(x, y)$  被分离出来。

$$LPF(f(x, y)) \approx \ln A(x, y) \quad (3-17)$$

最后通过指数变换得到  $A(x, y)$ 。

$$A(x, y) = \exp(LP F(f(x, y))) \quad (3-18)$$

在估计出  $B(x, y)$  和  $A(x, y)$  之后，我们就可以根据式 (3—14) 恢复出原图像。由于通常有一些噪声点出现在图像直方图的两端，这些点并不代表图像的重要特征，却使图像的动态显示范围缩小了，因此，我们根据图像像素值的频数设置最高灰度值和最低灰度值两个阈值，我们分别从灰度直方图的最大灰度级和最小灰度级开始，搜索第一个频数较大的

灰度级  $T_{hi}$  和  $T_{lo}$ 。然后利用式 (3—19) 将各点的灰度归一化到显示设备的显示范围内。

$$\hat{G}''(x,y) = \begin{cases} g_{lo}, & \hat{G}(x,y) \leq T_{lo} \\ g_{hi}, & \hat{G}(x,y) \geq T_{hi} \\ \frac{g_{hi} - g_{lo}}{T_{hi} - T_{lo}}(\hat{G}(x,y) - T_{lo}) + g_{lo}, & else \end{cases} \quad (3-19)$$

其中,  $g_{hi}$  和  $g_{lo}$  为显示设备灰度显示范围的上限和下限。

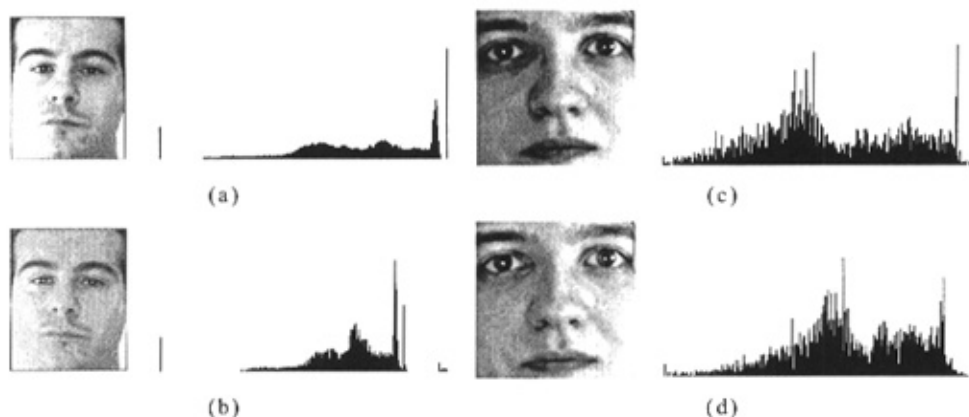


图 3—3 实验结果 (a)(c)原始图像及其直方图;

(b)(d)采用基于仿射变换光照模型的光线归一化方法处理后的图像及其直方图

Figure 3—3 Experiment Results (a)(c) Original image and its histogram;

(b)(d) Illumination normalization based on illumination affine transformation model and its histogram

[注] 第 1、3 列为人脸图像, 第 2、4 列为对应人脸图像的直方图

在实验中, 我们采用  $11 \times 11$  的高斯滤波器, 对图 3—3 (a) (c) 的图像进行处理, 其结果如图 3—3 (b) (d) 所示。可以看到, 与直方图均衡的处理结果相比较, 此方法的归一化效果更好, 但计算复杂度较高。表 3—1 列出了三种方法的计算复杂度。

表 3—1 三种光照归一化方法的计算复杂度比较

Table 3—1 Comparison of computing complexity about 3 kinds of illumination normalization methods

| 光照归一化方法     | 处理速度 <sup>[注]</sup> | 计算复杂度 |
|-------------|---------------------|-------|
| 传统的直方图均衡化   | 16ms                | 低     |
| 改进的直方图均衡化   | 16ms                | 低     |
| 基于仿射变换的光照模型 | 641ms               | 高     |

[注] 处理速度为在 Pentium4 1.8GHz CPU 上处理一幅 256×320 像素的图片所需的时间

### 3.1.4 实验结果

我们的实验采用 Gabor 滤波器提取表情特征<sup>[78]</sup>，识别器采用 k 均值距离分类器，使用 PCA 和 PCA+LDA 两种方法对特征进行降维处理，实验中的训练和测试的样本来自 AR 人脸数据库(AR Face Database)<sup>[79]</sup>，其识别率如表 3—2 所示。

表 3—2 人脸识别系统采用不同的光照归一化方法的识别率

Table 3—2 Recognition rate of different illumination normalization in face recognition system

| 降维方法        | PCA    | PCA+LDA |
|-------------|--------|---------|
| 未经处理的图像     | 91.79% | 96.48%  |
| 传统的直方图均衡化   | 92.96% | 97.87%  |
| 改进的直方图均衡化   | 93.18% | 97.97%  |
| 基于仿射变换的光照模型 | 92.43% | 96.59%  |

实验结果表明，使用光照归一化后人脸表情的识别率均有所提高，使用改进的直方图均衡化方法后的识别率也比使用传统的直方图均衡化方法的识别率要高，但是，改进的直方图均衡化要求原始图像包含 RGB 色彩空间信息，从而使其应用范围受到限制。而基于仿射变换光照模型的光照归一化方法的实时性比较差，其处理速度随着所采用的滤波器的

复杂度的增加而增加。同时，我们也注意到，由于在实验中我们对 Gabor 特征进行了降维处理，这个处理过程可能去掉了表示图像光照属性的特征向量<sup>[7]</sup>，因此，不同的光照归一化方法得出的效果也有一定的局限性。

## 3.2 几何归一化

### 3.2.1 Sobel 算子

边缘检测是图像处理中的重要内容。边缘是图像的最基本特征。所谓边缘，是指周围像素灰度有阶跃变化或屋顶变化的那些像素的集合。边缘与图像中物体的边界有关但又是不同的。图像的边缘对应着图像灰度的不连续性。显然图像的边缘很少是从一个灰度跳到另一个灰度这样的理想状况。真实图像的边缘通常都具有有限的宽度呈现出陡峭的斜坡状。

边缘的锐利程度由图像灰度的梯度决定。梯度是一个向量，指出灰度变化的最快的方向和数量。

$$\nabla f = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \quad (3-21)$$

梯度的大小和方向是由

$$\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \quad (3-22)$$

$$\theta = \tan^{-1}\left(\frac{\partial f / \partial y}{\partial f / \partial x}\right) \quad (3-23)$$

因此最简单的边缘检测算子是用图像的垂直和水平差分来逼近梯度算子：

$$\nabla f = (f(x, y) - f(x-1, y), f(x, y) - f(x, y-1)) \quad (3-24)$$

Sobel 边缘检测算子使用两个有向算子(一个水平的，一个是垂直的，般称为模板)，每一个逼近一个偏导数：

$$S_v = \begin{matrix} & -1 & -2 & -1 \\ 0 & 0 & 0 \\ & 1 & 2 & 1 \end{matrix} \quad (3-25)$$

$$S_H = \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} \quad (3-26)$$

如果我们用 Sobel 算子检测图像 M 的边缘的话，我们可以先分别用水平算子和垂直算子对图像进行卷积，得到的是两个矩阵，在不考虑边



界的情形下也是和原图像同样大小的  $M_1$ ,  $M_2$ , 他们分别表示图像  $M$  中相同位置处的两个偏导数。然后把  $M_1$ ,  $M_2$  对应位置的两个数平方后相加得到一个新的矩阵  $G$ ,  $G$  表示  $M$  中各个像素的灰度的梯度值(一个逼近)。然后就可以通过阈值处理得到边缘图像。总的过程是:

$$E = ((M \otimes S_v)^2 + (M \otimes S_h)^2) > Thresh^2 \quad (3-27)$$

这些模板是怎么来的呢?

我们假设图像的灰度满足下面这个关系:

$$M_{x,y} = \alpha x + \beta y + \gamma \quad (3-28)$$

则梯度是  $(\alpha, \beta)$ 。

显然, 当前像素  $3 \times 3$  邻域内像素值为

$$\begin{array}{ccc} -\alpha - \beta + \gamma & -\alpha + \gamma & -\alpha + \beta + \gamma \\ -\beta + \gamma & \gamma & \beta + \gamma \\ \alpha - \beta + \gamma & \alpha + \gamma & \alpha + \beta + \gamma \end{array} \quad (3-29)$$

定义垂直算子和水平算子形如:

$$\begin{bmatrix} -a & -b & -a \\ 0 & 0 & 0 \\ a & b & a \end{bmatrix} \quad (3-30)$$

$$\begin{bmatrix} -a & 0 & a \\ -b & 0 & b \\ -a & 0 & a \end{bmatrix} \quad (3-31)$$

之所以这样定义是为了满足对称性和电路设计的需要。利用这两个模板对当前像素进行卷积, 得到的方向导数为

$$g_x = 2\beta(2a+b) \quad (3-32)$$

$$g_y = 2\alpha(2a+b) \quad (3-33)$$

因此当前像素处的梯度的大小为

$$G = 2(2a+b)\sqrt{\alpha^2 + \beta^2} \quad (3-34)$$

显然要有:

$$2(2a+b)=1 \quad (3-35)$$

如果我们取  $a=1/8$ ,  $b=1/4$  则得到的就是  $1/8$  乘 Sobel 算子。

### 3.2.2 眼睛定位

因为本文要检测的目标就是眼睛, 根据对图像灰度的分析我们知道,

眼睛区域的灰度和周围像素的灰度不同，而且有较大变化，因此我们用 Sobel 边缘检测算子和区域生长的方法来检测眼睛。具体算法如下：

- (1) 在视频图像中检测出人脸（参见第二章）。
- (2) 使用 Sobel 算子检测边缘。
- (3) 将图像分割为 4 个部分，如图 3—4 所示。分别计算上半部分左右两个图像在垂直方向上的投影。

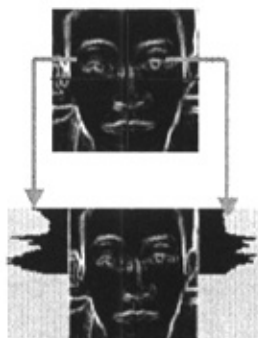


图 3—4 眼睛检测

Figure 3—4 Eyes detection

- (4) 找到投影值较大且靠近底部的点作为眼睛在垂直方向的坐标。
- (5) 从人脸图像的中间开始沿着垂直坐标所确定的方向做一条射线，当射线生长到射线上的最大灰度值和最小灰度值之差大于预定的阈值时，则射线的可变端点的坐标为眼睛的位置坐标。

