

103



摘 要

人脸表情含有丰富的人体行为信息,是人们进行非语言交流的一种重要方式。人脸表情识别在人机交互、智能机器人、安全驾驶、临床医学等领域有着非常广泛的应用前景。本论文主要研究内容如下:

1. 阐述了人脸表情识别的研究背景和意义,介绍了表情识别的国内外发展现状,并从特征提取和分类两方面对人脸表情识别领域的若干关键技术方法做了总结和概括。

2. 本文提出基于 Adaboost 和多掩模的人眼精确定位算法,该算法先通过 Adaboost 算法定位人眼大概区域,再利用矩形掩模和椭圆形掩模进一步缩小眼睛所在区域,最后计算像素点的分布中心得到眼睛坐标。

3. 针对局部二元模式(Local Binary Pattern, LBP)算法的三大不足,本文采用中心化二元模式(Center-based Binary Pattern, CBP)算法提取人脸表情特征。该方法与 LBP 相比,不但降低了特征向量的维数,而且将中心像素点的作用考虑在内,同时改写了符号函数提高了抗噪性。实验表明, CBP 算法在提取人脸表情特征方面有较快的速度和较强的抗噪性。

4. 在表情分类阶段,本文尝试分别采用中心最近邻分类器(Center Nearest Neighbor, CNN)和二对二多分类支持向量机方法(SVM)对表情进行分类,并将二对二多分类 SVM 方法、CNN 和其它方法进行比较实验,实验表明二对二多分类 SVM 方法不但识别速度快,而且有着较好的识别性能。

关键词: 表情识别 人眼定位 中心化二元模式 支持向量机

Abstract

Facial expression recognition is widely used in the fields of Man-Machine Interaction, intelligent robot, safe driving and clinical medicine. The main research contents in this paper are as follows:

1. The research background and the current research situation of the facial expression recognition are illustrated. The main technologies of facial expression feature extraction and classification in the facial expression recognition are generally described.

2. A new method of eye location is proposed. It detects a roughly eye area with Adaboost in the first. Then it reduces the area with rectangular mask and elliptical mask. At last eye locations are obtained by computing the distribution centre of pixels.

3. Considering the three deficiencies of local binary pattern (LBP), Center-based binary pattern (CBP) is adopted at the stage of feature extraction. Compared with LBP, CBP, which considers the center pixel, reduces the feature dimensions and changes the sign function. Experiment results show that the method has a fast speed and a good ability to overcome the noise.

4. Center nearest neighbor (CNN) classifier and two-versus-two multi-class SVM classifier are adopted to classify the expressions. Compared with the CNN and the other methods, the result of experimentation shows that two-versus-two multi-class SVM classifier has a fast classification speed and a high performance.

Keywords: Facial Expression Recognition Eye location

Center Binary Pattern

Support Vector Machine

目 录

第一章 绪论	1
1.1 研究背景与意义	1
1.2 人脸表情识别概述	2
1.2.1 国内外人脸表情识别的现状	3
1.2.2 人脸表情特征提取的方法	4
1.2.3 人脸表情特征分类的方法	6
1.2.4 人脸表情识别的难点	7
1.3 常用人脸表情数据库	8
1.4 本文研究内容及论文结构	9
1.4.1 本文研究内容	9
1.4.2 论文结构	10
第二章 人眼定位	11
2.1 Adaboost 算法	11
2.1.1 Haar-like 特征	11
2.1.2 Adaboost 算法	13
2.1.3 Adaboost 级联分类器	15
2.2 人眼定位	15
2.2.1 Sobel 算子	16
2.2.2 基于 Adaboost 算法和积分投影的人眼精确定位	16
2.2.3 基于 Adaboost 算法和多掩模的人眼精确定位	18
2.2.4 实验结果及分析	21
2.3 人脸图像的预处理	24
2.3.1 几何预处理	24
2.3.2 光照预处理	25
2.4 小结	26
第三章 基于中心化二元模式的人脸表情特征提取	27
3.1 局部二元模式(LBP)	27
3.2 中心化二元模式	29
3.2.1 LBP 算子的不足	29
3.2.2 中心化二元模式	30
3.3 实验结果及分析	32
3.3.1 实验参数选择	32
3.3.2 实验结果及分析	33
3.4 小结	34

第四章 基于中心最近邻和 SVM 的人脸表情特征分类..... 35

4.1 中心最近邻分类器..... 35

4.1.1 常用的距离度量函数 35

4.1.2 中心最近邻分类器 36

4.2 基于二对二支持向量机的表情分类..... 37

4.2.1 多分类问题的描述 37

4.2.2 多类别分类问题的解决方法 37

4.2.3 基于二对二支持向量机的表情分类 41

4.3 实验结果及分析..... 43

4.4 小结..... 44

第五章 总结与展望 45

5.1 总结..... 45

5.2 展望..... 45

致谢 47

参考文献 49

第一章 绪论

1.1 研究背景与意义

人脸表情在人们交流中起着十分重要的作用,是人们进行非语言交流的一种重要方式。人脸表情含有丰富的行为信息,是情感最主要的载体,对它的研究可以进一步了解人类的心理状态。通过表情,人们不仅能够准确地表达自己的思想感情,而且还可以辨别他人的内心活动。

随着科学的进步,计算机、机器人在一定程度上已经能够模仿人类行为,但是仍不具有像人一样的情感活动。早在 1997 年 Picard 指出情感计算(Affective Computing)是未来人工智能理论与应用的重要组成部分,其目的就是通过赋予计算机识别、理解、表达和适应人的感情能力,来建立和谐的人机环境,并使计算机具有更高、更全面的智能^[1]。人脸表情识别研究是实现智能人机交互(Human Computer Interface, HCI)的重要基础,对于实现人机交互以及图像理解有重大意义。随着计算机技术在图像分析和模式识别领域的进步,表情识别相关理论方法和实用性技术得到了迅速发展,在人机交互、智能机器人、安全驾驶、临床医学、心理学研究、人脸图像合成与动画、智能控制、虚拟现实等多个领域有着非常广泛的应用前景。

1. 智能人机交互

机器人走进人类日常生活,将人类从繁重、危险和单调的日常劳动中解放出来是人们追求的梦想。目前的机器人主要用于大型车间环境下从事重复和高精度的操作,难以适应人类的日常生活环境和任务要求。因此,研究人机交互是实现机器人生活化、大众化的关键,让计算机通过捕捉人脸表情变化,识别并理解人类情感,并针对人的情感做出不同的回应,就可以使机器人更好地服务于人类。由此可见,人脸表情识别是人机交互中非常重要的一个部分。

2. 智能监控

对于涉及重大安全的工作岗位,如火车调度员、长途汽车司机、客机机长等,如果工作人员疲劳工作,一旦发生险情后果将不堪设想。2008 年 2 月,美国发生罕见客机“飞过站”事故,夏威夷一架内陆客机竟然飞过目的地后才折返降落。美国联邦航空局调查之后发现,在飞机飞过机场上空时,正、副机长均在睡觉。所幸,这次事故没有酿成惨剧。无独有偶,2010 年 5 月印度航空公司一架从迪拜起飞的波音 737 客机在卡纳塔克邦的门格洛尔机场降落时坠毁,166 名乘客和机组人员中只有 8 人生还,成为印度十多年来死伤最严重的一起空难。事后工作人员调查事故原因,发现失事飞机的黑匣子中记录着飞机着陆前机长“响亮”的鼾声和粗重

的呼吸声。可见，表情识别非常适合这些安全领域的智能监控，通过监视工作人员的面部表情，一旦出现打瞌睡、疲劳等征兆，表情识别系统将会自动地发出报警，避免险情的发生。

3. 医疗、教育领域与日常监护

医疗领域中，在医护不在场的情况下，表情识别系统一旦发现病人面部出现痛苦的表情时立刻向医护人员发出报警，及时预防病人险情的发生。教育领域中，可以在课堂上根据学生的眨眼情况判断学生注意力的集中程度以及对课程的接受程度。在家庭、幼儿园和老人院的日常监护中，表情识别系统通过监控儿童、老人或需监护人员面部表情变化，及时发现其情绪和身体状况的变化，避免不必要事故的发生。

4. 娱乐

虚拟现实游戏中，如果游戏中的角色可以根据游戏者的不同面部表情做出不同的反应，对玩家而言这样的游戏肯定比传统设定好规则的游戏更有吸引力。在动画片制作方面，通过人脸表情识别合成技术为角色赋予丰富的表情变化，比采用手工方式设计更能提高效率。随着游戏产业和其他数字娱乐产业的飞速发展，表情识别具有相当大的应用价值。

总之，人脸表情分析技术是人机智能交互的重要基础，该技术有着广泛的应用前景。

1.2 人脸表情识别概述

1978 年 Suwa 等^[2]人对图像序列中的人脸表情进行分析和研究，将表情识别引入机器视觉领域。Fasel^[3]认为表情识别是基于视觉信息将脸部的运动或脸部特征的变形进行分类，因此可以说表情识别是一个模式分类问题。分类问题有三个基本要素：分类的类别、分类的特征和分类的方法。表情的类别一般为 4 到 8 类^[4]，分类特征和分类方法因出发点的不同而有差异^{[5][6]}。总之，表情识别的过程是将输入的表情图片通过某种分类方法将其归为某个类别或给出某种属性描述的过程。

人脸表情识别系统一般包含三个部分：人脸检测、人脸表情特征的提取和人脸表情的分类，如图 1.1 所示。

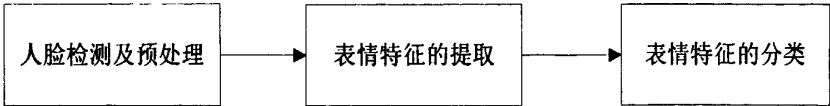


图 1.1 人脸表情识别系统

传统的人机交互，主要通过键盘、鼠标、屏幕等方式进行，只追求便利和准确，无法理解和适应人的情绪或心情。由于人类之间的沟通与交流是自然而富有感情的，因此，在人机交互的过程中，人们也很自然地希望计算机具有情感能力。

情感计算就是要赋予计算机类似于人一样的观察、理解和生成各种情感特征的能力,最终使计算机像人一样能进行自然、亲切和生动的交流。情感计算是一个新的研究领域,1997年 R. Picard^[1]给出了情感计算的定义,即情感计算是关于情感、情感产生以及影响情感方面的计算。情感计算技术通过传感器获取由人的情感所引起的表情及生理变化信号,利用“情感模型”识别这些信号,理解人的情感并做出适当响应。自此,情感计算研究才不断被学术界重视。

1.2.1 国内外人脸表情识别的现状

近年来,随着智能人机交互与情感计算技术的飞速发展,人脸表情识别已成为人们研究的热点,吸引了来自国内外不同领域学者和研究机构的关注。

在国外,美国、日本、英国等发达国家和印度、新加坡等发展中国家都有专门的科研机构对人脸表情进行研究。从事这方面研究的主要有美国麻省理工学院(MIT),卡内基梅隆大学(CMU),美国伊利诺斯大学(UIUD),瑞士的 IDIAP 研究院,日本的名古屋大学(Nagoya),东京大学(Tokyo)和大阪大学(OSAKA)等众多研究机构。由日本早稻田大学及日本机器人制造商 tmsuk 共同开发的 KOBIAN 仿人类表情机器人能表现出七种不同的表情。

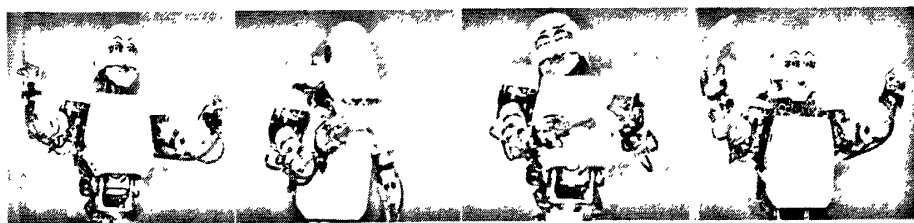


图 1.2 KOBIAN 仿人类表情机器人

在国内,情感计算相关领域的研究和应用正方兴未艾,国家自然科学基金委员会也将其列入发展的重点项目,得到了国家的大力支持。

2005年10月,首届国际情感计算与智能感知交互(International Conference on Affective Computing and Intelligent Interaction, ACII)会议在北京举行,由中国科学院自动化研究所、中国图像图形学会、国家自然科学基金委员会等知名科研机构主办,云集了国内外著名学者与优秀研究人员,促进相关科研人员在此领域的交流与合作,推动了国内在该领域研究的发展。以后该会议每两年举办一次,第四次国际 ACII 会议将于 2011 年 10 月在美国举行。

东南大学学习科学研究中心的主要研究方向是儿童面部表情与情感语音信息分析、情绪与认知、虚拟现实与脑科学可视化。2004 年该校研发了由多个摄像头和处理相关数据的计算机组成的表情识别系统,这个系统主要应用于儿童脑科学的研究。数据库中存储高兴、愤怒、悲伤、吃惊、厌恶、恐惧等 6 种表情有关数据和图片,测试者在该系统中做出任何一个表情,计算机将根据该表情的特

征在数据库中进行检索,核对数据库有关数据后确认是何种表情。

2005 年哈尔滨工业大学机器人工程技术中心研发出国内首台具有表情的机器人,这台机器人被命名为“百智星”,是利用语音识别、神经网络控制和人类仿生学设计完成的。该机器人不但具有丰富的面部表情,而且能模仿主人的声音唱歌、陪孩子说话和玩耍。

中国科学院自动化研究所设计的福娃机器人,具有语音识别功能,头和手臂都能活动,可以眨眼、转动眼睛、张嘴说话,并且能做出部分面部表情动作。其语音识别智能对话功能,能够听懂汉语、英语等应用广泛的大语种,并能同人进行交流。

清华大学、浙江大学、北京大学、华中科技大学等科研机构也在做相关研究。

智能人机交互与情感计算作为一门跨学科的交叉学科,在目前依旧缺乏智能化理论体系,尚处于探索阶段,但在国内外研究人员的不懈努力下已取得了很多成绩,推动了人脸表情识别技术的飞速发展。

1.2.2 人脸表情特征提取的方法

人脸表情特征提取方法主要分为两大类:一是基于形变特征,二是基于运动特征^{[7][8][9]}。形变特征是指在静态图像或在视频序列的一帧中提取的,相对于中性表情不同的表情形变特征;而运动特征是指在视频序列中提取的,当表情发生时,脸部产生的连续变化的特征。

1. 基于形变的表情特征提取

脸部表情特征的形变包括形状形变和纹理形变。基于形变的表情特征提取又可分为基于图像和基于模型两类方法。其中,基于图像的方法有主成分分析(Principle Component Analysis, PCA)、局部二元模式(Local Binary Pattern, LBP)、Gabor 小波变换等;基于模型的方法有点分布模型(Point Distribution Model, PDM)、活动外观模型(Appearance Active Model, AAM)等。

(1)主成分分析法

PCA 也称主分量分析,用降维的思想,把多指标转化为少数几个综合指标。主成分分析法是一种数学变换的方法,它把给定的一组相关变量通过线性变换转换成另一组不相关的变量,这些新的变量按照方差依次递减的顺序排列。在数学变换中保持变量的总方差不变,使第一变量具有最大的方差,称为第一主成分,第二变量的方差次大,并且和第一变量不相关,称为第二主成分。依次类推, N 个变量就有 N 个主成分。

在表情识别应用时,将高维的人脸图像经 K-L 变换得到一组新的正交基,通过分析人脸训练样本集的统计特性保留部分正交基,得到低维人脸空间,将待识

别表情图片投影到新的低维空间进行分类,取得了较好的识别效果^[10]。但是 PCA 方法在样本类间离散度增大时,样本类内离散度也在增大,而且,只考虑到图像数据中的两阶统计信息,没有利用高阶统计信息,许多研究者提出了改进方法,如二维主成分分析算法^[11]、基于独立分量分析(Independent Component Analysis, ICA)算法^[12]。

(2)局部二元模式法

局部二元模式(LBP)是一种有效的纹理描述算法,可以对灰度图像中局部邻近区域的纹理信息进行度量和提取。近几年 LBP 算子已经广泛应用于纹理分类、图像、人脸图像分析等领域^[13]。基于 LBP 的方法具有计算简便、抗光线干扰、判别力强等优点,正引起人们的高度关注。与 Gabor 小波变换法相比, LBP 算法能够更快的从人脸图像中提取出特征,而且维数小很多,同时保留了有效的人脸信息^[14]。施兴华^[15]结合经典的线性判别分析法,提出了 LDA+LBP 方法,该方法具有判别能力的同时,能有效地提取纹理特征。付晓峰等^[16]针对传统 LBP 算法的不足,对 LBP 算法进行改进,提出了中心化二元模式(CBP)。

(3)Gabor 小波变换法

Gabor 小波核函数具有与人类大脑皮层简单细胞的二维反射区相同的特性,同时 Gabor 滤波器在提取图像特征时能够捕捉图像中不同的空间频率、空间位置以及方向的局部特性信息^[17],通常采用多个不同尺度和通道的 Gabor 滤波器组成滤波器组,并根据图像的特点选择参数,从而使不同方向和尺度的 Gabor 滤波器覆盖整个频域,因此人脸表情图像和一组 Gabor 小波进行卷积的结果就是该人脸表情图像的 Gabor 特征。但是采用多尺度多方向的 Gabor 滤波器进行特征提取后,特征向量的维数很高,运算量大,识别一幅图片的时间较长。孙兴华、徐红侠等^[18]采用局部 Gabor 二进制模式直方图序列来降低特征向量的维数,降低运算量。

(4)点分布模型

点分布模型(PDM)是 Cootes、Taylor^[19]提出的一种参数化的形状描述模型,它是一套形状模型,代表了平均几何形状和来自一组训练形状的几何变量的统计模型。PDM 在计算机视觉中的统计外观学习和医学图像分割中已经成为一个标准,形状在这些方面能有效地帮助对噪声和低分辨率图像的理解。PDM 用一组离散的控制点来描述对象的形状,并用 PCA 方法建立起各个控制点的运动模型,对控制点的基准位置和移动模式做了约束,从而保证模型的整体形变始终在可接受范围内。此模型引入了与目标特征有关的信息,从而降低了对噪声和形变的影响,但通常都对图像质量有较高要求,由于脸部特征附近对比度比较低,边界搜索收敛慢,因此运算量一般都比较大大。

2. 基于运动的表情特征提取方法有:光流法、特征点跟踪法和差分图像法等。

(1)光流法^[20]

光流是空间运动物体在观测成像面上的像素运动的瞬时速度。光流的研究是利用图像序列中的像素强度数据的时域变化和相关性来确定各自像素位置的“运动”，即研究图像灰度在时间上的变化与景象中物体结构及其运动的关系。面部表情的运动信息可以通过光流估计的运动场来分析。由于传统光流法是基于灰度一致性假设前提下得到的稠密光流场，当光源不稳定或运动物体存在形变时，光流计算将不准确。人脸运动是典型的非刚体运动，采用传统光流法必然会导致光流场计算的不精确，从而影响面部表情的识别。为此 Wildes 等^{[21][22]}基于流体运动的连续方程提出了一种适合计算流体运动光流场的扩展光流算法，该算法对光源变化、运动物体的变形具有较好的鲁棒性。

(2)特征点跟踪法

特征点跟踪法是指利用灰度变化较大的眼角点、嘴角点等进行跟踪，并进一步得到脸部特征的位移或形变信息。选择脸部的显著特征点来进行运动估计，特征点一般都选取在脸部永久特征上。Tian 等^[23]人提出了一种不同的元素模型，其中嘴巴、眼睛、眉毛和颧骨被用作特征点跟踪，根据隐藏的面部特征的变形来获得模型的轮廓。近年来产生了一些改进算法，如 Fisher 脸方法等。特征点跟踪法只考虑用感兴趣的特征点来分析表情，而不考虑背景等无关信息，减少了运算量，但同时由于只标记了部分特征点，可能会丢失有用信息。

(3)差分图像法

差分图像法是指将视频序列中相邻的两帧做差分运算，即将对应像素相减，等于零的点代表“不变”区域，非零点代表“变化”区域。当该方法用于表情分析时，通常对含有表情的人脸图像序列与相同背景下中性表情的人脸图像进行差分运算，来检测图像序列中处于极大状态即表情信息最为丰富所对应的关键帧。在后继的表情识别中只需对检测到的关键帧进行处理，这样降低了运算量，减少了运算时间。Ma 等^[24]人将含表情的人脸图像和中性人脸图像相减，对差分图像做离散余弦变换，取低频部分作为特征向量对表情进行识别。

人脸表情十分丰富，所传递的信息复杂性很高，各种特征提取方法都有不同的适应环境和各自的优缺点，使用单一的特征方法精确地提取出表情信息是不现实的，因此利用数据融合理论，各种方法综合使用，扬长避短，是以后研究的趋势。

1.2.3 人脸表情特征分类的方法

表情分类是整个表情识别系统的最后一个环节，即设计一个分类器，将待识别表情特征归于某一类中。常用的分类器设计方法有以下几类。

1. 最近邻分类器

Cover 和 Hart^[25]于 1968 年提出最近邻分类器,该分类器是模式识别非参数法中最重要的方法之一。最近邻分类器是根据待识别样本与模板之间的距离作为判断依据。距离的测度主要有 L_1 范数、 L_2 范数、马氏距离和 Hausdorff 距离等。与隐马尔可夫、支持向量机等复杂分类器相比,其通用性好。但和某些针对特定问题而设计的分类器相比,其识别率要差。

2. 人工神经网络分类器

人工神经网络(Artificial Neural Networks, ANNs)也简称为神经网络(NNs)或称连接模型(Connectionist Model),它是一种模仿动物神经网络行为特征,进行分布式并行信息处理的算法数学模型。人工神经网络具有自学习和自适应的能力,可以通过预先提供的一批相互对应的输入—输出数据,分析掌握两者之间潜在的规律,最终根据这些规律,用新的输入数据来推算输出结果。用于表情识别的人工神经网络分类器主要有:多层感知器(Multilayer Perception, MLP),BP(Back Propagation)网络和径向基函数神经网络(Radical Basis Function, RBF)。用于人工神经网络方法进行表情识别的缺点是,当输入的训练样本集较大时,会出现过拟合等问题。当识别自然表情而不只是 6 种基本表情时,对分类器的训练会比较繁重。

3. 支持向量机分类算法

支持向量机(Support Vector Machine, SVM)是 Corinna Cortes 等^[26]于 1995 年首先提出的,它在解决小样本、非线性及高维模式识别中表现出许多特有的优势,并能够推广应用到函数拟合等其他机器学习问题中。支持向量机方法根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷,以获得最好的推广能力。由于其在解决小样本、非线性样本问题上的优势,该方法在表情识别中得到了广泛应用^{[27][28][29]}。

4. 隐马尔可夫模型分类算法

隐马尔可夫模型(Hidden Markov Model, HMM)是 Baum 等^{[30][31][32]}人最早在 1966 年提出的,该模型是一个统计模型,是马尔可夫链的一种,在 HMM 中观察到的事件与状态通过一组概率分布相联系,是一个双重随机过程。Sakaguchi 等^[33]首先将 HMM 用于人脸表情识别。杨国亮和于仲安^[34]提出基于改进光流算法和 HMM 的面部表情识别,该算法取得了较好的识别效果。

1.2.4 人脸表情识别的难点

近年来,人脸表情识别的研究发展很快,取得了不少成绩。但是,要让计算机能自动、实时和准确识别人脸表情还有相当大的距离。目前的研究还处于起步阶段,依然存在许多亟待解决的问题。

1. 人脸是非刚性的,不同脸型、不同年龄和不同性别的人的表情很难用一个

通用的模型来描述。

2. 人类表情十分丰富, 表情的变化细微而复杂。美国心理学家 Ekman 可以人为制造 10000 种表情, 大约 3000 种有实际意义。而现阶段研究的仅有 4 到 8 种基本表情, 而且这几种表情是夸张化了的表情, 学者对细微变化的表情研究现阶段还比较少。

3. 光照变化、多姿态、遮挡物等因素对表情识别算法的鲁棒性影响很大, 现有的图像处理技术的局限性给特征提取造成一定的困难。

4. 表情识别涉及图像处理、模式识别、生命科学和心理学等多个学科, 大部分学科刚刚起步, 理论和方法还不成熟, 需要进一步发展。

1.3 常用人脸表情数据库

本文对常用的人脸表情数据库做了一个统计, 有的可以直接从网上下载, 有些需要申请才能得到。

1. Yale 人脸数据库, 包含 15 人共 165 张灰度图, 每个人多姿态且表情各异。

2. BioID 人脸数据库, 该库包含了 1521 张, 大小为 384×286 像素的灰度图像。该数据库中人脸图像多姿态、多光照, 如图 1.3 所示。而且该数据库给出了数据库中所有人脸图像的眼睛坐标, 所以该数据库也被用来分析人眼精确定位算法的性能。本文对人眼精确定位性能的分析也是采用该数据库。



图 1.3 BioID 人脸数据库

3. FERET Face Database 为美国 National Institute of Standards and technology(NIST)中的 Information Technology(ITL)实验室所提供。该数据库有着相当丰富的人脸影像, 是目前所有资料库中最多的一个, 在此资料库中, 参与拍摄的人数共 1208 人, 每人拍摄条件及张数各不相同, 包含正面、侧面姿态、光照、表情、室内外及时间跨度变化。

4. 日本 Kyushu 大学的 JAFFE(Japanese Female Facial Expression)数据库, 是研究亚洲人表情的重要测试库之一。该库包含 213 张女性人脸表情图像, 共计 10 人, 每个人各有 7 种表情(包括中性表情), 如图 1.4 所示。表情是静态表情, 图像为 256 级灰度图像, 图像采集时对光照和头部位置都未加控制。JAFFE 库中已经给出了量化的日本女性对库中每幅图像属于某种表情的语意性描述。



图 1.4 JAFFE 六种基本人脸表情

5. Cohn-Kanade 数据库是基于 AU 编码的数据库, 含有 210 个对象, 大约 2000 个图像序列, 每个人有一系列的脸部活动, 包括单个 AU 或 AU 组合, 如图 1.5 所示。



图 1.5 Cohn-Kanade 六种基本人脸表情

6. MIT CBCL 人脸数据库集, 其中训练集包括 6977 张不同尺寸规格的图像, 2429 张人脸图像, 4548 张非人脸图像; 而测试集有 24045 张图像, 472 张人脸图像, 23573 张非人脸图像。

7. CAS-PEAL 人脸图像数据库中包含了 1040 名中国人共 99450 幅头肩部图像。所有图像在专门的采集环境中采集, 涵盖了姿态、表情、饰物和光照 4 种主要变化条件, 部分人脸图像具有背景、距离和时间跨度的变化。

1.4 本文研究内容及论文结构

1.4.1 本文研究内容

人脸表情识别是实现智能人机交互的前提和关键技术之一, 对于实现自然的人机交互以及自动的图像理解具有重大意义。本文对人脸表情识别中部分关键技术进行了研究, 主要研究内容有:

1. 针对人眼定位这一问题, 本文在分析了传统的基于灰度投影定位眼睛方法的缺点后, 提出了基于 Adaboost 和多掩模的人眼精确定位方法。实验表明, 该算法定位眼睛的效果优于传统的基于灰度投影法。

2. 分析了基于局部二元模式(LBP)的人脸表情特征提取方法, 针对 LBP 的不足, 本文采用了中心二元模式(CBP)算子提取人脸表情特征, 该算子有以下三大优势:(1)CBP 算子产生的直方图维数远小于传统 LBP 算子产生的直方图维数。(2)CBP 算子将中心像素点考虑在内, 通过比较“紧邻点对”之间的差异, 更好的捕捉到

梯度信息,提高了其鉴别力。(3)CBP 对白噪声反应不敏感。实验结果表明, CBP 在降维、辨别力和抗白噪声方面都优于 LBP。

3. 为了有效地将表情分类, 本文对 CNN 和二对二多分类 SVM 算法进行了深入研究, 并把 CNN、二对二多分类 SVM 算法和其它分类方法作对比实验, 实验表明, 二对二多分类 SVM 算法有较快的分类速度和较好的分类性能。

1.4.2 论文结构

本文共分五章, 各章内容安排如下:

第一章 绪论。简要介绍了课题的研究背景和意义, 综述了国内外的研究现状以及人脸表情识别中特征提取和分类的主要技术方法, 指出了研究的难点, 介绍了常用的人脸表情数据库。

第二章 人眼定位。该章首先介绍了 Adaboost 算法的基本原理, 接着提出基于 Adaboost 算子和多掩模的人眼精确定位方法, 并对此方法做了实验分析, 最后介绍了表情图像几何预处理和光照预处理的方法。

第三章 基于中心化二元模式的人脸表情特征提取。首先介绍了 LBP 算法的基本原理, 针对 LBP 算法的不足采用了 CBP 算法提取人脸表情特征, 实验证明 CBP 在提取人脸表情纹理特征方面比 LBP 更加有效, 并且有较强的抗噪性。

第四章 基于中心最近邻和 SVM 的人脸表情识别。本章对表情多分类问题展开研究, 首先介绍了 CNN 和常用的 SVM 分类方法, 接着分别尝试采用 CNN 分类器和二对二多分类 SVM 分类器对人脸表情特征进行分类, 并对实验结果进行了分析。

第五章 总结和展望。对本文工作进行总结, 并对下一步工作做出展望。

第二章 人眼定位

人眼定位是人脸识别和表情识别预处理的关键步骤, 是人脸旋转角度和尺度归一化的前提。人脸面部表情是通过眼睛、眉毛、嘴巴和脸颊肌肉变化来表现情绪状态。而人的眼神变化是面部表情最主要的体现, 其次才是嘴角和眉头肌肉的变化。眼睛在人脸中的位置比较固定, 两眼间距受光照和表情变化的影响最小, 双眼中心连线的方向随人脸的偏转而偏转, 可以作为图像旋转校正的依据。本章主要研究了人眼定位算法以及人脸表情图片的预处理方法。

2.1 Adaboost 算法

Viola 和 Jones^[35]于 2001 年引入积分图像(Integral Image)概念, 提出了基于 Haar-like 特征、级联结构的 AdaBoost 算法, 该方法利用积分图像快速计算图像的矩形特征, 然后通过 AdaBoost 算法选择少量的关键特征并构造分类器。在检测时, 将多个分类器组成一个级联分类器, 快速排除非人脸区域。该方法具有速度快、检测率高等优点, 使人脸检测从真正意义上走向了实用。