### 3.2.3 仿射变换

仿射变换就是一种实现几何变换的公式，它可以理解为对坐标进行放缩，旋转，平移后取得新坐标的值。一个变换  $\omega: R^2 \rightarrow R^2$  的形式为

$$\omega \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \quad (3-36)$$

其中， $a, b, c, d, e, f$  均为实数，则称  $\omega$  为二维仿射变换，这是一种最广泛的线性变化。

### 3.2.4 人脸图像的几何归一化

本文采用的将眼睛的坐标固定在输出图像的特定坐标的方法作几何归一化。具体的作法是：

- (1) 利用第 3.2.2 节介绍的方法检测得到两只眼睛的原始坐标  $(x_1, y_1)$  和  $(x_2, y_2)$ 。
- (2) 由 (3—37) 计算在两眼连线的垂直平分线上位于眼睛下方，与两只眼睛的距离等于两眼之间的距离点的坐标  $(x_3, y_3)$ 。

$$\begin{aligned} \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\ \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \end{aligned} \quad (3-37)$$

- (3) 将  $(x_1, y_1)$ 、 $(x_2, y_2)$  和  $(x_3, y_3)$  分别映射到新图像的坐标  $(22, 44)$ 、 $(74, 44)$  和  $(48, 87)$  上，即按照第 3.2.3 节介绍的仿射变换的方法计算出仿射变换参数  $a, b, c, d, e, f$ 。
- (4) 根据上述变换参数确定式 (3—36)，计算变换后的新图像，作为几何归一化后的图像。



图 3—5 几何归一化实验结果（左图为原始图像，右图为归一化后图像）

Figure 3—5 Experiment Result of Position Normalization (The left one is original image, the right one is normalized image)

### 3.3 本章小结

本章主要叙述了实时人脸识别系统中人脸图像的预处理，该部分包括了光照归一化和几何归一化。3.1 节介绍了三种光照归一化的方法，并对其实验结果进行了对比；3.2 节介绍了利用区域生长算法进行眼睛定位和利用仿射变换进行几何归一化的方法。实验表明，这一部分的工作为系统的后续处理做出了很大的贡献。

## 第四章 特征提取和特征选择

本文采用基于 Gabor 滤波器的特征提取算法, 即将 Gabor 滤波器与图像进行卷积, 提取特征矢量---Gabor 特征。然后, 对高维空间中的 Gabor 特征通过特征选择和压缩的方法投影到低维空间中。

### 4.1 Gabor 滤波器简介

Gabor 滤波器能够最好的兼顾信号在时域和频域中的分辨能力<sup>[80, 81]</sup>, 能充分描述图象的纹理信息, 它的小波特性和说明了 Gabor 滤波结果是描述图象局部灰度分布的有力工具。因此, 图象的纹理信息可以用 Gabor 滤波来抽取, 这是 Gabor 滤波器的一个广泛运用之处。Gabor 核函数由于去掉了直流分量, 对局部光照影响不敏感。Gabor 滤波结果可以描述不同方向上灰度的分布信息, 并可以容忍图象有一定的平移、旋转、亮度不均、尺度变化等情况。

与传统的傅立叶变换相比, Gabor 变换有以下几个优点: (1) 具有良好的时频局部化特性。即非常容易地调整 Gabor 滤波器的方向、基频带宽及中心频率从而能够最好地兼顾信号在时空域和频域中的分辨能力。(2) 多分辨率特性及变焦能力。即采用多通道滤波技术, 将一组具有不同时频域特性的 Gabor 小波应用于图像变换, 每个通道都能够得到输入图像的某种局部特征, 这样可以根据需要在不同粗细粒度上分析图像。此外, 在人脸特征提取方面, Gabor 小波变换与其它几种典型方法(如光流模型)相比, 一方面其处理的数据量较少, 能满足系统的实时性要求, 另一方面, 小波变换对光照变化不敏感, 且能容忍一定程度的图像旋转和变形, 当采用基于欧氏距离进行识别时, 表情模板和待识别表情不需要严格的对应, 故能够提高系统的鲁棒性。总体而言, Gabor 变换优于其他图像处理方法, 这也是本文采用 Gabor 滤波器提取人脸特征的原因所在。

#### 4.1.1 二维 Gabor 滤波器

在空间域 Gabor 滤波器可以看作是一个被 Gaussian 函数调制的正弦平面波<sup>[82]</sup>, 二维 Gabor 滤波器的定义为:

$$\psi(x, y, \omega, \theta) = \frac{1}{2\pi\sigma^2} e^{-((x \cos \theta + y \sin \theta)^2 + (-x \sin \theta + y \cos \theta)^2)/2\sigma^2} \times [e^{i\omega(x \cos \theta + y \sin \theta)} - e^{-\omega^2\sigma^2/2}] \quad (4-1)$$

式中  $(x, y)$  是空间域像素位置坐标,  $\omega$  是径向中心频率(尺度),  $\theta$  是 Gabor 滤波器的方向,  $\sigma$  是 Gaussian 函数沿  $x$  轴和  $y$  轴的均方差, Gabor 滤波器窗函数的宽度取  $\sqrt{2}\sigma$ 。二元 Gaussian 函数横截面的一般形式是椭圆, 在这里取圆形这一特殊形式。指数函数(即正弦平面波)经过 Gaussian 函数调制之后的正弦成分具有零均值, 而余弦成分不具有零均值, 所以减去  $e^{-\omega^2\sigma^2/2}$  以使正弦平面波具有零均值。 $\sigma$  和  $\omega$  的关系可以由下式来表示:

$$\sigma = \sqrt{2\ln 2} \left( \frac{2^\Phi + 1}{2^\Phi - 1} \right) / \omega \quad (4-2)$$

$\Phi$  是倍频程的带宽, 一般取 1~1.5, 在本文中取 1。

根据以上的叙述可知 Gabor 滤波器的特性由尺度  $\omega$  和方向  $\theta$  决定, 选取一系列的尺度  $\omega$  和方向  $\theta$ , 则可以获得一组的 Gabor 滤波器, 称之为 Gabor 滤波器组。在人脸表情识别中通常选用 5 个尺度和 8 个方向<sup>[83]</sup>, 即  $\omega_m = \omega_{(\max)} \times \lambda^{-(m-1)}$ ,  $m = \{1, 2, 3, 4, 5\}$ ,  $\lambda = \sqrt{2}$ , 和  $\theta_n = (n-1)\pi/8$ ,  $n = \{1, \dots, 8\}$ 。

选取最大的尺度  $\omega_{(\max)} = \pi/2$ , 5 个尺度和 8 个方向的 Gabor 滤波器组的实部如图 4—1 所示, 从图 4—1 中可以看出其强大的局部特性和方向选择性。

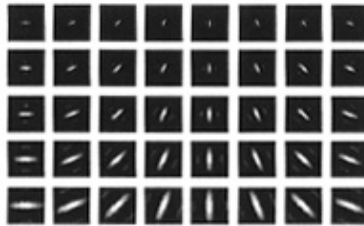


图 4—1 5 个尺度和 8 个方向的 Gabor 滤波器组的实部, 行对应不同的尺度, 列对应不同的方向

Figure 4—1 Reality of Gabor filter bank in 5 scale and 8 orientation, rows represent different scale and columns represent different orientation

#### 4.1.2 局部 Gabor 滤波器组

从图 4—2 可以看出, 对于相同方向不同频率的 Gabor 滤波器提取的特征值具有非常相似的特性, 特别是在相邻的两个频率上尤其明显, 可见其特征值具有很大的冗余性和相关性。Hongbo 等人提出了局部 Gabor

滤波器组的概念<sup>[87]</sup>。 $m$  个尺度和  $n$  个方向所构成的所有滤波器，称为全

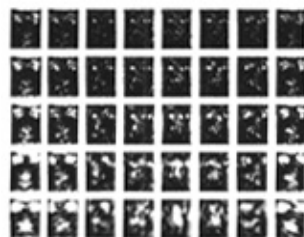


图 4—2 一幅人脸图像的 Gabor 特征表征

Figure 4—2 Gabor feature of a facial image

局滤波器组，用  $G(m \times n)$  表示。针对全局滤波器组  $G(m \times n)$  特征提取时间较长以及特征数据存在的冗余性，选用  $m$  个尺度和  $n$  个方向所构成的部分滤波器进行特征提取，称之为局部滤波器组，用  $LG(m \times n)$  表示。选取的局部滤波器组要覆盖  $n$  个方向，同一方向上只选取一个尺度或者增大尺度之间的差距，以减少特征值之间的冗余性，这样可以达到缩短特征提取时间、降低特征维数和减少存储量的目的，同时又保证了识别性能。图 4—3 给出了几种全局滤波器组和局部滤波器组实例。 $LG1(mx n)$  的选取方法是尺度  $m$  从小到大递增，到最大值后返回重新开始，对应的方向  $n$  从 1 开始一直到最大值。 $LG2(mx n)$  的不同之处在于尺度  $m$  从大到小递减，到最小值后返回重新开始。 $LG3(mx n)$  则是隔点选取尺度  $m$  和方向  $n$ 。局部滤波器组与全局滤波器组相比具有降低特征维数、缩短特征提取时间、减少计算量和存储量等优点。

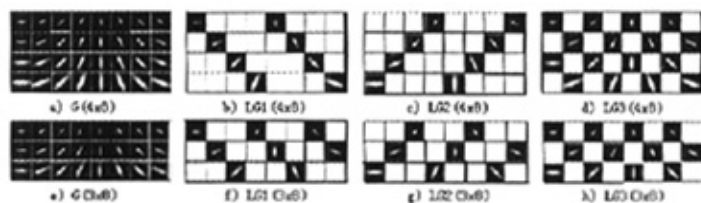


图 4—3 全局滤波器组和局部滤波器组实例，局部滤波器组  $LG(mx n)$  由显示的那些滤波器构成。

Figure 4—3 Examples of global filter bank and local filter bank, local filter bank  $LG(mx n)$  is consist of the filter displayed in the figure

## 4.2 Gabor 特征提取

图像的 Gabor 变换是原图像与一个 Gabor 滤波器组的卷积, 若  $I(x, y)$  是一幅灰度图像,  $\psi(x, y, \omega_m, \theta_n)$  为一个 Gabor 滤波器组, 则输出为:

$$O_{m,n}(x, y) = I(x, y) * \psi(x, y, \omega_m, \theta_n) \quad (4-2)$$

式中\*号是卷积运算符。图像的卷积输出为复数形式(包含实部和虚部), 复数的量值即为提取的 Gabor 特征值。对应图 4—1 的 Gabor 滤波器组, 图 4—2 为一幅人脸表情图像的 Gabor 特征值。可以看出, 在空间位置、大小和方向选择性上, 输出特征和 Gabor 滤波器组的特征是极其相似的。由于输出特征包含了不同的空间、大小和方向特征, 我们把每个 Gabor 滤波器的输出特征连接起来, 形成原始的 Gabor 特征矢量。在实际中 Gabor 特征提取时间长, 其特征维数相当高。例如, 预处理之后的图像大小为  $128 \times 96$ , 那么 40 个滤波器得到的 Gabor 特征矢量的维数一共是 491520 ( $128 \times 96 \times 5 \times 8$ ) 维。因此通常先对每个滤波器的输出特征进行采样, 然后经过归一化, 再连接起来形成新的 Gabor 特征矢量。设  $O_{m,n}^{(\rho)}$  表示采用和归一化后的  $O_{m,n}(x, y)$ , 则新的 Gabor 特征矢量为:

$$X^{(\rho)} = (O_{0,0}^{(\rho)t} \ O_{0,1}^{(\rho)t} \ \dots \ O_{m,n}^{(\rho)t})^t \quad (4-3)$$

式中,  $t$  表示转置。

## 4.3 特征选择

### 4.3.1 主分量分析 (Principle Component Analysis)

假设  $N$  个人脸图像, 每个图像用一个  $t$  维特征矢量  $x_i (i = 1, 2, \dots, N)$  表示, 主分量分析的方法是在原始的矢量空间中找到一个子空间, 该子空间基本矢量的方向为原始空间中的矢量方差最大的方向, 用  $W_{pca}$  代表将原始空间的  $t$  维矢量映射到特征子空间中的  $f$  维特征矢量的线性映射矩阵, 其中  $f < t$ 。在进行 PCA 变换之前, 我们首先对每个训练样本  $x_i$  进行归一化(即减去平均值, 然后除以方差), 转换为符合正态分布  $N(0, 1)$  的矢量  $X_i$ , 则训练样本集的总离散度矩阵表示为

$$S_t = \sum_{i=1}^N (X_i - \mu)(X_i - \mu)^T \quad (4-4)$$

式中  $\mu$  为样本的均值矢量,  $\mu = \frac{1}{N} \sum_{i=1}^N X_i$ , 矩阵  $S_t$  的计算复杂度为  $O(Nt^2)$ 。

$$\begin{aligned}
 S_t^T &= \left( \sum_{i=1}^N (X_i - \mu)(X_i - \mu)^T \right)^T = \sum_{i=1}^N ((X_i - \mu)(X_i - \mu)^T)^T \\
 &= \sum_{i=1}^N (X_i - \mu)(X_i - \mu)^T = S_t
 \end{aligned} \quad (4-5)$$

可见,  $S_t$  为对称矩阵 ( $t \times t$ ), 并且所有元素为实数。根据线性代数 Lagrange 定理, 任何一个实对称矩阵  $A$  都可以通过正交变换矩阵  $P$  化成对角形, 即:

$$P^{-1}S_tP = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_t \end{bmatrix} \quad (4-6)$$

式中,  $\lambda_i (i=1,2,\dots,t)$  为  $S_t$  的特征值, 可以证明<sup>[84]</sup>对称矩阵的特征值均非负值, 不妨令  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t \geq 0$ ; 特征值  $\lambda_i$  对应的正交归一化特征向量为  $\mu_i$ , 则  $\mu_1, \mu_2, \dots, \mu_t$  可以构成  $\mathfrak{R}^t$  空间的一组标准正交基。于是, 在新的正交空间中, 人脸样本  $X_i$  可以表示为:

$$X_i = \sum_{j=1}^t x_i(j) \mu_j \quad (4-7)$$

若选用前  $f$  个较大的特征值所对应的特征向量构造正交基 (即  $t \times f$  的线性变换矩阵  $W_{pca}$ ), 新的特征矢量可以表示为

$$Y_i = W_{pca}^T X_i \quad (i = 1, 2, \dots, N) \quad (4-8)$$

其中  $X_i$  为  $t$  维的列向量,  $W_{pca}^T$  为  $W_{pca}$  的转置,  $Y_i$  为  $f$  维的列向量。这样就可以把人脸特征数据由原来的  $t$  维降到  $f$  维 ( $f < t$ )。  $y_i$  对  $x_i$  近似的均方误差为:

$$\varepsilon = E\{Y_i - X_i\}^2 \quad (4-9)$$

主分量分析可以使均方误差最小, 即:

$$\varepsilon_{\min} = \sum_{i=f+1}^t \lambda_i \quad (4-10)$$

则人脸特征矢量的维数可由相对均方误差进行限制:

$$t = \frac{\sum_{i=f+1}^t \lambda_i}{\sum_{i=1}^t \lambda_i} \quad (4-11)$$

阈值  $t$  越大, 则  $Y_i$  的维数  $f$  越低;  $t$  越小, 则  $Y_i$  的维数越高。阈值  $t$  又称为最小能量比例系数, 当变换后某个特征值的能量在原始数据总能量所占的比例小于该系数时, 该特征值所对应的特征向量将被省略。



PCA 算法的复杂度为  $O(d^3)$ ，其中  $d$  为图像矩阵的尺寸大小。所以当  $S_t$  矩阵较大时，主分量分析的时间开销很大。从公式 4—4 可以看出，总离散度矩阵  $S_t$  的大小，由降维后的人脸 Gabor 特征矢量的维数决定。

#### 4.3.2 线性判别分析 (Linear Discriminant Analysis)

如果说主分量分析 (PCA) 是寻找最能描述原始数据的矢量空间，那么线性判别分析 (LDA) 则是寻找最能区分各类数据的矢量空间，其充分利用了训练样本的类别信息。LDA 的目标是最小化类内距离，同时最大化类间距离。在数学形式上说，类内散度矩阵  $S_w$  和类间散度矩阵  $S_b$  的定义如下：

$$\begin{aligned} S_w &= \sum_{j=1}^c \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T = H_w H_w^T \\ S_b &= \sum_{j=1}^c (\mu_j - \mu)(\mu_j - \mu)^T = H_b H_b^T \end{aligned} \quad (4-12)$$

其中  $x_i^j$  是类别  $j$  的第  $i$  个样本， $\mu_j$  是类别  $j$  的平均值， $\mu$  是所有类别的平均值， $c$  是类别个数， $N_j$  是类别  $j$  的样本个数。

要得到 LDA 的变换矩阵  $W_{lda}$ ，一种常用的方法就是最大化比率  $\det[S_b]/\det[S_w]$ 。如果  $S_w$  是非奇异矩阵，那么要使比率  $\det[S_b]/\det[S_w]$  达到最大值，只需求出矩阵  $S_w^{-1}S_b$  的特征矢量，选取最大的  $g$  个特征值对应的特征矢量作为列向量构成变换矩阵  $W_{lda}$ ，新的特征矢量可以表示为

$$y_i = W_{lda}^T x_i \quad (i = 1, 2, \dots, N) \quad (4-13)$$

其中  $x_i$  为  $t$  维的列向量， $W_{lda}$  为  $t \times g$  的变换矩阵， $W_{lda}^T$  为  $W_{lda}$  的转置， $y_i$  为  $g$  维的列向量。需要注意的是：采用 LDA 最多有  $c-1$  个非零的特征矢量，因此子空间最多是  $c-1$  维。

#### 4.3.3 线性判别分析的扩展

上节中介绍的 LDA 变换要求两个离散度矩阵中至少有一个是非奇异的，这给 LDA 的应用带来了很大的局限性。于是，一些方法被提出来解决这个问题，如伪逆 LDA<sup>[85]</sup>、规则化 LDA<sup>[86]</sup> 和 PCA+LDA<sup>[7]</sup> 等。

##### 4.3.3.1 伪逆 LDA

矩阵的奇异性问题可以用伪逆矩阵来解决。在计算矩阵  $S_w^{-1}S_b$  的特征矢量是时使用  $S_w$  的伪逆矩阵  $S_w^+$  来代替  $S_w^{-1}$ ，即通过求  $S_w^+S_b$  或  $S_b^+S_w$  的特征矢量还构造变换矩阵  $W_{lda}$ 。

伪逆矩阵可以采用奇异值分解的方法来计算。假设矩阵  $M$  的奇异值

分解为：

$$M = U\Sigma V^T \quad (4-14)$$

式中  $U$  和  $V$  为标准正交矩阵， $\Sigma$  为正对角矩阵。则  $M$  的伪逆矩阵可以用下式表示：

$$M^+ = V\Sigma^{-1}U^T \quad (4-15)$$

#### 4.3.3.2 规则化 LDA

另一个解决矩阵奇异性的方法是对  $S_w$  规则化，即将  $S_w$  加上若干个单位矩阵，使

$$S'_w = S_w + \sigma I_n, \quad \sigma > 0 \quad (4-16)$$

式中  $I_n$  为  $n$  阶单位矩阵。显然  $S_w + \sigma I_n$  是正定的，因此  $S'_w$  是非奇异的。通过用  $S'_w$  代替  $S_w$  来计算特征矢量并构造变换矩阵的方法称为规则化线性判别分析 (Regularized LDA, RLDA)。RLDA 的一个局限性是很难估计一个最优的规则化参数  $\sigma$ 。

#### 4.3.3.3 PCA 与 LDA 相结合的方法 (PCA+LDA)

为了保证  $S_w$  是非奇异矩阵，需要  $t+c$  个训练样本，当特征维数  $t$  特别大时，在实际应用中往往难以满足要求。为了解决训练样本不够的情况，在[7]中提出了中间过渡子空间，即先将  $t$  维的高维空间经过 PCA 降到  $f$  维的过渡子空间，再在此空间进行 LDA 变换，得到最终的  $g$  维子空间。更确切的说，可以表示为

$$z_i = W_{lda}^T W_{pca}^T x_i \quad (i = 1, 2, \dots, N) \quad (4-17)$$

同样，当样本的特征维数较大时，PCA 的计算复杂度比较大；而且如果过渡子空间的维数过小，原始特征中的一些有用的信息将被丢失，如果过渡子空间的维数过大，又达不到将  $S_w$  非奇异化的效果。实验表明，当 PCA 的最小能量比例为 0.1 时可以达到比较好的效果。

#### 4.3.4 基于 QR 分解的线性判别分析 (LDA/QR)

基于 QR 分解<sup>[85]</sup>的线型判别分析分为两步进行。第一步是通过 QR 分解最大化类间距离。这一步也可以作为一个独立的降维算法，我们称之为预判别 (pre-LDA/QR)，它的优点在于其较低的时间和空间复杂度。在后面的实验中我们也会将其作为一个独立的算法加以比较。第二步是在保持低空间和时间复杂度的基础上最小化类内距离。

##### 4.3.4.1 预判别算法

LDA/QR 的第一步的目的是找到一个最优的变换矩阵  $G$ ，使

$$G = \arg \max_{G^T G = I_r} \text{trace}(G^T S_b G) \quad (4-18)$$

这个最优化的  $G$  可以通过对 (4—12) 中的类间离散度矩阵  $S_b$  的先导矩阵  $H_b$  作 QR 分解得到。令  $H_b$  的 QR 分解为  $H_b = QR\Pi$ , 其中  $Q \in \mathbb{R}^{m \times t}$  为标准正交矩阵,  $R \in \mathbb{R}^{t \times k}$  为上三角矩阵,  $\Pi \in \mathbb{R}^{k \times k}$  为置换矩阵,  $t = \text{rank}(H_b)$ ; 那么对于任意正交矩阵  $W \in \mathbb{R}^{m \times t}$ ,  $G = QW$  为 (4—18) 的最优解<sup>[85]</sup>。下面作一个简单的证明。

设矩阵  $P = [Q, \hat{Q}]$  为正交矩阵, 其中  $\hat{Q} \in \mathbb{R}^{m \times (n-t)}$ , 则  $PP^T = P^T P = I_n$ ,