2.1.1 Haar-like 特征

1. Harr 函数^[36]

Harr 函数是一种定义在连续闭区间 $[0,1]$ 上的正交归一化函数, 该函数计算简单, 收敛均匀而且速度快, 在图像信息压缩和特征编码方面得到广泛应用。

Haar 函数 $h_k(z)$ 其中 $k=0,1,2,\dots,N-1$, 其中 $N=2^n$ 。整数 k 可被唯一地分解成:

$$k = 2^p + q - 1 \quad (2-1)$$

其中 $0 \leq q-1 \leq n-1$, 当 $p=0$ 时, $q=0$ 或 $q=1$; 而当 $p \neq 0$ 时, $1 \leq q \leq 2^p$ 。

Haar 函数定义为:

$$h_0(z) = h_{00}(z) = 1/\sqrt{N} \quad \text{当 } 0 \leq z \leq 1 \quad (2-2)$$

$$h_k(z) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2} & \frac{q-1}{2^p} \leq z \leq \frac{q-1/2}{2^p} \\ -2^{p/2} & \frac{q-1/2}{2^p} \leq z \leq \frac{q}{2^p} \\ 0 & \text{其它} \end{cases} \quad (2-3)$$

2. Haar 变换

一维 Haar 变换, 一维矢量 $v(x)$ 的 Haar 变换 $z(x)$ 定义为

$$z(x) = H \cdot v(x) \quad (2-4)$$

式中 H 为 Haar 变换系数矩阵。

将一维 Haar 变换推广到二维

$$F(u,v)=\frac{1}{N}H\cdot f(x,y)\cdot H^T \tag{2-5}$$

式中 $f(x,y)$ 为空间域数据矩阵， H 为 Haar 变化系数矩阵。

3. Haar-like 特征

Haar-like 特征也叫矩形特征，最初的 Haar-like 特征是由 Papageorgiou 等^{[37][38]}人提出。Haar 特征分为三类：边缘特征、线性特征、中心特征和对角线特征。特征模板内有白色和黑色两种矩形，并定义该模板的特征值为白色矩形像素和减去黑色矩形像素和。一般来说，人脸图像中的单个像素点值仅能反映该像素点的亮度或灰度信息，不能提供任何类别信息，只有对这些像素点值采用某种特征综合策略才可得到利于检测的信息。Haar-like 特征是指图像中两个或多个形状大小相同的矩形内所有像素灰度值之和的差值，是一种适合表征图像中代表不同区域变化的特征，利用 Haar-like 特征可以分辨人脸不同区域的变化信息及各器官之间的空间分布关系，所以 Haar-like 特征值能有效地区别人脸与非人脸区域。图 2.1 为 Viola 等^[35]人采用的 Haar-like 矩形模板。



图 2.1 Haar-like 矩形特征

4. Haar-like 特征快速计算

Viola 等^[35]人将积分图应用到特征值的计算中。积分图的引入可以只对图像进行一次遍历计算，每个特征值计算所需时间为常量，这使得计算和检测的速度大大提升。

对于图像内任意一点 (x,y) ，其积分图像值就是在 (x,y) 左上角所有像素之和，即：

$$ii(x,y)=\sum_{x'\leq x,y'\leq y}i(x',y') \tag{2-6}$$

式中 $ii(x,y)$ 表示原图像在 (x,y) 处的积分图像值， $i(x',y')$ 为点 (x',y') 处的原始图像灰度值。

$ii(x,y)$ 也可以通过下式迭代求出：

$$\begin{aligned} s(x,y) &= s(x,y-1)+i(x,y) \\ ii(x,y) &= ii(x-1,y)+s(x,y) \end{aligned} \tag{2-7}$$

其中 $s(x,y)$ 为点 (x,y) 所在纵坐标不超过该点所有像素灰度之和，称为“列积分和”，记为：

$$s(x, y) = \sum_{y' \leq y} i(x, y') \quad (2-8)$$

通过积分图像, 任何矩形的像素灰度累加可以通过 4 个点积分图像值的加减计算获得。如图 2.2 所示: 点 1 的积分图的值是矩形 A 内的像素值的和, 点 2 的积分图的值是矩形 A+B 内的像素值的和, 点 3 的积分图的值是矩形 A+C 内的像素值的和, 点 4 的积分图是矩形 A+B+C+D 的像素值的和, 所以矩形 D 的积分图像值为 $4+1-(2+3)$ 。

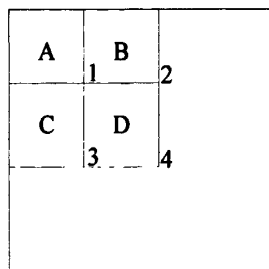


图 2.2 积分图示例

2.1.2 Adaboost 算法

Adaboost(Adaptive Boosting)算法^[39]是具有自适应的 boosting 算法, 是对 boosting 的改进。Adaboost 是一种迭代算法, 其核心思想是针对同一个训练集训练不同的分类器(弱分类器), 然后把这些弱分类器集合起来, 构成一个更强的最终分类器(强分类器)。理论上, 弱分类器的分类能力只要比随机猜测要好, 当弱分类器的个数趋于无穷时, 强分类器的错误率将趋于零。

1. 弱分类器

弱分类器的分类精度稍大于 50%, 不期望有很好的分类效果。弱分类器构造方法如下:

每个特征 f_j 所对应一个弱分类器 $h_j(x)$, 那么

$$h_j(x) = \begin{cases} 1 & p_j f_j(x) < p_j \theta_j \\ 0 & \text{其它} \end{cases} \quad (2-9)$$

其中, $p_j = \pm 1$ 表示不等式的方向, θ_j 是在训练样本上最小错误分类的阈值。 θ_j 可以通过下式得到:

$$\theta_j = \frac{N_1 \overline{m_1} + N_2 \overline{m_2}}{N_1 + N_2} \quad (2-10)$$

其中, N_1 为人脸样本个数, N_2 为非人脸样本个数。 $\overline{m_1}$ 为当前特征在所有人脸样本上的平均特征值, $\overline{m_2}$ 为当前特征在所有非人脸样本上的平均特征值。

若 $\overline{m_1} < \theta_j$, 则 $p_j = -1$; 此时当 $f_j(x) < \theta_j$ 时, $h_j(x) = 1$ 弱分类器将该图判断为人脸图像, 否则判断为非人脸图像。若 $\overline{m_1} > \theta_j$, 则 $p_j = +1$; 此时当 $f_j(x) > \theta_j$ 时, $h_j(x) = 1$ 弱分类器将该图判断为人脸图像, 否则为非人脸。

2. 强分类器

为了获得强分类器, 需要用 Adaboost 算法将多个弱分类器组合成一个强分类器。Adaboost 算法中不同的训练集是通过调整每个样本对应的权重实现的。最开始的时候, 每个样本对应的权重是相同的, 在此样本分布下训练出一个基本分类器 $h_1(x)$ 。对于 $h_1(x)$ 错分的样本, 则增加其对应样本的权重; 而对于正确分类的样本, 则降低其权重。这样可以使得错分的样本突出出来, 并得到一个新的样本分布。同时, 根据错分的情况赋予 $h_1(x)$ 一个权重, 表示该基本分类器的重要程度, 错分得越少权重越大。在新的样本分布下, 再次对基本分类器进行训练, 得到基本分类器 $h_2(x)$ 及其权重。依次类推, 经过 T 次这样的循环, 就得到了 T 个基本分类器, 以及 T 个对应的权重。最后把这 T 个基本分类器按一定权重累加起来, 就得到了最终所期望的强分类器。其算法如下:

(1) 给定 n 个样本 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), \dots, (x_n, y_n)$, 其中 x_i 是输入样本图像, $y_i = 1$ 表示正样本, $y_i = 0$ 表示负样本。

(2) 初始化权值:

$$W_{1,i} = \begin{cases} \frac{1}{2m} & y_i = 0 \\ \frac{1}{2l} & y_i = 1 \end{cases} \quad (2-11)$$

其中 l 、 m 分别为正、负样本的个数, $l + m = n$ 。

(3) 当 $t = 1, 2, \dots, T$ 时:

① 规范化权值:

$$W_{t,i} = \frac{W_{t,i}}{\sum_{j=1}^n W_{t,j}} \quad (2-12)$$

W_t 呈概率分布。

② 对每个特征 j , 训练一个分类器 h_j , 计算对应 W_t 的错误率:

$$\varepsilon_j = \sum_i W_{t,i} |h_j(x_i) - y_i| \quad (2-13)$$

③ 选取最小错误率的分类器 h_j 。

④ 更新权值:

$$W_{t+1,i} = W_{t,i} \beta_i^{1-\varepsilon_i} \quad (2-14)$$

若样本 x_i 被正确分类则 $\varepsilon_i = 0$, 否则 $\varepsilon_i = 1$; $\beta_i = \frac{\varepsilon_i}{1-\varepsilon_i}$ 。

(4) 最终构成的强分类器为:

$$h(x) = \begin{cases} 1 & \sum_{i=1}^T a_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^T a_i \\ 0 & \text{其它} \end{cases} \quad (2-15)$$

其中 $a_i = \log \frac{1}{\beta_i}$ 。

2.1.3 Adaboost 级联分类器

在得到强分类器后,使用级联结构可以使强分类器构成级联结构。Viola^[40]提出的级联结构的分类器是由一系列的分类器串联组成。对待识别的样本进行判别时,只有前面一级的分类器判决为正的样本才被送入后面的分类器继续处理,反之,则认为是负样本直接拒绝。整个分类器的输出是被所有分类器判决为正的样本。在级联结构中,前几级的分类器结构比较简单,特征数较少,但检测率很高,同时能够尽可能地滤除那些与目标差异很大的负样本。后面的分类器则使用更复杂的结果和更多特征,从而可以将那些与目标相似的负样本与正样本区分开来。所以级联结构的 Adaboost 分类器能快速排除背景区域,保留感兴趣区域,提高了检测的速度。

级联结构的 Adaboost 分类器如下图所示:

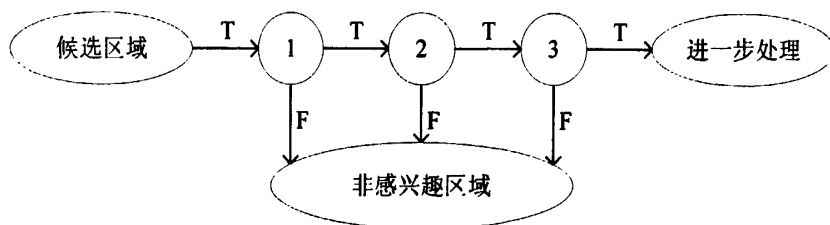


图 2.3 级联结构分类器

2.2 人眼定位

Reinders 等^[41]提出一种基于神经网络的眼睛定位算法,该方法将搜索窗口的像素作为神经网络的输入,如果该窗口包含眼睛图像,则神经网络的输出较大。Huang^[42]等提出一种基于遗传算法和决策树的眼睛定位方法,该方法采用一种混合遗传结构使基本视觉规则不断进化,最终得到可用于眼睛定位决策树形式的视觉规则。

投影法也是眼睛定位中经常用到的一种技术,Feng 和 Yuen^[43]提出将积分投影应用于人眼定位中。戴景文等^[44]提出一种基于投影峰的眼睛快速定位算法,该算法根据人脸的几何分布特征设置眼睛候选窗口,再通过直方图阈值分割出眼睛,最后经过灰度积分投影得到一系列投影峰,经过比较筛选得到眼睛的精确位置。Kith 等^[45]对投影法加以改进,提出在灰度图上基于特征和外观的眼睛定位方法,

该方法先使用基于特征的方法,利用投影函数确定左右眼所在可能区域;然后滤除所有非眼睛区域,仅保留左右眼区域,并且所有可能眼睛区域经过 PCA 处理;最后用混合投影函数定位瞳孔中心。

眼睛定位的一般步骤是先通过某种算法,找出眼睛的候选区域,然后用其它算法对眼睛精确定位。本文先采用 Adaboost 算法找出眼睛的候选区域,再根据眼睛的结构特征,用 Sobel 算子对眼睛候选区域处理,最后分别采用基于积分投影和基于多掩模的方法对人眼精确定位。

2.2.1 Sobel 算子

Sobel 算子是图像处理中常用的算子之一,主要用作边缘检测,它是一离散差分算子,用来计算图像亮度函数的梯度近似值。该算子包含两组 3×3 的模板,分别为水平模板及垂直模板,将之与图像作平面卷积,可分别得出水平及垂直的亮度差分近似。如果以 A 代表原始图像, G_x 、 G_y 分别代表水平梯度图、垂直梯度图。其计算公式如下:

$$G_x = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix} * A \quad (2-16)$$

$$G_y = \begin{pmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} * A \quad (2-17)$$

可用式(2-18)计算图像中每一个像素点的梯度值,式(2-19)用来计算梯度方向。

$$G = \sqrt{G_x^2 + G_y^2} \quad (2-18)$$

$$\theta = \arg \tan\left(\frac{G_y}{G_x}\right) \quad (2-19)$$

2.2.2 基于 Adaboost 算法和积分投影的人眼精确定位

图像投影法在提取图像特征时是一种非常有效的方法。通常用两个 1-D 正交投影函数来表达一幅 2-D 图像,从而减小了运算量。1998 年 Feng 和 Yuen^[43]提出将积分投影应用于人眼定位中。人脸图像中的眼睛区域灰度值通常比周围区域的灰度值要低小,利用该特征常使用灰度积分投影的方法来定位人眼。

若 $I(x, y)$ 为点 (x, y) 的像素灰度值,则在区间 $[x_1, x_2]$, $[y_1, y_2]$ 的垂直积分投影和水平积分投影函数分别为 $S_v = \sum_{y=y_1}^{y_2} I(x, y)$, $S_h = \sum_{x=x_1}^{x_2} I(x, y)$ 。积分投影函数可以反映出图像在水平或者垂直方向上总体灰度值情况,因此可以通过对眼睛区域进行

积分投影来判断瞳孔位置。该算法对左右眼分别处理，这样做的好处是对正脸且左右倾斜角度不大的情况下，该法有较好的适应性。主要步骤如下（以计算左眼瞳孔坐标为例）：

1. Adaboost 算法检测人眼区域。
2. Sobel 算子对检测到的人眼区域进行边缘检测。
3. 二值化。
4. 对二值化图像作垂直积分投影和水平积分投影。
5. 分别搜寻垂直积分投影图和水平积分投影图双峰间的谷底，便可求出对应左眼瞳孔的横坐标和纵坐标。

右眼瞳孔坐标的求取步骤同上。灰度积分投影算法流程图如图 2.4 所示：

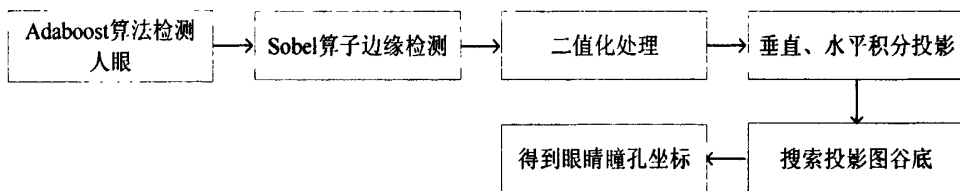


图 2.4 灰度积分投影算法流程图

理想情况下，眼睛中心横坐标和纵坐标分别为垂直灰度投影图和水平灰度投影图双峰间的谷底所对应的坐标。但是在实际处理时，有以下三个不利因素：1. 由于刘海、眉毛、鬓角等因素干扰，使得它们的灰度值甚至比瞳孔的灰度值还低；2. 光照影响使眼睛内部分区域产生阴影，这对灰度投影是极其不利的；3. 投影得到的灰度直方图是离散不平滑的，而且由前两个原因造成直方图曲线非严格单调上升或者非严格单调下降，造成出现多个极大峰的情况，很难确定哪两个峰值中间的谷底是瞳孔中心坐标。

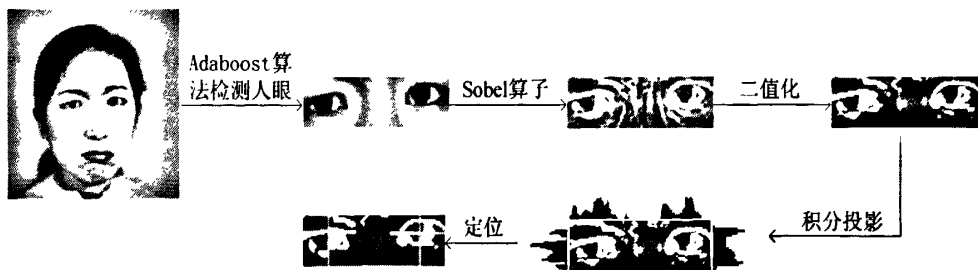


图 2.5 灰度积分投影算法示意图

图 2.5 为灰度积分投影算法示意图，由定位结果可以看出左右眼瞳孔定位结果都存在偏差，而且右眼纵坐标定位偏差较大，这是因为定位眼睛的窗体过大，在边缘特征提取时将右眼眉毛的边缘特征提取出来对眼睛定位造成影响。

为了减小光照不均使眼睛内产生阴影和眉毛对瞳孔定位的影响，以及避免投影直方图离散、非严格单调的问题，本文在求取瞳孔坐标时，分别用式(2-16)和式(2-17)对 Adaboost 算法检测出来的人眼大概区域进行处理，得到眼睛区域纵向轮廓

比较明显的水平梯度图和眼睛区域横向轮廓比较明显的垂直梯度图。在有眉毛干扰的情况下，垂直梯度图将眉毛的边缘特征也提取出来，但对单只眼睛而言，眉毛是左右对称的，对求取眼睛的横坐标影响较小；在光照不均引起眼睛内区存在阴影的情况下，水平梯度图将阴影的边缘特征也提取出来，对单只眼睛而言，阴影是上下对称的，对求取眼睛的纵坐标影响较小。因此可以将梯度图二值化后，分别用纵向轮廓明显的人眼区域来确定眼睛的横坐标，而用横向轮廓明显的人眼区域确定眼睛的纵坐标。主要步骤如下（以计算左眼瞳孔坐标为例）：

1. Adaboost 算法检测人眼区域。

2. Sobel 算子水平模板对检测到的人眼区域处理得到水平梯度图；Sobel 算子垂直模板对人眼区域处理得到垂直梯度图。

3. 分别对垂直梯度图、水平梯度图二值化。

4. 对二值化后的垂直梯度图求取左眼瞳孔横坐标，对其左眼区域求其白色像素点横坐标的“分布中心”（Distribution Center, DC），在下文中，该方法被称为 DC 法。计算方法为：若 $I(x, y) = 255$ ，则 $x_{cl} = \frac{1}{N * M} \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} x$ ，其中 N 和 M 分别为左眼区域的高和宽，下文相同；

对二值化后的水平梯度图求取左眼瞳孔纵坐标，对其左眼区域求其白色像素点纵坐标的“分布中心”，计算方法为：若 $I(x, y) = 255$ ，则 $y_{cl} = \frac{1}{N * M} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} y$ ；

右眼瞳孔坐标计算方法同上。由图 2.6 可以看出该算法对存在眉毛和眼睛内区有阴影干扰的情况下有一定的适应性。

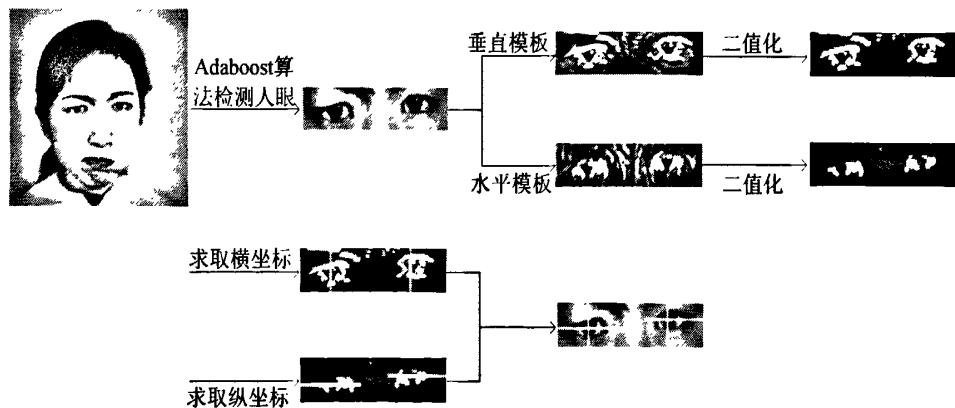


图 2.6 DC 法人眼精确定位示意图

2.2.3 基于 Adaboost 算法和多掩模的人眼精确定位

掩模运算是指用选定的图像、图形或物体，对待处理的图像(全部或局部)进行遮挡，来控制图像处理的区域或处理过程。用于覆盖的特定图像或物体称为掩模

或模板。数字图像处理中,图像掩模主要有以下几个作用:1.提取感兴趣区域。用预先制作的感兴趣区域掩模与待处理图像相乘,得到感兴趣区域图像,感兴趣区域内图像值保持不变,而区域外图像值都为0;2.屏蔽作用。用掩模对图像上某些区域作屏蔽,使其不参加处理或不参加处理参数的计算,或仅对屏蔽区域作处理或统计;3.结构特征提取。用相似性变量或图像匹配方法检测和提取图像中与掩模相似的结构特征;4.特殊形状图像的制作。

眼睛区域计算白色像素点 DC 法虽然对存在眉毛干扰的情况下有一定的容忍度,但是眉毛对眼睛定位的干扰不容忽视,为了进一步减小其干扰,本文对 DC 法进一步改进,提出了基于 Adaboost 和多掩模的人眼精确定位算法。该算法为了缩小眼睛区域的范围,在 DC 法定位到眼睛大概坐标的情况下,分别构造出矩形掩模和椭圆形掩模,然后在掩模区域对眼睛进行精确定位。

1. 矩形掩模的获取

以左眼为例,在 DC 法定位到左眼坐标 (x_l, y_l) 的情况下,以 x_l 为中心左右扩展 $\frac{1}{6}W$,以 y_l 为中心上下扩展 $\frac{1}{3}H$ 便得到矩形掩模(其中 W, H 为 Adaboost 算法检测出来的眼睛区域的宽和高),该矩形掩模能消除大部分眉毛区域。

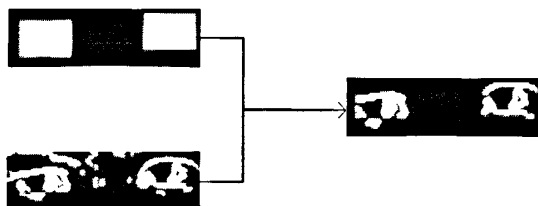


图 2.7 矩形掩模缩小眼睛范围示意图

矩形掩模的不足在于:人眼为非矩形,而是类椭圆形,该算法虽然消除了大部分眉毛区域,但是一部分眼睛轮廓外非眼睛区域经过矩形掩模处理后也被当成了眼睛区域保留下来。为了解决矩形掩模的不足,本文在矩形掩模的基础上再使用椭圆形掩模来进一步缩小眼睛范围。

2. 椭圆形掩模的获取

对二值化人眼区域 Sobel 算子梯度图作相关处理可得到椭圆形掩模,该算法也是对左右眼分别处理,主要步骤如下:

(1)两次查找眼睛轮廓,并用最小二乘法拟合椭圆^{[46][47][48]}。第一次查找轮廓:允许存在较多的轮廓细节,将轮廓上的点存起来,找到最大轮廓和次最大轮廓,椭圆拟合,填充拟合椭圆内部为白色。第二次查找轮廓:目的是为了消除点集较少的轮廓,使特征明显的轮廓更加突出,再次找到最大轮廓和次最大轮廓,椭圆拟合,填充拟合椭圆内部为白色。

(2)腐蚀。由于光照影响,在检测轮廓时可能将非眼睛轮廓检测出来,腐蚀能有效减小非眼睛区域轮廓的影响。

(3)膨胀。膨胀恢复腐蚀所产生的眼睛区域变小问题，经过膨胀后获得椭圆形掩模。

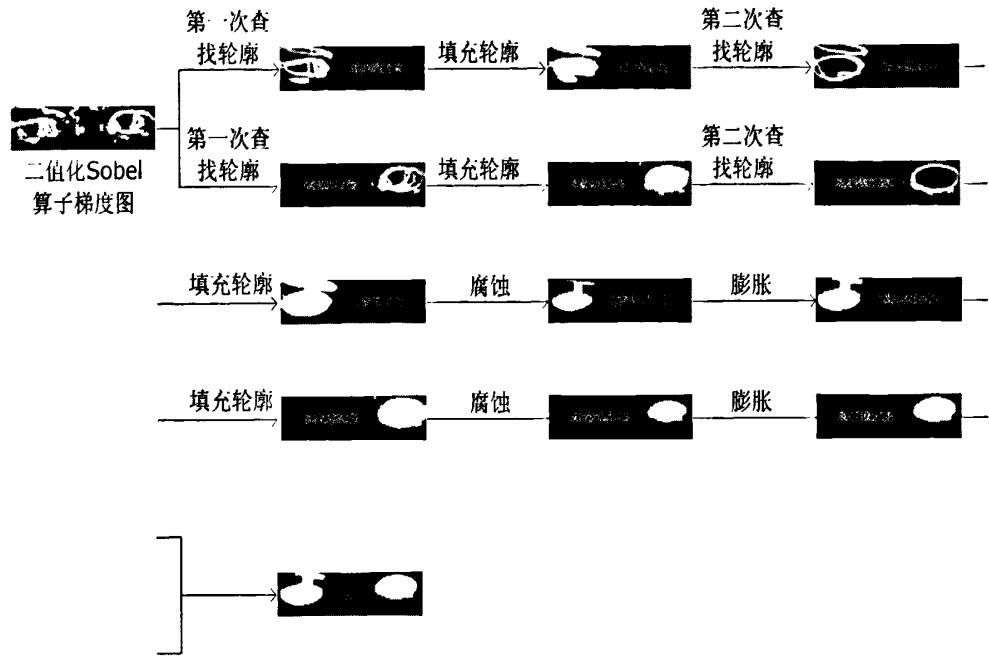


图 2.8 椭圆形掩模获取示意图

3. 矩形掩模和椭圆形掩模的结合

由图 2.8 可知椭圆形模板有效地缩小眼睛的范围，但是仍存在眉毛的干扰，所以为了减小眉毛的干扰，则将矩形模板和椭圆形模板相结合，产生一个新的掩模，在新的模板下对眼睛精确定位。图 2.9 为矩形掩模和椭圆形掩模共同作用产生的新掩模。

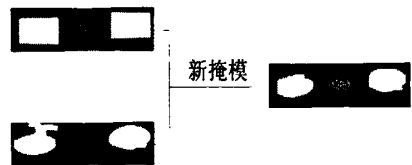


图 2.9 最终掩模示意图

4. 基于多掩模的人眼精确定位

由矩形掩模和椭圆形掩模产生的新掩模能有效地去除非眼睛区域，该掩模对眼睛定位的精确度有了进一步提升。以左眼为例，计算方法如下：若 $(x_i, y_i) \in R_l$ 且 $I(x_i, y_i) = 0$ ，则 $x_{cl} = \frac{1}{N} \sum_{i=0}^{N-1} x_i$ ， $y_{cl} = \frac{1}{N} \sum_{i=0}^{N-1} y_i$ 。其中 R_l 为左眼掩模区域， $I(x_i, y_i)$ 为 (x_i, y_i) 点的灰度值， x_{cl} 、 y_{cl} 为左眼瞳孔的横坐标和纵坐标。右眼瞳孔坐标计算方法同上。

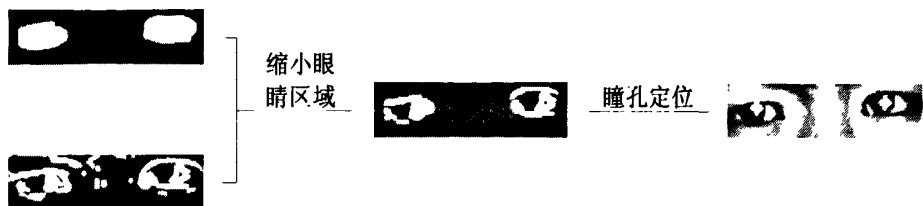


图 2.10 基于多掩模的人眼精确定位示意图

2.2.4 实验结果及分析

1. 归一化误差

本文采用基于实际眼睛坐标和实验眼睛坐标之间的距离的相对误差来衡量定位眼睛的性能。如图 2.11(a)所示, C_r 、 C_l 为实际眼睛中心, $\tilde{C}_r$ 、 $\tilde{C}_l$ 为实验眼睛坐标中心, d_r 、 d_l 分别为 C_r 到 $\tilde{C}_r$ 的欧氏距离和 C_l 到 $\tilde{C}_l$ 的欧氏距离。本文使用 $\max(d_r, d_l)$ 的最大距离来衡量算法性能, 该距离除以实际眼睛距离 $\|C_r - C_l\|$ 得到归一化距离, 使它与图像尺寸和图像中人脸大小无关。该距离可表示为:

$$e = \frac{\max(d_r, d_l)}{\|C_r - C_l\|} \quad (2-20)$$

在下文中将距离测度 e 称为归一化误差。如图 2.11(b)所示, 通常情况下, 人脸内眼角的距离等于单只眼睛的宽, 归一化误差 $e = 0.25$ 等于单只眼睛宽的一半。在该测度中, $e \leq 0.25$ (四分之一眼距) 大概是眼睛中心到眼角之间的距离, $e \leq 0.1$ 是虹膜的范围, $e \leq 0.05$ 是瞳孔的范围。

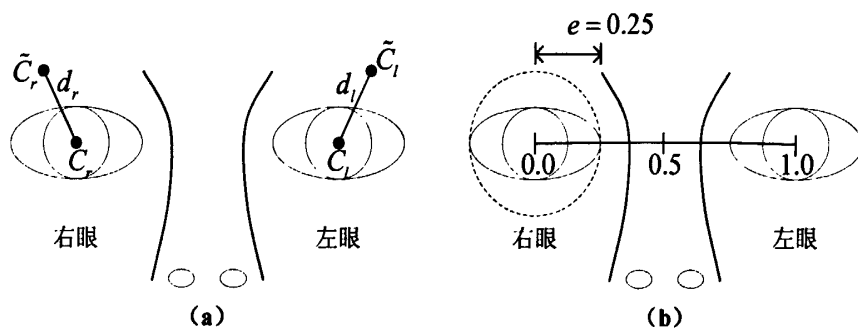


图 2.11 人眼关系图和归一化误差

2. 实验结果及分析

本实验在 JAFFE 人脸表情库和 BioID 数据库上进行算法性能分析。JAFFE 人脸表情库中, 实际眼睛坐标是手工标定; 而 BioID 数据库中的人眼实际坐标是数据库已给出的, 只需将实际坐标与实验坐标利用归一化误差来衡量算法的性能。