所以  $S_b = H_b H_b^T = (QR\Pi)(\Pi^T R^T Q^T) = QRR^T Q^T = P\Sigma P^T$

其中  $\Sigma = \begin{pmatrix} RR^T & 0 \\ 0 & 0 \end{pmatrix}$

$\therefore \text{trace}(G^T S_b G) = \text{trace}(\tilde{G}^T \Sigma \tilde{G})$ , 其中  $\tilde{G} = P^T G$ 。

$\because P$  为正交矩阵且  $G$  为标准正交矩阵

$\therefore \tilde{G}^T \tilde{G} = G^T P P^T G = G^T G = I_t$

$\therefore \text{trace}(G^T S_b G) = \text{trace}(\tilde{G}^T \Sigma \tilde{G}) \leq \text{trace}(\Sigma) = \text{trace}(RR^T)$

当  $\tilde{G} = \begin{pmatrix} W \\ 0 \end{pmatrix}$ , 其中  $W \in \mathbb{R}^{m \times t}$  为任意正交矩阵时上式等号成立。

因此,  $G = P\tilde{G} = QW$  为式 (4—18) 的最优解, 其中  $W$  为任意正交矩阵。证毕。

为了减少计算量, 在实验中我们通常选取  $W$  为单位矩阵。预判别算法的步骤如表 4—1 所示。

表 4—1 预判别算法

Table 4—1 pre-LDA/QR algorithm

|                  |  |
|------------------|--|
| 预判别算法:           |  |
| 输入: 矩阵 $A$       |  |
| 输出: 降维后的矩阵 $A^L$ |  |
| 1.               | 根据式 (4—12) 构造矩阵 $H_b$  |
| 2.               | 对 $H_b$ 作 QR 分解, 使 $H_b = QR\Pi$ , 其中 $Q \in \mathbb{R}^{m \times t}$ 为标准正交矩阵, $R \in \mathbb{R}^{t \times k}$ 为上三角矩阵, $\Pi \in \mathbb{R}^{k \times k}$ 为置换矩阵, $t = \text{rank}(H_b)$ |
| 3.               | 求最优化的矩阵 $G$  |
| 4.               | 求降维后的矩阵 $A^L = G^T A$  |

在预判别算法中, 用 QR 分解求变换矩阵  $G$  的计算复杂度为  $O(k^2 n)$ , 求降维后的矩阵  $A^L = G^T A$  的计算复杂度为  $O(knN)$ , 因此, 整个预判别算法的计算复杂度为  $O(knN)$ 。由于忽略了类内信息, 预判别算法比其他线性判

别分析算法的计算复杂度要低，效率更高。

#### 4.3.4.2 LDA/QR 算法

LDA/QR 算法的第二步是在上一步的基础上进一步解决最小化类内距离的问题，具体做法是将类内离散度信息包含到矩阵  $W$  中。在第一步中要求  $W$  为正交矩阵，在这一步，我们放宽对  $W$  的限制，即  $W$  可以为任意矩阵。我们可以找到一个变换矩阵  $G$  满足  $G=QW$ ，其中  $W$  为任意矩阵，即

$$\begin{aligned} G^T S_b G &= W^T (Q^T S_b Q) W, \\ G^T S_w G &= W^T (Q^T S_w Q) W, \end{aligned} \quad (4-19)$$

用  $\tilde{S}_b = Q^T S_b Q$  代表经过降维处理的类间离散度矩阵， $\tilde{S}_w = Q^T S_w Q$  代表经过降维处理的类内离散度矩阵，于是，计算最优化的  $G$  问题相应地转化为计算最优化的  $W$  的问题。即求

$$W = \arg \min_W \text{trace}((W^T \tilde{S}_b W)^{-1} (W^T \tilde{S}_w W)) \quad (4-20)$$

由于  $\tilde{S}_b$  是非奇异的，所以我们可以通过求  $\tilde{S}_b^{-1} \tilde{S}_w$  的特征矢量计算出最优化的  $W$ 。LDA/QR 算法的步骤如表 4—2 所示。

表 4—2 LDA/QR 算法

Table 4—2 LDA/QR algorithm

|                 |  |
|-----------------|--|
| LDA/QR 算法：      |  |
| 输入：矩阵 $A$       |  |
| 输出：降维后的矩阵 $A^L$ |  |
| 1.              | 根据式 (4—12) 构造矩阵 $H_b$ 和 $H_w$  |
| 2.              | 对 $H_b$ 作 QR 分解，使 $H_b = QR\Pi$ ，其中 $Q \in \mathfrak{R}^{m \times t}$ 为标准正交矩阵， $R \in \mathfrak{R}^{t \times t}$ 为上三角矩阵， $\Pi \in \mathfrak{R}^{t \times t}$ 为置换矩阵， $t = \text{rank}(H_b)$ |
| 3.              | $Z \leftarrow H_w^T Q$   |
| 4.              | 求经过降维处理的类间离散度矩阵 $\tilde{S}_b = RR^T$   |
| 5.              | 求经过降维处理的类内离散度矩阵 $\tilde{S}_w = Z^T Z$  |
| 6.              | 计算 $\tilde{S}_b^{-1} \tilde{S}_w$ 的特征值和特征向量，取其最大的 $t$ 个特征值对应的特征向量 $\{w_i\}_{i=1}^t$  |
| 7.              | 求最优化的矩阵 $G = QW$ ，其中 $W = [w_1, \dots, w_t]$   |
| 8.              | 求降维后的矩阵 $A^L = G^T A$  |

上表中，第 2 步的 QR 分解的计算复杂度为  $O(k^2 n)$ ；第 3 步矩阵乘法的计算复杂度为  $O(Nnk)$ ；第 4 步的计算复杂度为  $O(t^2 k)$ ，由于  $H_b$  的阶数  $t$  不大于  $k-1$ ，所以可以近似认为是  $O(k^3)$ ；第 5 步的计算复杂度为  $O(k^2 N)$ ；

第 6 步是计算一个  $k \times k$  的矩阵的特征分解，其计算复杂度为  $O(k^3)$ ；第 7 步矩阵乘法的计算复杂度为  $O(nt^2) = O(nk^2)$ ；最后第 8 步的矩阵乘法的计算复杂度为  $O(tnN) = O(knN)$ 。

由于特征维数  $n$  和样本数  $N$  通常远大于类别数  $k$ ，LDA/QR 算法的计算量主要集中在第 3 步和第 8 步。因此，LDA/QR 算法的总的计算复杂度为  $O(Nnk)$ ，这和判别算法的计算复杂度是一样的。

表 4—3 复杂度比较

Table 4—3 Complexity comparison

| 方法         | 时间复杂度     | 空间复杂度   |
|------------|-----------|---------|
| PCA        | $O(N^2n)$ | $O(nN)$ |
| PCA+LDA    | $O(N^2n)$ | $O(nN)$ |
| pre-LDA/QR | $O(nNk)$  | $O(nk)$ |
| LDA/QR     | $O(nNk)$  | $O(nk)$ |
| RLDA       | $O(N^2n)$ | $O(nN)$ |

[注]  $N$  为训练样本的数量， $n$  为样本的维数， $k$  为样本的类别数

表 4—3 列出了本文介绍的几种算法的时间复杂度和空间复杂度，可以看出，在样本数量较大的情况下，LDA/QR 体现了较为优越的性能。

## 4.4 实验结果

我们的实验采用 Gabor 滤波器提取表情特征<sup>[78]</sup>，分类器采用  $k$  均值距离分类器，实验中的训练和测试的样本来自 AR 人脸数据库 (AR Face Database)<sup>[79]</sup>，我们选用 AR 数据库中的 134 个人，每个人包含 7 幅脸部没有遮掩物的正面图像，使用各种特征选择方法的识别率如表 4—4 和表 4—5 所示。

实验结果表明，使用 PCA+LDA、RLDA 和 LDA/QR 的特征选择方法可以得到较高的识别率，其中 RLDA 需要耗费更多的计算时间，PCA+LDA 次之，LDA/QR 的计算复杂度最低，由于目前本系统的训练样本数相对较小，从 4.3 节的分析可以看出，PCA+LDA 的计算复杂度和 LDA/QR 的计算复杂度相差不大，而 PCA+LDA 的特征选择方法的识别率要比 LDA/QR 高 0.2 到 0.75 个百分点，因此，本系统采用 PCA+LDA 作为降维的特征选择方法。

表 4—4 使用全局 Gabor 滤波器时不同特征选择方法的识别率比较

Table 4—4 Comparison of different feature selection with global Gabor filter

|         | PCA    | PCA+LDA | RLDA   | pre-LDA/QR | LDA/QR |
|---------|--------|---------|--------|------------|--------|
| 原始图像    | 91.79% | 96.48%  | 96.77% | 92.11%     | 95.74% |
| 光照归一化图像 | 92.96% | 97.87%  | 97.45% | 93.39%     | 97.23% |

表 4—5 使用局部 Gabor 滤波器时不同特征选择方法的识别率比较

Table 4—5 Comparison of different feature selection with local Gabor filter

|         | PCA    | PCA+LDA | RLDA   | pre-LDA/QR | LDA/QR |
|---------|--------|---------|--------|------------|--------|
| 原始图像    | 90.09% | 95.95%  | 96.16% | 90.83%     | 95.20% |
| 光照归一化图像 | 92.75% | 97.33%  | 97.33% | 93.07%     | 96.91% |

## 4.5 本章小结

人脸图像经过预处理后，首先要通过 Gabor 滤波器进行特征提取，再对提取出的高维 Gabor 特征进行降维，即特征选择。本章主要讲述了特征提取和特征选择的过程，4.1 节首先介绍了实验采用的 Gabor 滤波器组及其特性，4.2 节则介绍了特征提取的方法，4.3 节主要叙述了几种特征选择的算法，并对它们的性能作了比较，4.4 节列出了特征提取和降维的实验结果。

## 第五章 实时人脸识别系统的整体设计和实现

前面几章分别对基于视频的人脸检测、人脸图像的预处理、特征提取和特征选择的基本原理及算法进行了详细的分析和研究。在基于视频的实时人脸识别系统中，除了以上介绍的实现功能的算法外，视频的实时捕捉也是很重要的一步，下面我们先介绍本系统中用到的 DirectX 视频捕捉技术<sup>[90]</sup>，然后介绍系统的设计和实现。

### 5.1 DirectX 视频捕捉技术

Microsoft DirectX 是一系列为创建电脑游戏和其他高性能多媒体应用程序而准备的低级应用程序编程接口 (APIs)。它包括对二维和三维图形、音效和音乐、输入设备的支持，以及对诸如多人游戏之类的网络应用程序的支持。它由许多部分组成，以我们用到的 DirectX8.1 为例，它包括 DirectX Graphics, DirectX Audio, DirectInput, DirectPlay, DirectShow 以及 Direct Setup。这里，用于图像捕捉的是 DirectShow，所以这里只对它进行介绍。