(1) JAFFE 数据库上的实验

JAFFE 数据库共 313 张人脸表情图片, 包含 10 个人的 7 种正面表情, 选取每个人每种表情 3 张, 共 310 张, 图片格式为 256×256 的灰度图像。

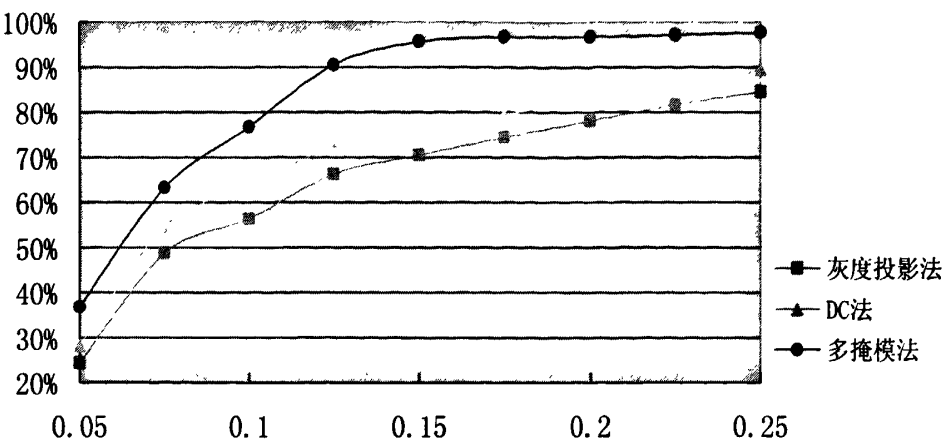


图 2.12 JAFFE 库上不同算法的性能

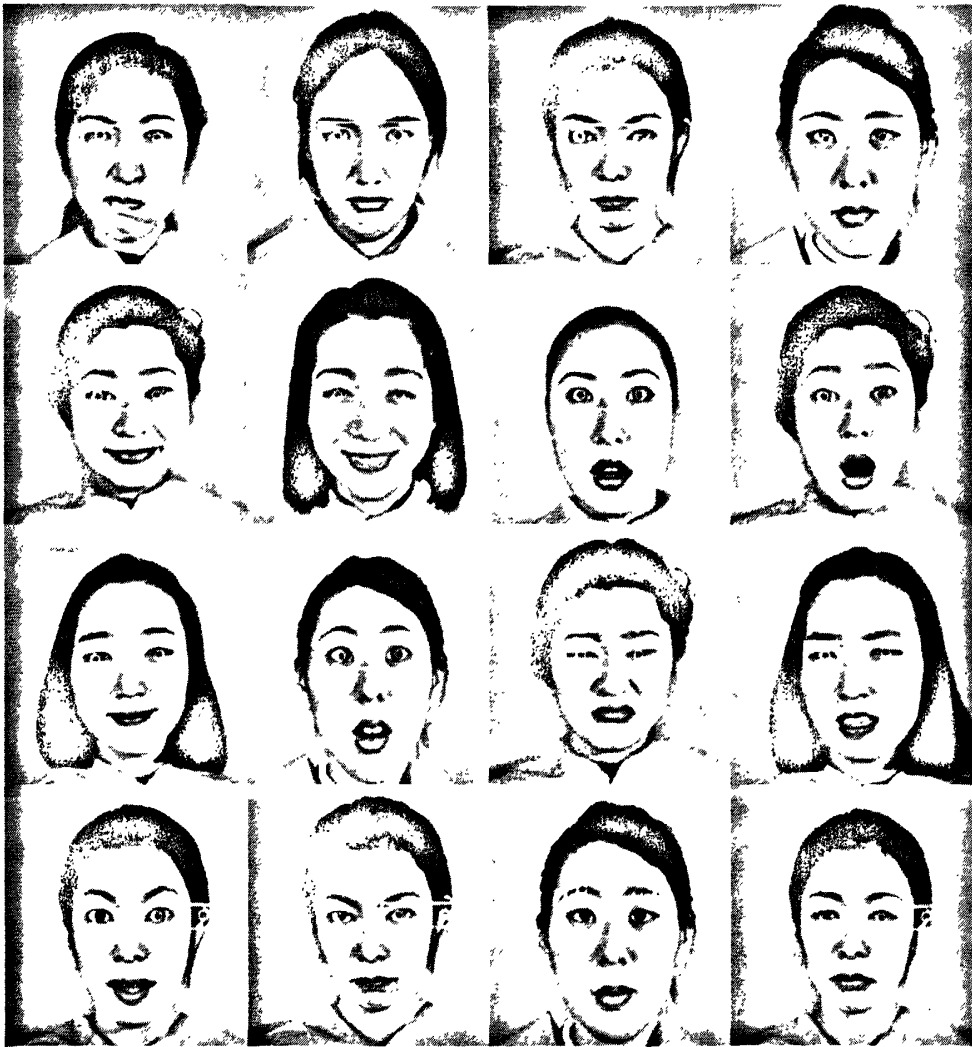


图 2.13 多掩模法在 JAFFE 库人眼定位示例图

(2) BioID 数据库上的实验

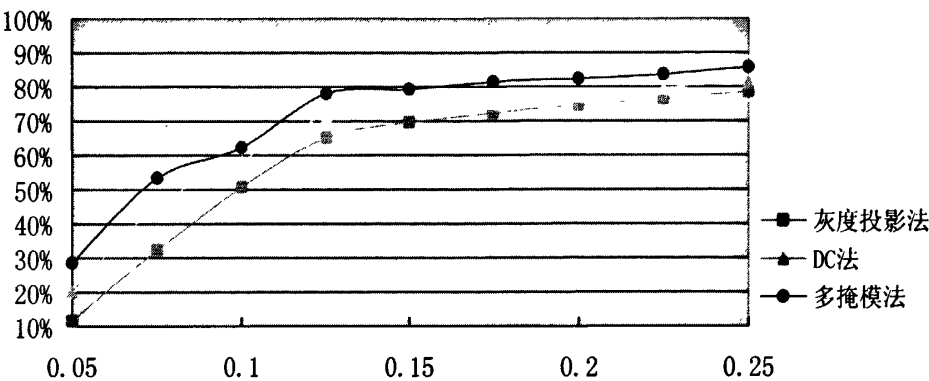


图 2.14 BioID 库上不同算法的性能



图 2.15 多掩模法在 BioID 库人眼定位示例图

BioID 人脸数据库包含由 23 个测试者提供的 1521 张 384×286 灰度自然场景

下的人脸图像。

从图 2.12 和图 2.14 可以看出，本文基于多掩模的人眼精确定位算法在 JAFFE 数据库中，有较高的识别率，该数据库是在较理想的环境中拍摄的，无遮挡物并且光照均匀。但是在 BioID 人脸库中，识别率下降，主要原因有：1. BioID 数据库强调的是自然场景下的人脸，因此该数据库中的图像在光照、背景方面变化幅度较大；2. 该数据库中有较多的戴眼镜的人脸图像存在，眼镜反光造成部分眼睛定位不准。

2.3 人脸图像的预处理

在定位到眼睛的位置后，需要根据一定的比例裁剪出人脸图像进行后续处理。由于图像采集设备、光照条件、人脸相对镜头位置差异，应该对裁剪出的人脸图片进行预处理。在实际应用中，预处理工作包含两个方面：一方面，获取的图像信息不符合后续操作的具体要求而进行的预处理，例如尺度，角度的归一化。另一方面，原始图像由于噪声、光照等原因，图像质量不高，需进行光照预处理，以利于提取感兴趣的信息。

2.3.1 几何预处理

1. 旋转校正

在精确定位得到眼睛坐标后，依据左右瞳孔的坐标及其连线相对于水平方向偏转的角度对人脸图像进行旋转矫正。

偏转角度可由下式得出：

$$\theta = \arctan\left(\frac{y_r - y_l}{x_r - x_l}\right) \tag{2-21}$$

其中 $(x_l, y_l), (x_r, y_r)$ 分别为左右瞳孔坐标。当 θ 为正时顺时针旋转图像，为负时逆时针旋转图像。

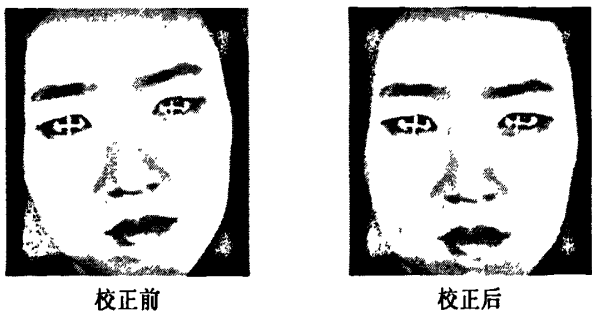


图 2.16 人脸图像旋转矫正

2. 图像裁剪与尺度归一化

由于背景和头发等非人脸区域的存在对表情识别存在影响，故需要从含有人

脸图像的图片中裁剪出人脸。但裁剪出的人脸图像大小不一，为了方便后续的特征提取，还需要对裁剪出的人脸图像进行归一化。目前较常用的裁剪规则是图 2.17 所示的比例。该方法根据人脸面部的结构特征，设左右眼中心的距离为 d ，垂直方向向上取 $0.5d$ ，向下取 $1.5d$ ；水平方向以两眼连线的中点为基准，向左、向右各取 d 。

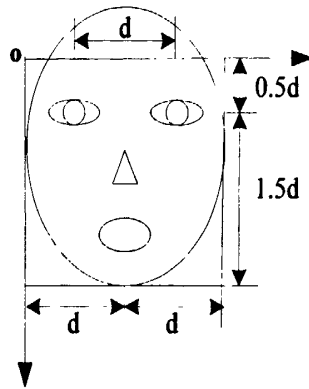


图 2.17 人脸图像裁剪比例示意图

该比例模型的裁减比例能有效去除头发、肩颈等对识别起到干扰作用的区域，同时又合理地分割出人脸区域。

裁剪得到的人脸图像大小不一，需要经过缩放变换将样本图像统一处理为尺寸为 100×100 的标准图像。



图 2.18 经裁剪和尺度归一化后的图像

2.3.2 光照预处理

表情数据库中部分图像存在光线较暗，图像质量差等缺陷，因此需要对这部分图像进行光照补偿，增强对比度，以达到改善图像质量的目的。

本文采用直方图均衡化(Histogram Equalization)把原始图像的灰度直方图从比较集中的某个灰度区间变成在全部灰度范围内的均匀分布。

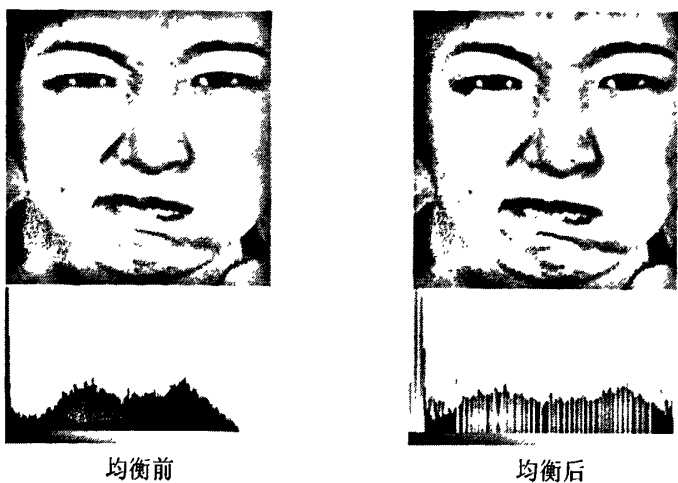


图 2.19 经尺度归一化和直方图均衡化后的图像

2.4 小结

本章首先详细介绍了 Adaboost 算法的基本原理和 Adaboost 级联分类器的构造方法，接着详细论述了灰度投影法定位人眼的主要方法，并详细阐述了本文提出的基于 Adaboost 和多掩模的人眼精确定位方法的算法实现。实验表明，本文提出的算法人眼定位效果良好。在精确定位到眼睛的坐标后，根据一定的比例裁剪出人脸图像，对人脸图像进行预处理，方便后续处理。

第三章 基于中心化二元模式的人脸表情特征提取

人眼坐标得到后可以根据一定的比例裁剪出人脸区域,进而提取人脸表情特征。人脸表情特征提取方法主要有:基于 PCA 的方法^[11]、基于 Gabor 的小波变换法^[17]、基于 AAM 的方法^[49]、基于 LBP 的方法^[14]等。基于 PCA 的方法能有效地将高维人脸图像投影到低维空间中,但是在样本类间离散度增大时,样本类内的离散度也在增大。基于 Gabor 小波变换采用多尺度多方向 Gabor 滤波器进行特征提取后,特征向量维数很高,运算量大,不适合建立高效快速的人脸表情识别系统。基于 AAM 的方法虽能获得可靠的人脸面部特征,但计算复杂,初始化参数难以获取。基于 LBP 的方法具有计算简单、抗光线干扰强等优点,与基于 Gabor 小波方法相比, LBP 能更快的提取人脸面部特征,而且维数较小,同时保留了人脸的有效信息。

3.1 局部二元模式(LBP)

局部二元模式(LBP)是一种有效的统计型纹理描述算子,可以对灰度图像中局部近邻区域的纹理信息进行度量和提取。近年来 LBP 算子已经广泛地应用于图像检索、纹理分类、人脸图像分析等领域^{[50][51]}。LBP 算子是一种用来描述图形纹理特征的算子,它通过对图像中的一点与其周围点的灰度值大小关系的准确描述来表征图像中局部纹理的空间结构,具有一定的旋转和光照不变性。采用 LBP 算子对一幅图像处理时,先要确定像素的环形邻域,该邻域是由半径为 R 的圆周上对称分布的 P 个邻域点构成, R 和 P 可以为任意值。如图 3.1 所示,若近邻点的位置不在像素点中心时,灰度值由线性插值方法确定。

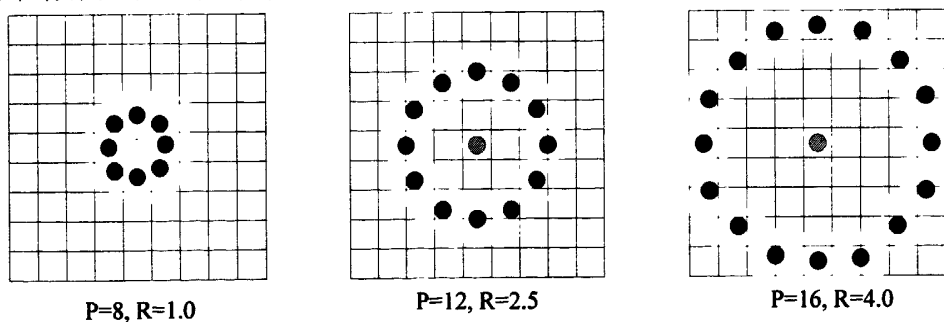


图 3.1 (R, P) 取不同值时环形邻域

假设中心像素点坐标为 (x_c, y_c) , 在半径为 R 的 P 个邻域点坐标为 (x_p, y_p) , 可通过式(3-1)计算:

$$\begin{aligned} x_p &= x_c + R \cos(2\pi p / P) \\ y_p &= y_c + R \sin(2\pi p / P) \end{aligned} \quad (3-1)$$

如果中心点像素的灰度值为 g_c ，其邻域点的灰度值为 $g_p (p=0,1,...,P-1)$ ，则像素 g_c 的局部灰度联合分布定义为：

$$T=t(g_c, g_0-g_c, ..., g_{P-1}-g_c) \tag{3-2}$$

假设 $g_p - g_c$ 与 g_c 是相互独立的，则式(3-2)可变为：

$$T=t(g_c)t(g_0-g_c, ..., g_{P-1}-g_c) \tag{3-3}$$

其中 $t(g_c)$ 代表图像的全局亮度，没有提供有益于纹理分析的信息，因此去掉 $t(g_c)$ 将不影响图像局部结构的描述。式(3-3)可用下式描述：

$$T=t(g_0-g_c, g_1-g_c, ..., g_{P-1}-g_c) \tag{3-4}$$

T 是一个纹理算子，该算子有着高度的鉴别力，它用一个 P 维直方图统计所有可能出现的不同邻域像素分布情况。由式(3-3)可看出，对于灰度无变化区域在各个方向上的灰度差值均为 0。对于渐变边缘，该纹理算子在梯度方向上有最大灰度差值，但在边界方向上的值为 0。 $g_p - g_c$ 的符号函数不受平均亮度变化的影响，联合差值分布不受灰度等级变化的影响，则可用灰度差值的符号函数来代替真实的灰度差：

$$T=t(s(g_0-g_c), s(g_1-g_c), ..., s(g_{P-1}-g_c)) \tag{3-5}$$

其中符号函数 $s(x)$ 的定义为：

$$s(x)=\begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3-6}$$

为每个符号函数 $s(g_p - g_c)$ 赋予一个权重因子 2^p ，则可以得到 LBP 码：

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \tag{3-7}$$

6	7	2
8	7	15
4	10	5

例子

0	1	0
1		1
0	1	0

门限

二进制: 10101010

1	2	4
128		8
64	32	16

权值

十进制: LBP码=2+8+32+128=170

图 3.2 LBP 编码示例图

图 3.2 为 ($P=8, R=1.0$) 时 LBP 码的计算方法。由图 3.2 可看出，当图像中灰度值次序不变时，该 LBP 码就不变，因此 LBP 算子对灰度尺度变换具有不变性。

若 LBP 算子的二进制码被看成是一个首尾相连的环形时，至多存在两位由 1 到 0 或由 0 到 1 的转换，则该 LBP 码被称为统一 LBP 码^[48]。例如 00000000、11111111 无码元变化，00001111、00011100 两次码元变化，都属于统一模式；而 11100110、01100110 四次码元变化，都属于非统一模式。统一模式 LBP 与传统 LBP 相比不仅能有效地压缩数据，而且只检测重要的纹理特征。采用 LBP 码可以用一幅图像中

所有像素点的 LBP 码来表示人脸图像。

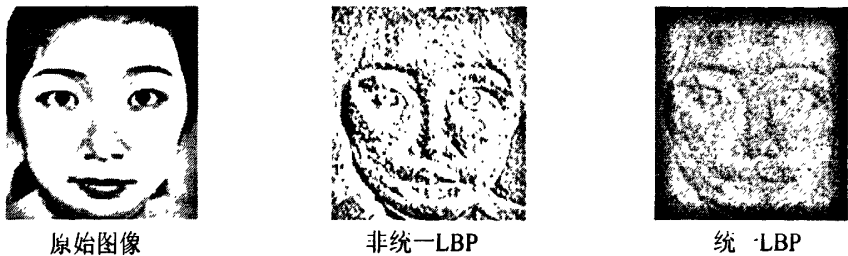


图 3.3 LBP 处理过的图像

为了有效的表示人脸的空间信息，人脸图像被划分为多个子块，每个子块的 LBP 直方图代表了该区域内的纹理变化情况。由于人脸上的不同器官的表情特征在表情识别中贡献不同，因此对不同的器官赋予不同的权重。如眼睛、嘴巴对表情识别的贡献较大，赋予它们较大的权重；而额头、脸颊、下巴这些对表情识别贡献小的部位则赋予较小权重。如图 3.4(b)所示，黑色区域对应权值为 0，深灰色权值为 1，浅灰色权值为 2，白色权值为 4。

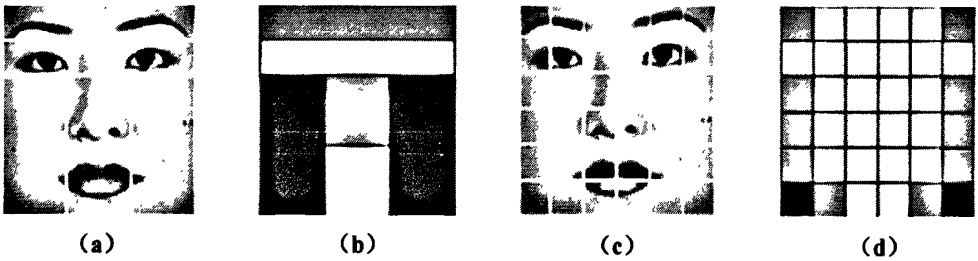


图 3.4 (a)人脸被分为 6 个子区域；(b)不同区域集合

(c)人脸被分为 6×6 子块； (d)分块加权

对每个子块分别进行 LBP 处理，处理后的 LBP 码乘以相应的权值得到的 LBP 直方图作为表情的特征向量既能反映表情的局部信息，又加入了整体信息。

3.2 中心化二元模式

3.2.1 LBP 算子的不足

虽然 LBP 算子能有效地描述图像局部邻邻区域的纹理特征，但是传统 LBP 算子有着自身的不足：

1. 产生的直方图维数过长，导致表情识别速度低下。以图 3.4 为例，人脸图像划分的子块数为 36，若 LBP 算子中的 P 值取 8，则有 256 个不同的 LBP 码。如果选择非统一模式对图像进行特征提取，则有 $256 \times 36 = 9216$ 维特征向量；如果选择统一模式，则有 $59 \times 36 = 2124$ 维特征向量。
2. LBP 算子没有考虑到中心像素点的作用，在某些情况下遗漏了局部信息。

以 $LBP(8,1)$ 为例, 由式(4-6)仅得出 256 种 LBP 码, 但是共有 512 中模式。其原因是仅仅考虑到了周围邻近像素点的作用, 没有将中心像素点作用考虑在内。然而, 某些情况下中心像素点能提供比邻近像素点更多的信息。

3. 产生的二值数据对噪声反应敏感。传统的 LBP 算子考虑到邻近像素与中心像素点的作用, 如果 $g_c - g_p > 0$, 则 $s(g_c - g_p) = 1$; 但关系是相互的, 也应该有 $g_p - g_c < 0$, $s(g_p - g_c) = 0$ 。

3.2.2 中心化二元模式

针对 LBP 算子的不足, 本文采用 CBP 算子^[52] (Center-based Nearest Neighbor) 提取人脸表情特征。CBP 算子是 LBP 算子的改进, CBP 算子如图 3.5 所示。

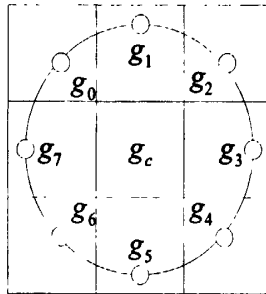


图 3.5 CBP(8,1)算子

$$CBP(8,1) = s(p_0 - p_4)2^0 + s(p_1 - p_5)2^1 + s(p_2 - p_6)2^2 + s(p_3 - p_7)2^3 + s(p_c - p_7)2^4 \quad (3-8)$$

其中 $p_r = (p_0 + p_1 + \dots + p_7 + p_c) / 9$ 。

CBP 算子针对 LBP 三点不足有以下改进:

1. 维数比 LBP 低, 而且 CBP 能捕捉到更好的梯度信息。
2. 考虑到中心像素点的作用并赋予其最高权重。
3. 改写符号函数 $s(x)$, 降低白噪声对图像的影响, 改写如下:

$$s(x) = \begin{cases} 1 & |x| \geq C \\ 0 & |x| < C \end{cases} \quad (3-9)$$

其中 C 为阈值。

CBP 算子可由式(3-10)表示。

$$CBP(P,R) = \sum_{p=0}^{(P/2)-1} s(g_p - g_{p+(P/2)})2^p + s(g_c - \frac{1}{P+1}(g_c + \sum_{p=0}^{P-1} g_p))2^{P/2} \quad (3-10)$$

其中, P 和 R 分别为环形邻域中紧邻点的个数和圆周半径, g_c 为中心像素点灰度值, $g_p (p=0,1,\dots,P-1)$ 为周围邻近像素点灰度值。

通过 CBP 与 LBP 的比较可知, CBP 比 LBP 具有三大优势:

1. CBP 算子产生的直方图维数远小于 LBP 算子产生的直方图维数。以 $(P=8,R=1)$ 为例,在对图像不分块的情况下,非统一模式 LBP 算子有 256 维,统一模式 LBP 算子 59 维,而 CBP 算子为 32 维。此外, CBP 通过比较经过中心像素点且处于同一条直线上的“近邻点对”之间的差异能更好地捕捉到梯度信息,大大提高 CBP 的鉴别能力。

2. CBP 算子将中心像素点的作用考虑进来且赋予最高权重,增强了中心像素点的作用且有利于提高鉴别力。图 3.6 为 LBP 与 CBP 鉴别能力的比较, CBP 中的阈值常数 C 取 10。第一列为两个局部结构信息不同的 3×3 像素块,第二列为 LBP 码,第三列为 CBP 码。由图 3.6 可以看出两个不同的像素块经 LBP 编码后得出相同的结果,但事实上这两个原始像素块的局部结构信息完全不同;经过 CBP 编码得出不同的结果,所以 CBP 算子能更精确地描述图像的局部结构信息。

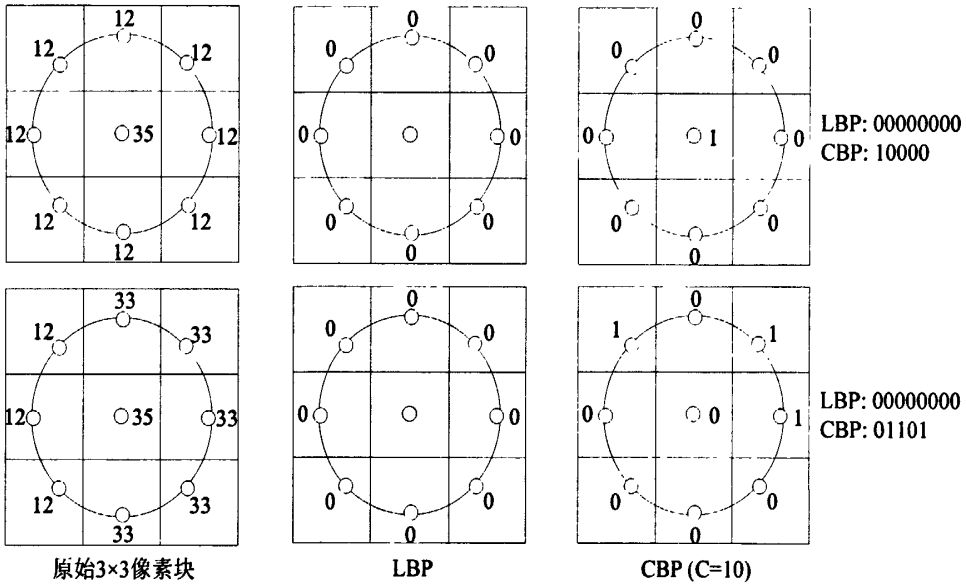


图 3.6 LBP 与 CBP 的比较

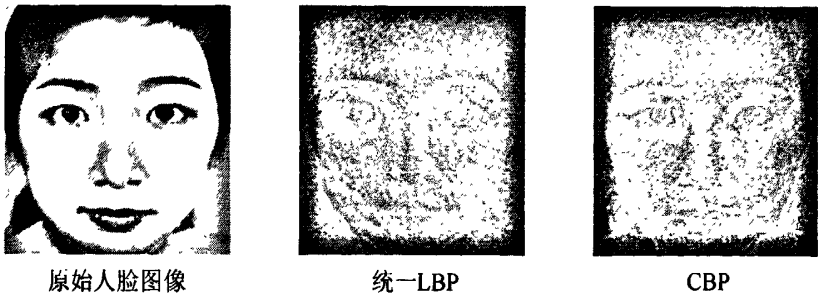


图 3.7 图像经 LBP 和 CBP 处理后的结果

图 3.7 为经过统一模式 LBP 和 CBP 算子处理过的人脸图像,可以看出, CBP 处理后的眼睛、嘴巴纹理,尤其是纹理细节方面明显优于统一模式 LBP,所以 CBP

算子比 LBP 算子有更强的鉴别力。

3. CBP 算子对白噪声反应不敏感。该优点可以通过实验验证得到。

3.3 实验结果及分析

3.3.1 实验参数选择

本章将 LBP、加权 LBP、CBP 和加权 CBP 算法在 JAFFE 人脸表情数据库上进行比较实验。所用图片是经过预处理的人脸表情图片。所有分类器均采用最近邻分类器(NN)，四种方法依次表示为：LBP+NN、加权 LBP+NN、CBP+NN 和加权 CBP+NN。本实验采用欧式距离度量样本特征间的距离。

LBP+NN 方法中，选用统一 LBP 算子，且 $(P=8, R=2)$ 。将预处理后的人脸图像按图 3.4(c)所示分块，对每个子块提取 LBP 直方图，并将其连接起来作为人脸图像 LBP 特征(59×36 维)。在识别阶段，将测试样本先预处理，后提取其 LBP 特征，将 LBP 特征送入最近邻分类器，用欧氏距离度量 LBP 特征之间的距离，最后将测试样本归到与之距离最小的训练样本的类中。

在加权 LBP+NN 方法中，选用统一 LBP 算子，且 $(P=8, R=2)$ 。将预处理后的人脸图像按图 3.4(c)所示分块，对每个子块提取 LBP 直方图并乘以相应权重(如图 3.4(d)所示)，把子块的 LBP 直方图连接起来作为整幅人脸图像的特征(59×36 维)。在识别阶段，处理方法与 LBP+NN 方法相同。

CBP+NN 方法中，选取 $(P=8, R=2)$ ，阈值常数选 $C=0$ 。将预处理后的人脸图像按图 3.4(c)所示分块，将每个子块的 CBP 直方图连接起来作为人脸 CBP 特征(32×36 维)。在识别阶段，先将预处理后的测试样本进行 CBP 特征提取，再将其 CBP 特征送入最近邻分类器中，用欧氏距离度量 CBP 特征之间的距离，将测试样本归到与之距离最小的训练样本的类中。

在加权 CBP+NN 方法中，选取 $(P=8, R=2)$ ，阈值常数选 $C=0$ 。将预处理后的人脸图像按图 3.4(c)所示分块，对每个子块提取 CBP 直方图并乘以相应权重，把子块的 CBP 直方图连接起来作为整幅人脸图像的特征(32×36 维)。在识别阶段，处理方法与 CBP+NN 方法相同。