Microsoft DirectShow 应用程序界面 (API) 是 Microsoft Windows 平台的一种媒体流体系。它提供高品质的媒体流捕捉和回放并支持宽松的格式变化，包括高级流格式 (ASF)，Motion Picture Experts Group (MPEG), Audio-Video Interleaved (AVI), MPEG 声音层-3 (MP3) 和 WAV 文件。此外，它还支持使用 Windows 驱动模块 (WDM) 或老的 Windows 视屏驱动捕捉。

DirectShow 整体是以 Microsoft COM 原件完成，它可以说是数个 COM 原件组成的，使用 DirectShow 进行开发就是好比在程序中产生所需要的 COM 原件。DirectShow 是 DirectX 的一部分，要进行程序开发必须先到微软的官方网站下载 DirectX SDK，然后进行安装即可。

多媒体技术的发展给人类社会生活带来了深刻的变化，作为一门崭新的领域，多媒体技术刚开始兴起的时候的确风起云涌，发展迅速。然而，经历了早期的迅速发展期之后，多媒体技术的发展速度明显放缓。当前多媒体技术面临的如下几个重大挑战：

1. 多媒体流包含大量必须迅速处理的数据。
2. 音频、视频和任何附加流必须同步开始和结束，并以相同的速度播放。

3. 数据流可能来自许多不同的地方，包括：本地媒体文件、计算机网络、陆地广播、录像机或其他媒体设备。

4. 存在多种格式，如：音频—视频交叉存取（AVI）、高级数据流格式（ASF）、电影图片专业组（MPEG）、数码电视（DV）和电影 JPEG（MJPEG）。

5. 程序无法知道终端用户系统上的硬件设备。

针对以上这些问题，DirectShow 提出了自己的解决方案。它的主要设计目的是用来简化 Windows 平台下多媒体程序的开发，避免了数据传输、硬件差别和同步播放的复杂。

DirectShow 包括了支持多媒体捕捉和调整设备的过滤器，其中调整过滤器是基于 Windows 驱动模型（WDM）的，象过滤器一样支持 Windows 传统视频（VFW）捕捉卡和视频压缩过滤器（VCM）接口。图 5-1 显示了一个程序中，DirectShow 组件、一些 DirectShow 支持的硬件和软件组件之间的关系。

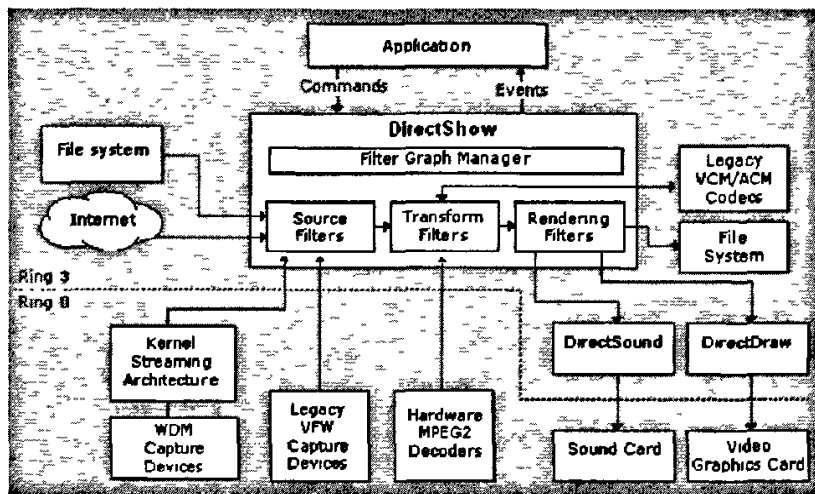


图 5-1 DirectShow 程序的构架<sup>[90]</sup>

Figure 5-1 Framework for Programs Based-on DirectShow

如图 5-1 所示，DirectShow 使程序可以播放来自不同媒介的文件和数据流，包括本地文件、本地 CD 和 DVD 驱动器、网络上的远程文件和 Windows 传统视频捕捉卡。DirectShow 自带一些文件格式的压缩器和解压缩器，而且很多第三方的硬件和软件解码器也和 DirectShow 兼容。另外，DirectShow 支持基于视频压缩管理器和音频压缩管理器接口的传统 VFW 多媒体数字信号解码器。



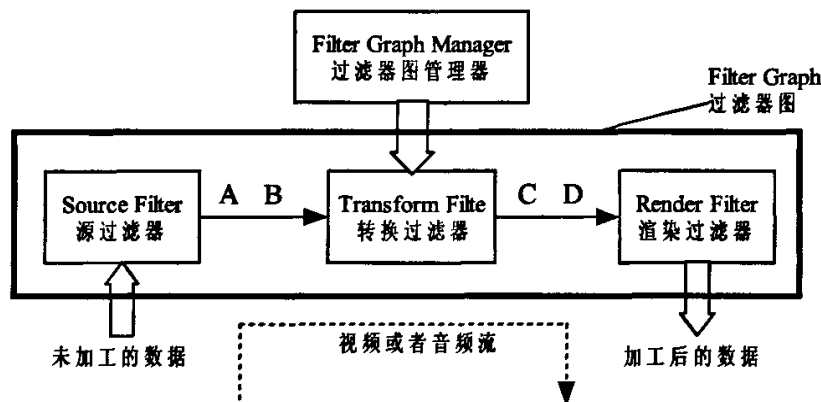


图 5-2 DirectShow 工作原理图<sup>[89]</sup>  
Figure5-2 Diagram of DirectShow

图 5-2 给出了 DirectShow 的工作原理图。其中，这里面有很多关于 DirectShow 的重要概念，这里将给予具体的说明。

■ 过滤器(Filter)

图 5-2 中大黑体方框为三个过滤器(Filter)。过滤器是 DirectShow 的基本组成部分。DirectShow 将多媒体数据的处理过程成分离的步骤，然后一个过滤器处理一个(或多个)步骤。这使应用程序可以将过滤器混合搭配从而对不同的媒体格式使用不同的硬件和软件设备执行不同类型的操作。虽然每一个过滤器都有内在完全不同的功能，但从应用程序的角度来看，它们都只是具有一定标准特征的 DirectShow 过滤器而已。这些标准特征包括：对 IBaseFilter 接口的支持和同一个或多个 DirectShow 过滤器连接的一个或多个输入/输出针脚。所有的 DirectShow 过滤器都可以大致归成以下三类：源过滤器，转换过滤器和渲染过滤器。当然，除此以外，还有一些具备其他特殊功能的过滤器，比如视频捕获程序要用到的 SampleGrabber 过滤器。

■ 过滤器图(Filter Graph)

图 5-2 中的大黑体方框就是一个过滤器图。过滤器图的定义相当简单，你可以认为一个过滤器图就是所用到的一系列过滤器的集合。当然，过滤器之间必须连接起来才是一个完整的过滤器图。

■ 过滤器图管理器(Filter Graph Manager)

图 5-2 中最上方的方框就代表着一个过滤器图管理器。可以认为过滤器图管理器是 DirectShow 中对过滤器图进行管理的 COM 组件。应用程序无需管理一个过滤器图中单个的过滤器，DirectShow

提供一个叫做过滤器图管理器的高级部件来替你完成这个任务。过滤器图管理器控制通过图形的数据流动。你的应用程序只需调用某些高级 API 函数, 比如“RUN”(移动数据以通过图形)或者“STOP”(停止数据流动), 就可以了。如果你想更直接地得到流操作的控制权的话, 你可以通过 COM 接口直接访问过滤器。过滤器图管理器会传递事件通告给应用程序, 于是你的应用程序就可以响应这些事件, 比如流的结束。此外, 过滤器图管理器简化了构造一个过滤器图的过程。比如说, 你可以指定一个文件名, 然后过滤器图管理器就会根据这个文件名来播放那个文件。

#### ■ 引脚(Pin)

图 5-2 中的过滤器都有出口或者入口, 即是图中的 A,B,C,D 处, 这一些都是 Filter 的 Pin (引脚)。如图 5-2 所示, 过滤器图中的多媒体数据从一个源过滤器开始, 经过零个或多个中间过滤器向下级传输, 最终到达渲染过滤器。引脚是处理过滤器间数据传输的底层细节。引脚是一个支持 IPinCOM 接口的 COM 对象, 它有一个方向 (输入或输出), 而且它是和图中的一个特定过滤器相关连的。引脚表示和另一个过滤器的连接点。上级过滤器中的输出引脚和下一个下级过滤器的输入引脚相连。引脚知道它们支持的媒体类型并会在两个过滤器初始连接时协商媒体类型。当媒体类型协商完成后, 引脚会进一步商讨当过滤器图运行时如何传输数据。

通过图 5-2 可以看出, 一个典型的 DirectShow 应用程序要完成三个如下的基本步骤:

- 1) 创建一个过滤器图管理器的实例。
- 2) 使用过滤器管理器图来构造一个过滤器图。
- 3) 控制过滤器图并响应事件。

## 5.2 Intel OpenCV 简介

OpenCV 是英特尔®开源计算机视觉库, 它由很多 c 函数和一部分 C++ 类构成, 实现了图像处理和计算机视觉方面的很多通用算法。OpenCV 是一个跨平台的中、高层 API 构成, 目前包括 300 多个 C 函数。它不依赖与其它的外部库, 但是也可以使用某些外部库。OpenCV 对非商业用途和商业用途都是免费 (FREE) 的。它为 Intel® Integrated Performance Primitives (IPP) 提供了透明的用户使用接口, 可以实时自动

为某些特定的处理器加载 IPP 库。

OpenCV 库致力于解决机器视觉领域中实时处理的问题。典型的应用如人机交互 (HCI)、物体识别、分割和识别、人脸识别、手势识别、运动跟踪、运动识别、运动重构等机器视觉系统。它几乎涵盖了机器视觉领域的各个方面，其内容如表 5—1 所示。

表 5—1 OpenCV 库的功能和内容

Table 5—1 Functionality and Description of OpenCV Library

| 功能       | 内容  |
|----------|---|
| 图像处理     | 图片的新建、内存分配、释放；快速的像素访问宏；                                   |
| 数据结构     | 静态结构、动态存储；  |
| 轮廓处理     | 图像轮廓的寻找、显示、处理、细化等；  |
| 几何特征     | 轮廓分析、直线和椭圆的匹配、凸起轮廓的寻找等；                                   |
| 特征算法     | 一维、二维图像的剃度算法；Canny、Hough 等线条寻找算法；轮廓的寻找与跟踪；                |
| 图像统计     | 感兴趣图像区域的计算、取中值、最小值、最大值、瞬间差值等；                             |
| 金字塔处理    | 色彩和特征的分割；   |
| 形态学      | 膨胀、腐蚀、开闭运算、顶帽、黑帽处理；                                       |
| 背景差分     | 图像累加、求方差、求平均；   |
| 距离变换     | 距离变换算法  |
| 线性处理     | 二值化、二值化反转、低通、置零等；   |
| 组件连接     | 4 领域、8 领域连接；  |
| 照相机定标    | 单精度定标、双精度定标、旋转矩阵和旋转向量间的转换、校正相机镜头变形、发现棋盘内部角点的大概位置；         |
| 形态观察     | 计算几何、极限几何等；   |
| 运动模板     | 去掉影像(silhouette) 以更新运动历史图像、计算运动历史图像的梯度方向、将整个运动分割为独立的运动部分； |
| CAMSHIFT | 寻找目标中心，计算目标尺寸和方向  |
| 动态轮廓     | Snakes 算法   |
| 光流       | HS, L-K, BM 和 L-K 变化算法；                                   |
| 预估器      | Kalman 和 Condensation 滤波器；                                |
| POSIT    | 1、2 维图像中基于 6 个自由度的模板估计                                    |
| 直方图      | 直方图基本处理、比较 2 个直方图、反向投影、EMD(Earth Mover's Distance)算法      |
| 手势识别     | 基于序列图像的手检测、计算对象的单应矩阵、计算点集的最外部                             |

|         |   |
|---------|---|
|         | 边界等；                                      |
| 矩阵运算    | 奇异值分解法 (SVD)、逆矩阵、交叉向量、马氏距离、特征值和特征向量、透视变换。 |
| 特征对象    | 计算协方差矩阵、计算特征对象、比较、协方差、分解和透视等；             |
| 隐马尔可夫变换 | 新建、释放、向量观测、DCT 变换、快速分割、训练和测试；             |
| 画图      | 在图片中画直线、矩形、圆、椭圆、多边形、文本；                   |
| 系统功能    | 读取优化代码、获得处理器信息；                           |
| 运算      | 求绝对值差、模板匹配、求方差、相位、抽样等等；                   |