本章采用的精度测试方法是十折交叉验证(10-fold cross validation)，即将经过预处理后的人脸表情数据库中的每种表情分成 10 份，轮流将其中 9 份做训练，1 份做测试，10 次结果的平均值作为对算法精度的估计。实验在操作系统为 Windows XP，CPU: 3.1GHz，2G 内存的 PC 机下进行。

3.3.2 实验结果及分析

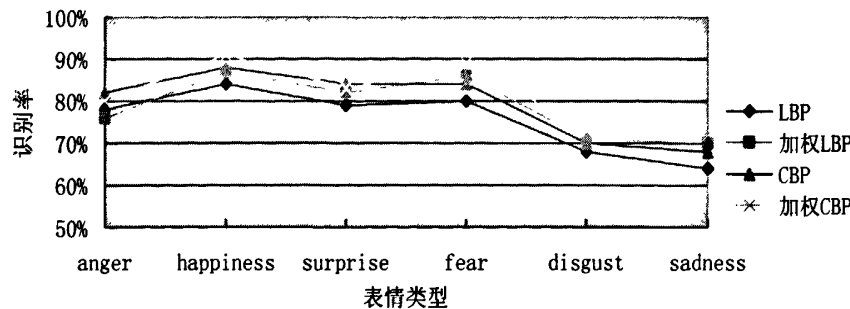


图 3.8 采用欧氏距离时各种方法在 JAFFE 库上的识别率

从图 3.8 可以看出，在六种基本人脸表情中高兴、愤怒、恐惧、吃惊的识别率要比厌恶、悲伤的识别率高。

表 3.1 采用欧氏距离时各种方法在 JAFFE 库上的结果

方法	平均识别率	识别一个样本所用时间(秒)
LBP+NN	75.5%	0.0182
加权 LBP+NN	78.6%	0.0189
CBP+NN	79.3%	0.0059
加权 CBP+NN	81.3%	0.0065

从表 3.1 可见，CBP+NN 方法的识别率高于 LBP+NN，加权 CBP+NN 方法的识别率高于加权 LBP+CNN，主要原因有：1. LBP 计算的是近邻像素点与中心像素点的差异，而 CBP 比较的是“近邻点对”之间的差异，CBP 更好的捕捉到梯度信息，有利于 CBP 鉴别力的提高；2. LBP 中没有考虑中心像素点的作用，而 CBP 考虑到中心像素点的作用并赋予最大权重，进一步提高了 CBP 的鉴别力。对同一副图像而言，分别采用 LBP 和 CBP 进行特征提取，LBP 特征 2124 维，而 CBP 特征为 1152 维，特征明显减少，提高了识别速度，这一特点可由表 3.1 中识别一个样本所用时间得到验证。

在实际应用中，图像容易受到噪声的污染，下面对各种方法的抗噪性进行对比实验。



图 3.9 (a)原始图像; (b)加入噪声后的图像

表 3.2 有噪声情况下采用欧氏距离时各种方法在 JAFFE 库上的结果

方法	平均识别率	识别一个样本所用时间（秒）
LBP+NN	68.5%	0.0203
加权 LBP+NN	71.6%	0.0212
CBP+NN	75.2%	0.0081
加权 CBP+NN	79.6%	0.0092

由表 3.2 可见，在图像被噪声污染的情况下，CBP+NN、加权 CBP+NN 方法的识别率均比 LBP+NN、加权 LBP+NN 识别率高，说明 CBP 比 LBP 的抗噪性要高，从而验证了 CBP 改写的符号函数有利于降低噪声对图像的污染。

3.4 小结

人脸表情的特征提取是表情识别中十分重要的一个环节，本章首先介绍了 LBP 算法的基本原理，然后介绍了 CBP 算法对 LBP 算法的三大改进。在特征提取阶段，本文采用中心化二元模式的方法提取表情特征，通过与 LBP 算法及其它算法的比较实验，CBP 在纹理特征提取和速度上有较大优势；同时，CBP 和 LBP 对有噪声污染图像的对比实验可以看出，CBP 比 LBP 有更好的抗噪性。

第四章 基于中心最近邻和 SVM 的人脸表情特征分类

在绪论中本文已经对表情分类算法进行了综述,目前表情分类的方法主要有:人工神经网络、隐马尔可夫模型、支持向量机等。人脸表情识别属于小样本问题,目前人脸表情数据库不多,各种表情样本比较少,因此本章对中心最近邻分类器和 SVM 分类器进行了尝试。CNN 不仅能解决小样本问题,而且能减小最近邻特征线^[53](NFL)的计算量,CNN 分类器也是在此基础上发展而来。人脸表情分类属于多分类问题,支持向量机是在统计学习理论的基础上发展起来的一种分类方法,将支持向量机用于表情识别很有实际意义。支持向量机既可以解决小样本问题也可以解决非线性分类的问题,能有效避免经典学习方法中过学习、维数灾难和局部极小等。因此 SVM 成功应用于人脸检测、人脸识别和表情识别等领域。

本章详细介绍和分析了 CNN 和各种多分类支持向量机方法,以及尝试将 CNN 和二对二多分类 SVM 用在表情分类中。实验表明二对二多分类 SVM 效果优于 CNN 和其它分类方法。

4.1 中心最近邻分类器

4.1.1 常用的距离度量函数

通常在获得待识别人脸表情图像的特征向量后,需要一个分类准则,按照某种准则将该待识别的表情图像划分到某种表情中。距离判别法是最简单的分类标准。常用的距离度量函数有以下几种:

1. L_1 范式

L_1 范式也称为和范式,将像素间的绝对值的差值相加, L_1 范式距离表示为:

$$L_1(x, y) = \sum_{i=1}^N |x_i - y_i| \quad (4-1)$$

2. L_2 范式

L_2 范式又称为欧氏距离,将像素的差的平方求和再开根号。 L_2 范式距离表示为:

$$L_2(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (4-2)$$

3. 最小距离

最小距离先计算各类训练样本的平均值,测试样本与训练样本平均值像素的差的平方求和再开根号。样本 X 与第 i 类的距离定义为:

$$d_i(x, y) = \sqrt{(X - \bar{X}_i)^T (X - \bar{X}_i)} \quad (4-3)$$

其中 $\bar{X}_i$ 为第 i 类的样本均值, T 为向量的转置。

4.1.2 中心最近邻分类器

确定了适当的距离度量后, 就要选择合适的距离分类器。常用的距离分类器有:

1. 最近邻分类器: 最常采用的分类器, 即把待识别的样本归为与之最近的已知样本的类别中。

2. k -近邻分类器: 分类器取待识别样本的 k 个近邻, 进行投票, k 个近邻中属于哪一类的已知样本最多, 就将待识别样本归到该类中。

3. 中心最近邻分类器: 首先求出已知样本的类中心, 再将待识别样本归到与已知类别类中心最近的类中。

中心近邻距离^[54](Center-based Nearest Neighbor, CNN)是定义了一条中心线(Center-based Line, CL), CL 连接的是一个已标记样本点和此样本点所在类的类中心。类 c 的类中心 o^c 可由下式计算得到:

$$o^c = \frac{\sum_{i=1}^{N_c} x_i^c}{N_c} \quad (4-4)$$

其中 N_c 为类 c 的样本数, x_i^c 为类 c 的某训练样本。

中心线 $\overline{x_i^c o^c}$ 是一条穿过 x_i^c 和 o^c 的直线, 这条线捕捉到 x_i^c 和 o^c 之间的交互信息, 以实现更好的分类效果。从 x 到中心线 CL 的距离定义为:

$$d(x, \overline{x_i^c o^c}) = \|x - p^{c,i}\| \quad (4-5)$$

其中, $p^{c,i}$ 是 x 在 CL 上的投影, $\|\cdot\|$ 是欧氏距离。

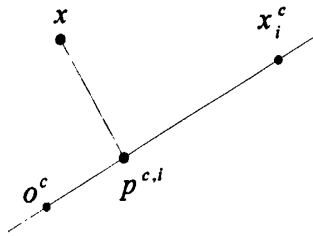


图 4.1 中心线 FL 和 CNN 距离 $d(x, \overline{x_i^c o^c})$

投影点 $p^{c,i}$ 根据下式计算得到:

$$p^{c,i} = x_i^c + \lambda(o^c - x_i^c) \quad (4-6)$$

其中 $\lambda \in \mathbb{R}$, 被称为位置因子, 可由式(4-7)求出:

$$\lambda = \frac{(x - x_i^c)^T (o^c - x_i^c)}{(o^c - x_i^c)^T (o^c - x_i^c)} \quad (4-7)$$

T 为向量的转置。那么 CNN 距离定义如下:

$$d_{CNN}^c = \min_{1 \leq i \leq N} d(x, x_i^c o^c) \quad (4-8)$$

根据最近邻准则, 中心最近邻决策准则为: 如果 $d_{CNN}^k = \min_{c=1,2,\dots,C} d_{CNN}^c$, 则测试样本 x 被归为第 k 类中, $k \in \{1, 2, \dots, C\}$ 。

中心最近邻分类器可定义为: 样本集合 $X = \{x_1, x_2, \dots, x_i, \dots, x_N\}$, 其中每个样本 x_i 所属类别已知, 共 C 类。由式(4-4)可以算出第 c 类的类中心, 并且可得到 N_c 条中心线, 测试样本 x 到 N_c 条 CL 的距离由式(4-5)得到, 由式(4-8)求得这些距离中的最小值, 可得到 C 个 d_{CNN}^c ($c=1, 2, \dots, C$), 最后利用中心最近邻决策准则找出最小的 d_{CNN}^c , 那么测试样本 x 被归为最小的 d_{CNN}^c 类中。

4.2 基于二对二支持向量机的表情分类

4.2.1 多分类问题的描述

SVM 是针对两类而提出二分类方法, 但在实际中往往需要解决多类别的分类问题。目前, 多分类 SVM 方法可分为两大类: 一类是一次性求解方法, 是直接考虑多类别, 方程求解要同时考虑所有样本和类别, 在解决和应用上过于困难和复杂; 另外一类是通过组合一系列二分类支持向量机分类器, 将多类别问题转化为多个二分类问题。后者在训练速度和分类准确率上都优于前者, 更适合于实际应用^[55]。因此, 目前主要的多类别分类算法都是将多类别问题转化为多个二分类问题。

对于多类别分类问题, 根据给定的样本训练集 (x_i, y_i) , $i=1, 2, \dots, n$ (n 为样本数), $x_i \in X \subseteq R^n$ 为样本, $y_i \in Y = \{1, 2, \dots, l\}$ 为类别标号, 每个训练样本分别属于 l 个类别中的一个。

$$T = \{(x_1, y_1), \dots, (x_n, y_n)\} \in (X, Y) \quad (4-9)$$

构造一个决策函数:

$$f(x): X = R^n \rightarrow Y \quad (4-10)$$

对于给定的训练样本, 该决策函数能正确地预测出所属类别。可见, 求解多分类问题, 实际是寻找一个能将分布在 R^n 上的点分成 l 部分的规则。

多类别样本分类问题和两类别样本分类问题有一定的关联性。如果 l 类别样本分类可分, 那么 l 类中任意两个类别可分; 反之, 在 l 类可分问题中, 如果其中任意两两可分, 那么可通过一定的组合法则, 能将 l 类别可分问题转化为两两可分问题。所以, 可通过多个二分类 SVM 按照一定组合法则来解决多类别分类问题。

4.2.2 多类别分类问题的解决方法

目前, 在解决多分类问题时主要是通过将多分类问题转化成多个二分类 SVM

的组合进行处理。根据所采用的组合策略不同,研究者已提出许多解决多分类问题的方法,主要有:一对一方法^[56]、有向无环图方法^[57]、二叉树法^[58]和一对余方法等。

1. 一对一方法

一对一(One-versus-one)方法是在各个类别之间构造分类器,每次从多个类别中挑选出两个不同类别,对这两类进行训练构造两类分类器。对 l 类样本共要构造 $C_l^2 = l(l-1)/2$ 个分类器,组合这些两类分类器并进行投票,得票最多的类即为样本点所属的类。

一对一方法虽然在训练过程中构造 $l(l-1)/2$ 个分类器,但是每个分类器只考虑两类样本,单个分类器的决策边界比一对余算法简单,容易训练,速度快。一对一方法也存在一些缺陷:1. 对未知样本做 $l(l-1)/2$ 次决策,当类别数 l 增加时会影响分类速度;2. 若单个二分类器不规范,会导致整个 l 类分类器趋于过学习;3. 投票可能出现多个类别票数相同的情况,因而可能存在误分,影响分类精度。

2. 有向无环图方法

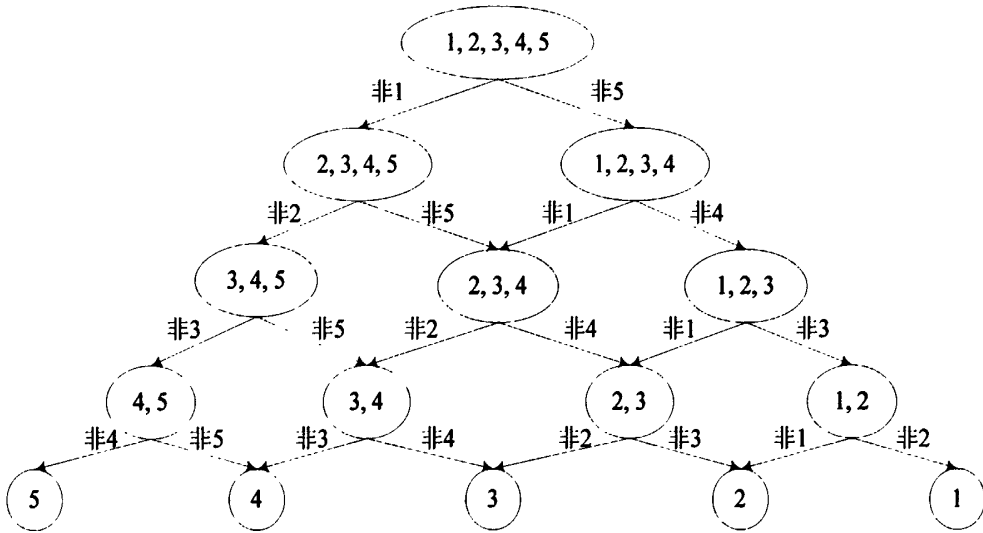


图 4.2 采用 DAG 分类的决策过程

有向无环图(Directed Acyclic Graph, DAG)方法是针对一对一方法的缺陷提出来的。该方法在训练阶段和一对一方法相同,对 l 类样本也要构造 $l(l-1)/2$ 个分类器。但在分类阶段,该方法构造一种两向有向无环图分类器实现多分类,包括 $l(l-1)/2$ 个内部结点和 l 个叶子结点,每个内部结点都是一个二分类的分类器,叶子结点是最终的类别。对待识别样本分类时,首先从根结点开始,根据分类器的输出决定待识别样本该送入左结点或右结点进行下一步分类,直至叶子结点,该叶子结点表示的类为待识别样本所属的类。分类时只需 $l-1$ 个决策函数即可完成分类,与一对一 SVM 分类相比,在分类过程中减少了重复操作,大大提高了分类速度,而且分类精度比一般二叉树方法高。这类方法的缺点是未考虑分类错误传递