OpenCV 库主要包括四个部分：（一）CXCORE，主要包括一些数据的基本结构、数组的操作、内存的动态分配、数据的保存和错误的处理等；（二）CV，图像处理和机器处理的基本算法，主要包括图像处理、结构分析、运动分析、对象跟踪部分、图像处理、结构分析、运动分析和对象跟踪部分等；（三）HighGui，主要包括基本的图形用户接口、图片的读取和存储、视频的输入输出及图片格式的转换等；（四）CVCAM，跨平台的视频捕捉模块，主要则从摄像头中读取视频流。

OpenCV 库提供了功能强大的机器视觉处理基本算法和方便的人机图形界面，使用也非常方便。OpenCV 在进行图像和视频处理时，需要用到 DirectX 的库，一般在使用时应先编译 DirectX 库和 OpenCV 库，然后把库文件和头文件加到工程中，在调用相应的库函数就可以了。

OpenCV 库中提供的基于类 Haar 特征的层叠式分类器的函数：

- LoadHaarClassifierCascade
- ReleaseHaarClassifierCascade
- CreateHidHaarClassifierCascade
- ReleaseHidHaarClassifierCascade
- HaarDetectObjects
- SetImagesForHaarClassifierCascade
- RunHaarClassifierCascade
- GetHaarClassifierCascadeScale
- GetHaarClassifierCascadeWindowSize

### 5.3 人脸识别系统的设计

本文设计并实现了一个基于视频的实时人脸识别系统，在此系统中我们采用 DirectShow 的视频捕捉技术，当捕捉到一幅图像后，首先利用第二章介绍的 Intel OpenCV 视觉库提供的函数检测人脸的位置和大小，提取人脸图像，然后对人脸图像进行预处理，包括眼睛的定位、光照归一化和几何归一化，再结合第三章介绍的特征提取和特征降维的方法提取人脸的低维 Gabor 特征，最后将特征输入分类器对该特征进行分类，由分类器输出识别结果。本系统采用的分类器为欧氏距离分类器。系统框图如下图所示：

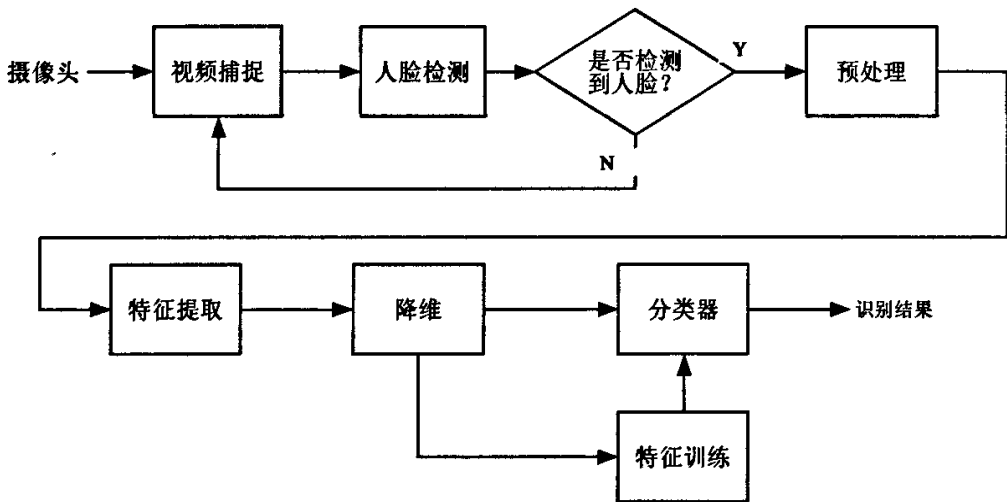


图 5—3 实时人脸识别系统框图

Figure 5—3 Framework of real-time face recognition system

在基于视频的实时人脸识别演示系统中，我们采用普通的 USB 摄像头，其采集到的图像的分辨率为  $320 \times 240$ 。为了提高人脸检测的速度，我们规定只对不小于  $20 \times 20$  的人脸进行检测。在人脸检测部分可以同时检测到图像中的多个人脸，但是为了识别的方便我们只是对检测到的第一个人脸进行识别，同时，为了提高识别的准确性，我们采用多帧识别结果投票的策略（如图 5—4 所示），取连续 5 帧的人脸图像的识别结果中分到识别结果数目最多的类别作为最终的识别结果输出。

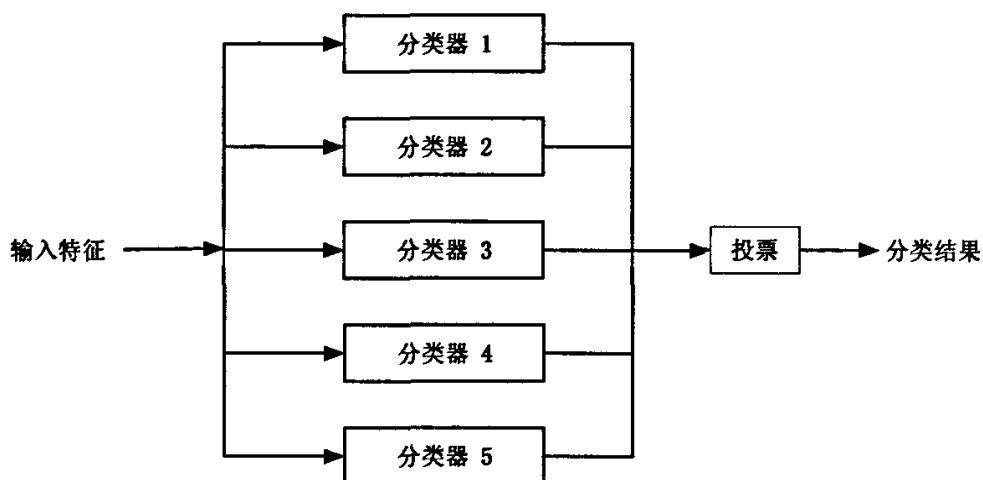


图 5—4 分类器策略

Figure 5—4 Strategy of classifier

## 5.4 人脸识别系统的实现

本系统以 Microsoft Visual C++ .NET 为开发平台完成人脸图像的采集、检测、预处理、特征提取和选择、识别全过程，由于 C++ 具有代码效率高，内存操作方便和功能强大的特点，所以能够弥补图像处理中数据量大、操作复杂、耗时长等带来的不便。结合使用 Matlab 作为特征选择训练的工具，利用 Matlab 强大的矩阵运算功能进行离线计算。

下图是实时人脸识别系统的界面：

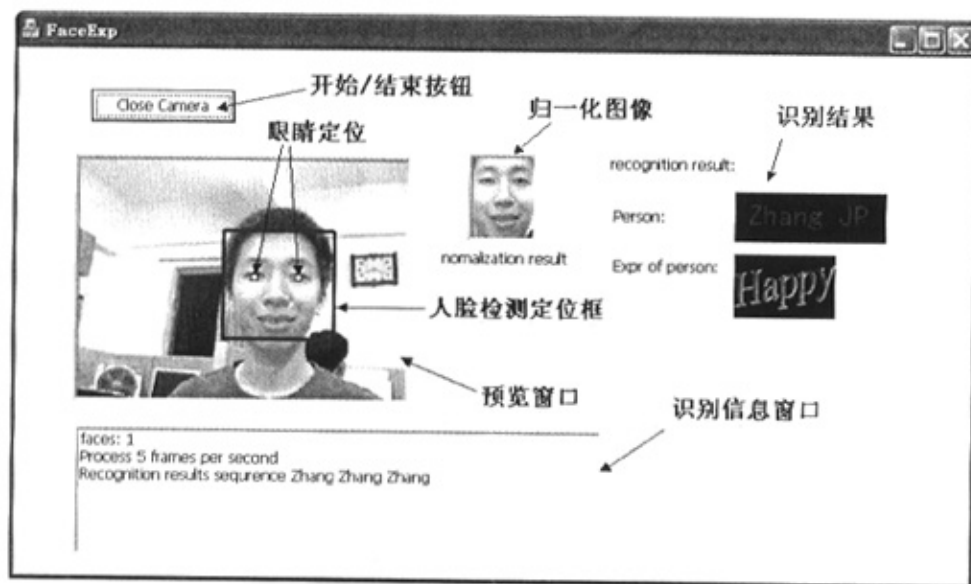


图 5—4 人脸识别系统演示程序界面

Figure 5—4 Interface of face recognition system demo

本实时人脸识别系统的建立环境为：

- 图像采集：普通 USB 摄像头
- 图像分辨率：320×240（即摄像头监视区域大小）
- CPU：P4 3.0G
- 内存：512MB
- 软件开发平台：Microsoft Visual C++ .NET

本系统的处理速度如表 5—2 所示：

表 5—2 处理速度

Table 5—2 Speed of process

| 图像捕捉  | 人脸检测   | 预处理   | 特征提取和选择  | 分类识别  |
|-------|--------|-------|----------|-------|
| <29ms | ~100ms | <29ms | 50~100ms | <10ms |

表 5—2 给出了实时人脸识别系统的平均处理速度，这里由于图像捕捉和随后的几个处理过程是同步进行且相互影响的，所以上表给出的处理速度为每个处理过程单独进行时的平均处理速度。从表中可以看出本系统整个处理过程所需要的时间在 200ms 左右，即 5 帧/秒的速度，已经基本达到实时系统的要求。

## 5.4 本章小结

本章主要讲述了如何利用前几章所介绍的技术来建立一个实时人脸识别系统。由于本系统的输入是基于 USB 的低分辨率摄像头，因此在本章首先讲述了基于 DirectX 的视频捕捉技术和 Intel OpenCV 函数库。在接下来的两节中，分别详细介绍了实时人脸识别系统的大概框架设计和系统的实现。经过实验证明，建立起来的系统每秒钟能识别 5 帧以上人脸图像，基本达到了实时系统的要求。



## 结论

本文主要研究基于视频的实时人脸识别系统的设计和实现。人脸识别系统的构成分为图像捕捉、人脸检测、图像预处理、特征提取和特征选择、以及人脸识别等部分, 本文重点讨论了人脸检测、光照归一化和特征选择等问题, 同时也介绍了其它部分的内容, 最后设计出了一个实时系统实现人脸识别的功能。本系统使用普通的 USB 摄像头, 通过预先训练好的模板可以实时识别摄像头前的人脸。

本文的主要工作如下:

- 概要介绍了人脸识别的研究背景和常用的人脸识别方法, 包括基于几何特征与模板匹配的人脸识别方法, 基于特征脸的识别方法, 基于人脸图像低频带信息的识别方法和基于神经网络与支持向量机的识别方法, 并指出它们的优缺点及使用范围。
- 介绍了几种目前应用最广泛的人脸检测方法, 包括基于肤色特征的方法, 基于模板匹配的方法, 基于特征脸的方法和基于神经网络的方法, 并且利用 OpenCV 计算机视觉库实现了一个基于类 Haar 特征的层叠式分类器, 达到了实时检测人脸的目标。
- 针对不同光照条件下的人脸识别效果比较差的问题, 详细介绍了几种不同的光照归一化方法, 并对这些方法做了大量的实验, 通过实验对比总结不同归一化方法对人脸识别的影响。
- 为了解决特征提取后特征维数较大的问题, 通过分析 PCA 和 LDA 等几种典型的特征选择算法, 通过实验对比, 选取了 PCA+LDA 作为系统的特征降维的解决方案。
- 通过研究与分析, 设计和实现了一个基于视频的实时人脸识别系统, 该系统可以完成图像捕捉、人脸检测、人脸定位、光照和几何归一化、Gabor 特征提取、降维和人脸识别的全过程, 并且达到了实时系统的要求。

在论文工作中, 不仅达到了毕业设计的预期目标, 更重要的是学到了相当多的宝贵知识, 总结起来主要有一下几点:

- 了解了一些当前图像处理和机器视觉领域的发展现状和存在的问题, 为以后的科研工作打下良好基础。
- 锻炼了自己独立研究的能力, 学会了科学研究的思维方法和良好的工作态度, 为以后的进一步深造、工作产生深远的影响。

- 通过对整个实时系统的实现，积累了一定的软件编程的知识和经验，对以后的工作大有益处。

由于受时间、学识所限，本文所设计的基于视频的实时人脸识别系统还存在着一些问题，还有很大的改进空间。主要问题在于：

- 对于视频中较小的人脸（小于  $20 \times 20$ ）检测率较低，虽然可以通过减小检测窗口大小的方法来提高识别率，但是对于系统来说其检测时间过长，从而影响了整个系统的实时性。
- 目前系统只能检测基本正面的人脸，如果是人脸偏向太大甚至是侧面，或者正面人脸在平面范围内旋转过大，人脸将无法被正确检测出来，这个问题可以通过训练侧面和旋转的人脸模板来解决。
- 本文只是提取人脸的 Gabor 特征，如果能结合其他方法提取更有效、更能表征人脸的特征，就会进一步提高识别率。
- 本文采用的光照归一化方法对于光照变化不太大的场合其效果是明显的，但是如果光照变化过大，对系统的识别率的影响是明显的，因此还需找到一种更为有效的归一化方法。
- 对于大样本的人脸数据库，特征选择的训练时间将比较长，而且加入新的样本需要对特征选择的模板重新进行训练。
- 分类器的策略也有进一步的提高空间。
- 人脸识别系统在室外环境中的应用可以进一步研究。

随着微电子、计算机、网络技术的发展，传统的磁卡、IC 卡、密码等身份鉴别方法正面临着严峻的考验。人脸识别技术的出现和日新月异的发展有望在不久的将来解决这些问题。因为，任何一个人的人脸信息特征是时刻跟随自己的，不用担心会丢失和被盗；另外，由于人脸特征相对比较复杂，仿造困难，作为安全密钥具有更高的可靠性和安全性。因此，随着人们对人脸识别技术研究的不断深入，随着计算机科学技术的快速进步，相信在不远的将来，基于人脸信息识别的信息安全技术一定会得到广泛的应用，造福于人类。

## 参考文献

- [1] Ming-Hsuan Yang, David J.Kriegman, and Narendra Ahuja, Detecting Faces in Images: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.1, Jan. 2002 Page(s):34-58.
- [2] Fei Zuo, Peter H.N de With, Real-time embedded face recognition for smart home, IEEE Transaction on Consumer Electronics, Volume 51, Issue 1, Feb. 2005 Page(s):183-190.
- [3] W. Zhao, R. Chellappa, P.J. Philips and A. Rosenfeld, Facerecognition: A literature survey, ACM Computing Surveys, Vol. 35, no.4, pp. 399-458, Dec. 2003.
- [4] T. Kanade, "Computer Recognition of Human Faces", Basel and Stuttgart: Birkhauser, 1973.
- [5] M.D. Kelly, "Visual Identification of People by Computer", Technical Report IA-130, Stanford AI Project, Stanford, CA, 1970.
- [6] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol.3, pp.72-86, 1991.
- [7] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces:: Recognition Using Class Specific Linear Projection", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol.19, pp.711-720, 1997.
- [8] L. Wiskott, J.-M. Fellous, and C.Von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 19, pp.775-779, 1997.
- [9] R. Brunelli, T. Poggio, Feature recognition: Features versus templates, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10):1042~1052, 1993.
- [10] R. J. Baron, Mechanisms of human facial recognition, Int. J. Man Machine Studies, 1981, 15:137~178.
- [11] A. L. Yuille, Deformable templates for face recognition, Cognitive Neuroscience, 3(1):59~70, 1991.
- [12] A. Yuille, D. Cohen, P. Hallian, Facial feature extraction by deformable templates. Technical report 88.2, Harvard Robotics Lab, 1988.
- [13] C. L Humg, C.W Chen, Human facial feature extraction for face interpretation and Pattern Recognition, 25(12):1435~1444, 1992.

- [14] X. Xie, R. Sudhakar, H. Zhuang, On improving eye feature extraction using deformable templates, *Pattern Recognition*, 1994, 27(6):791~799.
- [15] F. I. Parke, Computer Generated Animation of Faces, In: *Proceedings ACM AnnualConference*, 1972.1:pp.451~457.
- [16] F. Leymarie, M. D. Levine, Tracking deformable objects in the plane using an active contour model, *IEEE Transactions on PAMI*, 1993, 15(6):617~634.
- [17] D. J. Williams, M. Shah, A fast algorithm for active contour and curvature estimation, *CVGIP: Image Understanding*. 1992,55:14~26.
- [18] X. Li, N. Roeder, Face contour extraction from front-view images, *Pattern Recognition*, 28(8):1167~1179, 1995.
- [19] B. Olstad, A. H. Torp, Encoding of a priori information in active contour models, *IEEE Transactions on PAMI*, 1996, 18(9): 863~872.
- [20] G. Chow, X. Li, Towards a system for automatic facial feature detection, *Pattern Recognition*, 1993,26(12): 1739~1755.
- [21] 山世光,高文,陈熙霖. 基于面部纹理分布和变形模板的面部特征提取, *软件学报*, 2001, 12(4): 570~577.
- [22] Kim Tae Kyun, Kee Seok Cheol, Kim Sang Ryong. Real-time normalization and feature extraction of 3D face data using curvature characteristics, *Proceedings of the 10th IEEE International Workshop on Robot and Human Communication*, 2001, pp. 74~79.
- [23] M. Pardas, M. Losada, Facial parameter extraction system based on active contours, *Proceedings of the 2001 IEEE International Conference on Image Processing*, 2001, pp.058~1061.
- [24] A. Nikolaidis, I. Pitas, Facial feature extraction and pose determination, *Pattern Recognition*, vol.33, no.11, Nov., 2000, pp.1783~1791.
- [25] L. Sirovich, M. Kirby, Application of Karhunen-Loeve procedure for the characterization of human faces, *IEEE Transactions on PAMI*, 1990, 12(1): 103~108.
- [26] D. L. Swets and J. Weng, Using Discriminant Eigenfeatures for Image Retrieval, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp.891~896, Aug. 1996.
- [27] H. Peng, D. Zhang. Dual eigenspace method for human face recognition, *Electronics Letters*, vol. 33, no. 4, Feb.13, 1997, pp.283~284.
- [28] P. Belhumeur, J. Hespanha, and D. Kriegman, Eigenface vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.19 no.7, pp.711~720, 1997.

- [29] M. H. Yang, N. Ahuja, D. Kriegman, Face recognition using kernel eigenfaces, Proceedings of the 2000 IEEE International Conference on Image Processing, Vancouver, BC, vol. 1, pp.37~40, Sep.10~13,2000.
- [30] A. Pujol, J. Vitria, F. Lumberras, J. J. Villanueva, Topological principal component analysis for face encoding and recognition, Pattern Recognition Letters, vol.22, no. 6~7, pp.769~776, May, 2001.
- [31] Shakunaga Takeshi, Shigenari Kazuma. Decomposed eigenfaces various lighting conditions. Proceedings of the 2001 IEEE Comp for face recognition under outer Society Conference on Computer Vision and Patern Recognition, Kauai, II, vol. 1, pp. 864~871, Dec.8~14,2001.
- [32] 张辉,周洪祥,何振亚. 基于主元分析神经网络的人脸特征提取及识别研究,模式识别与人工智能, 1996.9(1):52~57.
- [33] 彭辉等. 基于 K-L 变换的人脸自动识别方法,清华大学学报(自然科学版),1997, 37(3):67~70.
- [34] V. Brennan and J. Principe, Face classification using PCA and multi-resolution, Proceedings of the IEEE Workshop on Neural Networks in Signal Processing, 1998,506~515.
- [35] L. D. Harmon, The recognition of faces, Science of America, 1973, 229:71-82.
- [36] A. P. Ginsburg, Visual information processing based on spatial filters constrained by biological data, AMRL Technical Report, 1978, pp.78~129.
- [37] J. Sergent, Microgenesis of face perception, in H. D. Ellis, M. A. Jeeves, F. Newcombe, A.Young (Eds.), Aspects of Face Processing, Nijhoff, Dordrecht, 1986.
- [38] C. Nastar, B. Moghaddam, A. Pentland, Flexible images: matching and recognition using learned deformations, Computer Vision and Image Understanding, 1997, 65 (2):171~191.
- [39]C. Nastar, N. Ayache, Frequency-based non-rigid motion analysis, IEEE Trans. Pattern Anal. Mach. Intel., 1996, 18(11):1067~1079.
- [40]S. T. Bow, Pattern Recognition and Image Preprocessing, Marcel Dekker, Inc., 1992, pp. 203~274.
- [41]B. Li, Y. H. Liu. When eigenfaces are combined with wavelets, Knowledge-based Systems, 2002, 15:343~347.
- [42]J. H. Lai, P. C. Yuen, and G. C. Feng, Face recognition using holistic Fourier invariant features,Pattern Recognition, 2001,34(1):95~109.
- [43] G. W. Cottrell, M. Fleming, Categorization of faces using unsupervised feature extraction, Proc. Intl Neural Networks Conf, Paris, 1990:65~70.

- [44] N. Intrator, D. Reisfeld, Y. Yeshum, Face recognition using a hybrid supervised /unsupervised neural network, Pattern Recognition Letters, vol. 17, no. 1, 1996, pp. 67~76.
- [45] S. Lawrence, et al, Face recognition: A convolutional neural-network approach, IEEE Transactions on Neural Networks, vol. 8, no. 1, 1997, pp. 98~113.
- [46] Y. Dai, Y. Nakano, Recognition of facial images with low resolution using a Hopfield memory model, Pattern Recognition, vol.31, no.2, Feb, 1998, pp.159~167.
- [47] C. Comes and V. Vapnik, Support vector networks, Machine Learning, 1995, 20:1~25.
- [48] 边肇祺,张学工等. 模式识别(第二版),北京:清华大学出版社,2000.
- [49] P. C. Yuen, J. H. Lai, Face representation using independent component analysis, Pattern Recognition, 2002, 35(6):1247~1257.
- [50] K. Lee. SVM-based face verification with feature set of small size, Electronics Letters, vol. 38, no. 15, 2002, pp.787~789.
- [51] G. Guo, S. Z. Li and K. Chan, Face Recognition by Support Vector Machines, Proceedings of the 4th IEEE Intern. Conf. on Automatic Face and Gesture Recognition, 2000, 196~201
- [52] K. Jonsson, J. Kittler, Y. P. U., J. Matas, Support vector machines for face authentication, Imaging and Vision Computing, 2002, 20:369~375.
- [53] M. J. Lyons, Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Dec, 1999, 21(12):1357~1362.
- [54] Z. M. Hafed, M. D. Levine, Face recognition using the discrete cosine transform, International Journal of Computer Vision, 2001, 43(3):167~188.
- [55] F. Samaria, Face Recognition Using Hidden Markov Models, PhD thesis, University of Cambridge, 1994.
- [56] R. Swiniarski, Data mining methods in face recognition, Proceedings of SPIE, Applications of Artificial Neural Networks in Image Processing V, San Jose, CA, USA, vol. 3962, pp.52~59, Jan. 27~28, 2000.
- [57] 梁路宏, 艾海舟, 徐光, 张钹, 人脸检测研究综述, 计算机学报, 2002, 25 (5): 449-458.
- [58] J.-C. Terrillon, M.N. Shirazi, H. Fukamachi, S. Akamatsu, Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, Automatic Fourth IEEE International Conference on Face and Gesture Recognition, 28-30 March 2000

Page(s):54-61.

[59] C. Garcia, G. Zikos, G. Tziritas, Face detection in color images using wavelet packet analysis, IEEE International Conference on Multimedia Computing and System, Volume 1, 7-11 June 1999 Page(s):703-708.

[60] M. Abdel-Mottaleb, A. Elgammal, Face detection in complex environments from color images, International Conference on Image Processing, Volume 3, 24-28 Oct. 1999 Page(s): 622-626.

[61] J. Karlekar, U.B. Desai, Finding faces in color images using wavelet transform, International Conference on Image Analysis and Processing, 27-29 Sept. 1999 Page(s): 1085-1088.

[62] B.D. Zarit, B.J. Super, F.K.H. Quek, Comparison of five color models in skin pixel classification, International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 26-27 Sept. 1999 Page(s): 58-63.

[63] V. Govindaraju, Locating human faces in photographs, International Journal of Computer Vision, vol. 19, no.2, pp. 129-146, 1996.

[64] Miao J, Yin BC, Wang KQ, Shen LS, Chen XC. A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. Pattern Recognition, 1999,32(7):1237-1248.

[65] A.L. Yuille, D.S. Cohen, P.W. Hallinan, Feature extraction from faces using deformable templates, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 4-8 June 1989 Page(s):104-109.

[66] K-K. Sung, T. Poggio, Example-based learning for view-based human face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, Jan. 1998 Page(s):39-51.

[67] 李士进, 杨静宇, 陆建峰, 基于奇异值特征和隐马尔可夫模型的人脸检测, 中国图象图形学报, 2001, 6 (7): 681-688.

[68] G. Dai, C. Zhou, Face recognition using support vector machines with the robust feature, IEEE International Workshop on Robot and Human Interactive Communication, Oct.31-Nov.2 2003 Page(s):49-53.

[69] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 10, No. 1, Jan. 1998 Page(s): 23-38.

[70] C.P. Papageorgiou, M. Oren, T. Poggio, A general framework for object detection, Sixth International Conference on Computer Vision, 4-7 Jan. 1998 Page(s):555-562.

- [71] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Volume 1, 2001 Page(s):I-511 - I-518.
- [72] M. Kearns, L.G. Valiant, Cryptographic Limitations on Learning Boolean Formulae and Finite Automata, Journal of ACM, Vol.41, No. 1, Jan. 1994 Page(s): 433-444.
- [73] Haitao Wang, Stan Z.Li, Yangsheng Wang, Weiwei Zhang, "Illumination Modeling and Normalization for Face Recognition", IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Page(s):104 – 111, 17 Oct. 2003.
- [74] Shiguang Shan, Wen Gao, Bo Cao, Debin Zhao, "Illumination Normalization for Robust Face Recognition Against Varying Lighting Conditions", IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Page(s):157 – 164, 17 Oct. 2003.
- [75] Yangrong Ling, Xiangrong Yin, Bhandarkar, S.M., "Sirface vs. Fisherface: recognition using class specific linear projection", International Conference on Image Processing, Volume 3, Page(s):III - 885-8, 14-17 Sept. 2003.
- [76] Jiang Duan, Guoping Qiu, "Novel Histogram Processing for Colour Image Enhancement", Third International Conference on Image and Graphics, Page(s):55 – 58, 18-20 Dec. 2004.
- [77] Juhua Zhu, Bede Liu, Stuart C.Schwartz, "General Illumination Correction and its Application to Face Normalization", IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume 3, Page(s):III - 133-6, 6-10 April 2003
- [78] Lin-Lin Huang, A. Shimizu, H.Kobatake, "Classification-Based Face Detection Using Gabor Filter Features", Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Page(s):397 – 402, 17-19 May 2004.
- [79] A.M. Martinez and R. Benavente. "The AR Face Database". CVC Technical Report #24, June 1998
- [80] 叶敬福, 詹永照, 基于 Gabor 小波变换的人脸表情特征提取, 计算机工程, 2005, 31 (15): 172-174.
- [81] P. Kruizinga, N. Petkov and S.E. Grigorescu, Comparison of texture features based on Gabor filters, Proceedings of the 10<sup>th</sup> International Conference on Image Analysis and Processing, 27-29, Sept. 1999 pp. 142-147.
- [82] T.S. Lee, Image representation using 2D Gabor wavelets, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 10, Oct. 1996 Page(s):959-971.



- [83] C. Liu, H. Wechsles, Independent Component Analysis of Gabor Features for Face Recognition, IEEE transactions on Neural Networks, Vol. 14, No. 4, July 2003 Page(s):919-928.
- [84] 罗家洪, 矩阵分析引论, 广州: 华南理工大学出版社, 2000: 133-135.
- [85] J. Ye, Q. Li, A two-stage linear discriminant analysis via QR-decomposition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 6, June 2005 Page(s):929-941.
- [86] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Regularized D-LDA for face recognition, IEEE International Conference on Acoustics, Speech, and Signal Processing Vol. 3, 6-10 April 2003 Page(s):III-125-8.
- [87] Hongbo Deng, Lianwen Jin, Li-Xin Zhen, Jian-Cheng Huang, A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA, to appear in International Journal of Information Technology, 2005.
- [88] Y. Freund, R.E. Schapire, A Decision-theoretic Generalization of On-line Learning and An Application to Boosting, Second European Conference on Computational Learning Theory, March, 1995.
- [89] 徐睿, 基于低分辨率摄像头的手指输入人机交互系统, 华南理工大学硕士学位论文, 2004.
- [90] 郭宇聪, 基于 VFW 和 DirectShow 的视频捕捉, 华南理工大学学士学位论文, 2002.

## 攻读硕士学位期间发表的论文

| 序号 | 作者（全体作者，按顺序排列） | 题    目                    | 发表或投稿刊物名称、级别 | 发表的卷期、年月、页码                                 | 相当于学位论文的哪一部分（章、节） | 被索引收录情况 |
|----|----------------|---------------------------|--------------|---|-------------------|---------|
| 1  | 张鞠鹏、金连文        | 几种光照归一化方法的比较研究及其在表情识别中的应用 | 信号处理         | 第 21 卷<br>第 4A 期<br>2005 年 8 月<br>257-260 页 | 3.1 节             |         |

## 致谢

首先，我要感谢我的导师金连文教授，本论文的工作得到了金老师的悉心指导。在我就读硕士研究生的三年时间里，金老师渊博的知识、严谨的治学态度和勇于开拓的创新精神以及对我的积极鼓励和无微不至的关怀，给我留下了深刻的印象。金老师在向我传授专业知识的同时，也向我传授了开展科研工作的方法和做人的道理，这将让我终身受益。在完成我的硕士论文工作时，每当遇到问题，他总是能给我很好的启发，使我能顺利的解决问题，他的教导使我受益匪浅，因此在这里我向金老师表达衷心的感谢。

此外，感谢华南理工大学电子与信息学院各位领导和老师在我攻读硕士学位期间给予的关怀和帮助，特别要感谢罗莉芳老师和马丽红老师的帮助。同时感谢黄建成博士，镇立新博士以及 Motorola 中国研究院对本项目工作的支持。

感谢实验室的徐睿、沈东升、龙钧宇、洪穗、陈彩虹、杨端端、陈彬、武锐、倪天龙、陈东明等师兄和师姐对我工作的支持和帮助；感谢邓洪波、龙腾、冯万仁、罗旺亮、庄崇彪、韩洪欣、熊波等同学，和他们一起度过的三年研究生岁月将是我一生中最美好的回忆；感谢杨军、王晓辉、罗鑫平、彭毅、戴雪峰、俎小娜、张彬、于建锋、郭锐、何智杰、纪培金、朱星华等师弟和师妹。感谢 JinGroup 的所有成员，组成了了一个团结、友爱、上进的集体，使我的研究生生活丰富多彩，深有意义。还要感谢电信 03 硕的所有同学，特别是从本科以来的同学余康桦，和他们学习和相处的日子是我的一大财富。还有很多朋友们一直对我的关心和支持，在此一并表示感谢。

最后，我要感谢远在家里的父母二十多年来对我的培养，他们一直是我战胜各种困难的精神支柱和坚强后盾，我前进的每一步都凝聚着他们的心血和期望。同时，也感谢所有关心和帮助过我的亲人和朋友，并借此机会向他们献上我最深切的挚爱和祝福。

张鞠鹏

2005-5-13