对后续产生影响, 并且当类别数较多时训练时间较长, 计算量大。

3. 二叉树方法

二叉树法(Binary Tree, BT)将 SVM 和二叉树相结合, 该方法将所有类别分为两个子类, 每个子类又划分为两个次级子类, 一直划分, 直到划分出最终类别, 每次划分后两类分类问题的规模逐级减小。每个决策点用 SVM 实现分类, 对于 l 类问题, 需要构造 $l-1$ 个分类器。二叉树法根据其分类时的结构可分为两种: 偏态二叉树和正态二叉树。如图 4.3 所示:

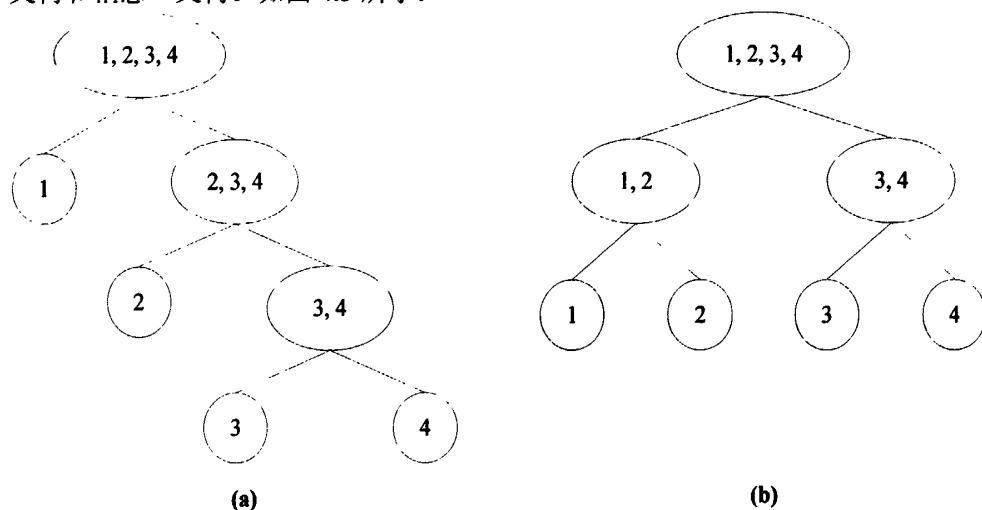


图 4.3 (a) 偏态二叉树 (b) 正态二叉树

二叉树法分类时不存在不可分区域, 不需要遍历所有的分类器, 对于分类器来说按照树的层次由上向下, 分类器中支持向量数逐层减少, 训练分类器时所需要的训练样本较少, 训练时间少, 有较高的分类效率。但是二叉树的结构对多分类的分类精度影响较大, 对于不同结构的二叉树, 会得到不同的分类模型, 即使结构相同, 若类别被判别出的顺序不同, 各处分类器精度不同, 也会导致最终分类结果不同; 同时二叉树分类结构存在误差积累问题。

4. 一对余方法

一对余方法(One-versus-the-rest), 或称为一对多方法, 对于 l 分类问题, 一对余方法将其转化为 l 个二分类问题, 构造 l 个二分类子分类器。在构造第 i 个支持向量机子分类器时, 将属于第 i 类别的样本数据标记为正类, 不属于 i 类别的样本数据标记为负类。测试时, 对测试数据分别计算各个分类器的决策函数值, 并选取最大函数值对应的类别为测试数据的类别。

一对余方法简单有效, 在测试阶段仅使用 l 个分类器, 分类速度快, 可用于大规模的数据。但是由于每次构造分类器时要将整个样本集作为训练样本, 当样本集过大时, 会影响训练速度; 同时当类别数较多时, 某一类的训练样本将大大小于其它类别训练样本的总和, 这种训练样本间的不平衡性将对其分类精度产生影

响。

5. 二对二方法

二对二方法是以四类别分类问题为研究基础，通过较少数量子分类器的组合，利用每个子分类器的输出结果进行综合判断，实现多分类^[59]。

四类别分类时用二对二方法只需构造两个子分类器，构造第一个子分类器时，取四类中的第一类、第三类样本作为正样本，其余两类作为负样本，正样本标记为 1，负样本标记为-1。构造第二个分类器时，将第一类、第二类作为正样本，其余两类为负样本，正样本标记为 1，负样本标记为-1。分类器最终的输出可以得到四种结果：1 和 1，1 和-1，-1 和 1，-1 和-1。

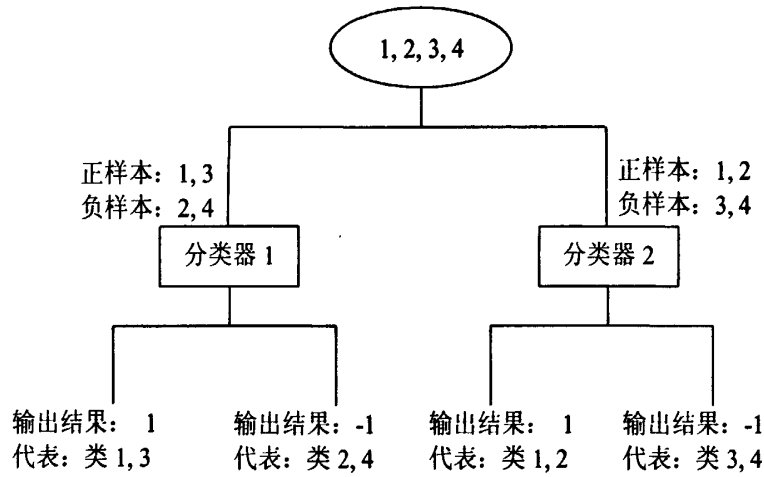


图 4.4 四类别二对二法决策方案

这四种不同的结果分别对应着不同的类别，对应关系如表 4.1 所示：

表 4.1 决策结果分析

分类器 1	分类器 2	分类结果
1	1	1
-1	1	2
1	-1	3
-1	-1	4

二对二多类分类法充分利用了每个子类分类器的识别结果，将 2 个子分类器组合在一起，解决了四类别的分类问题；如果采用一对一方法需要 15 个子分类器，一对多方法则需要 6 个子分类器，决策有向无环图法也需要 6 个子分类器。由于多类分类器的训练时间由子分类器的个数和子分类器的训练时间共同决定，所以该方法有提高速度的优势。

二对二法构造出 2 个子分类器就可以识别出 4 类，可以将其扩展到解决 $4 \times n$ 种类别的分类问题。比如：8 类别分类问题，用集合表示为 {1,2,3,4,5,6,7,8}。首先

将其组合为四类别分类问题,表示为 $\{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}\}$ 。先构造 2 个子分类器可以得到 4 个子类,表示为 $\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}$ 。这是 4 个二分类问题,再构造 4 个子分类器就可以实现分类,总共用了 6 个分类器。如果采用一对多方法需构造 8 个分类器,一对一方法需构造 28 个分类器。

二对二方法也可以解决 4^n 个类别的分类问题,先将 4^n 个类别分解为 $\{4^{n-1}, 4^{n-1}, 4^{n-1}, 4^{n-1}\}$ 4 个子类,构造两个子分类器就可以识别出这四类,再以此方法将每个 4^{n-1} 继续分解,直至识别出每一类。需要构建的子分类器数目 N 和类别数之间的关系为:

$$N = 2 + 2 \times 4 + \cdots + 2 \times 4^{n-1} \quad (4-11)$$

其中 $n=1,2,\dots$ 为自然数。

二对二方法在分类时,对未知样本首先从根结点开始,根据分类结果选择使用下层左结点或右结点继续分类,直到叶子结点,该叶子结点所表示的类即为未知样本所属的类。二对二方法有以下优点:1. 所需训练两类支持向量机的数目少,提高了训练速度;2. 测试时,不需要计算所有的分类判别函数,而且问题规模逐级下降,可节省测试时间;3. 消除了决策时存在测试样本同时属于多类或不属于任何一类的情况。但是,二对二方法也存在误差积累的缺点。所以,如何设计合理的决策方案也是该方法的一个关键。

4.2.3 基于二对二支持向量机的表情分类

目前,表情分类还主要是对六种基本表情进行分类:愤怒(angry)、高兴(happy)、吃惊(surprise)、厌恶(disgust)、恐惧(fear)、悲伤(sad)。因为二对二方法在解决实际多分类问题具有明显的优越性,本文采用二对二方法对六种基本表情进行分类识别。

二对二多分类方法存在误差积累现象,即分类误差在越靠近根结点的地方,其误差的积累效应越明显,分类性能就越差。为了避免这种问题,本文考虑在决策根结点用近似最优方法将多类别样本分成两组,并使类聚中心距离最大,且每组样本数据分歧最小。根据六种基本表情识别率不同,设计出较为合理的决策方案,获取最优分类性能。

六种基本表情属于非 4^n 种类别的分类,所以先将其分为近似的 4 类,然后再用二对二方法分类。从第四章中六种基本表情的识别结果可以发现高兴、恐惧比其它表情识别率高,所以本章设计二对二多分类 SVM 时在根结点挑选高兴和恐惧为正样本,愤怒、悲伤、厌恶和惊讶为负样本,使根结点处的分类错误最小,减小误差积累。六种基本表情的决策方案如下图所示:

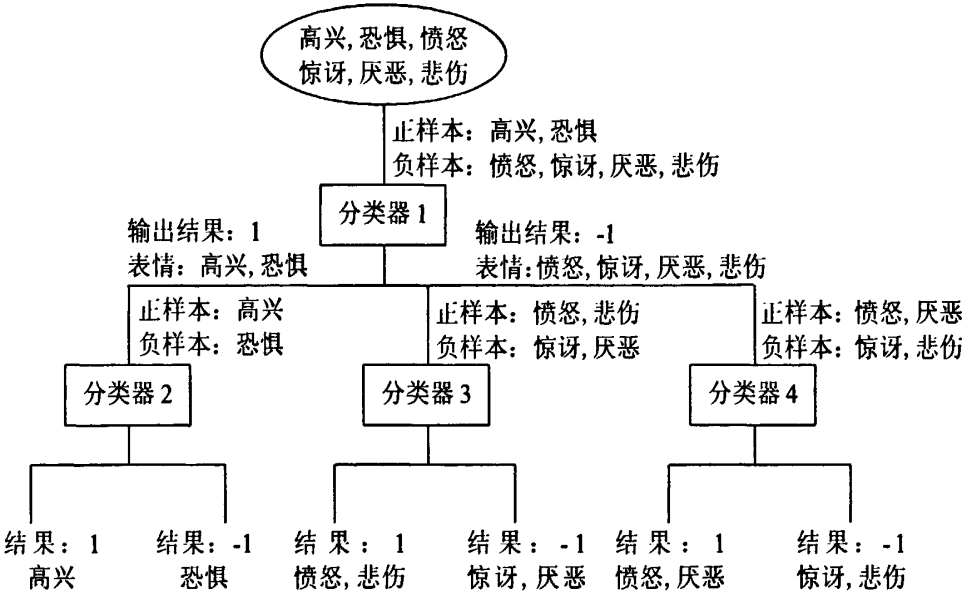


图 4.5 六种基本表情二对二法决策方案

分类器最终输出结果与分类表情的对应关系为表 4.2 所示：

表 4.2 六种基本表情决策结果分析

分类器 1	分类器 2	分类器 3	分类器 4	分类结果
1	1			高兴
1	-1			恐惧
-1		1	1	愤怒
-1		1	-1	悲伤
-1		-1	1	厌恶
-1		-1	-1	惊讶

二对二法对六种基本表情进行训练和测试的主要步骤如下：

1. 二对二 SVM 分类器的训练

①给定训练样本集 $X = \{x^1, x^2, ..., x^C\} (C = 6)$ 共六种基本表情, 对样本集预处理。

②用 CBP 算法对人脸表情图像进行特征提取。

③将表情特征向量转换成 libSVM^[60]中所要求的数据格式, 选用 RBF 核函数, 采用交叉验证选择四个分类器最佳的惩罚因子 c 和高斯核函数 λ , 利用 libSVM 对样本特征进行训练, 最终训练出四个二对二 SVM 子分类器。

2. 样本测试

①对测试人脸表情图片进行预处理。

②提取人脸表情图像的 CBP 特征。

③将表情特征向量转换成 libSVM 中所要求的数据格式。

④将转换后的数据送入训练好的四个 SVM 分类器中, 根据分类器的输出将测

试样本分类。

4.3 实验结果及分析

实验采用 JAFFE 人脸表情数据库对 CNN 算法和二对二多分类 SVM 方法(简称为 2to2SVM)进行测试。

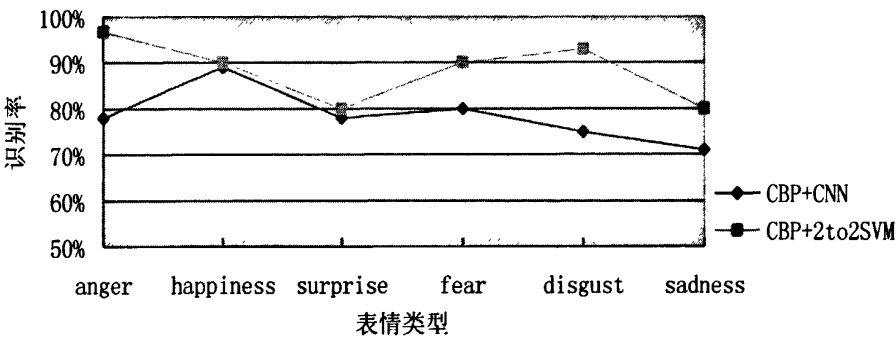


图 4.6 CBP+CNN 和 CBP+2to2SVM 在 JAFFE 库上的识别率

表 4.3 本文所用算法与其他算法的比较(同为 JAFFE 库)

方法	平均识别率	识别一个样本所用时间(秒)
HOLA 特征+Fisher 判别 ^[61]	69.4%	无
LBP+由粗到精分类 ^[62]	77.0%	无
改进 LBP ^[63]	81.0%	无
Gabor+LDA ^[64]	90.6%	无
LBP+LDA ^[15]	89.6%	0.0070
CBP+CNN	83.6%	0.0107
CBP+2to2SVM	88.3%	0.0066

由图 4.6 和表 4.3 可知，CBP+2to2SVM 比 CBP+CNN 方法的识别率高，并且在识别速度上 CBP+2to2SVM 较快，一方面是因为 SVM 在解决小样本问题上具有优势，另外一方面是二对二多分类方法的使用大大较少了分类器数量。CBP+2to2SVM 在识别率上比 LBP+LDA 方法稍低，但是在速度上却比 LBP+LDA 方法快，可根据实际应用中 对识别率和速度两个方面不同的要求，选择不同的算法。CBP+2to2SVM 方法在某些表情如愤怒、厌恶有较好的识别效果，而有些表情则很难被识别如悲伤，因此导致了总体识别率不高。

本文将 2to2SVM 与其他多分类方法也做了对比，采用 CBP 算子提取人脸表情特征，分别用一对一、一对余、DAG、偏态二叉树、正态二叉树和本文 2to2SVM 分类方法进行表情分类。实验结果如表 4.4 所示：

表 4.4 2to2SVM 与其他多分类方法的比较(同为 JAFFE 库)

方法	平均识别率	识别一个样本所用时间(秒)
一对一	84.5%	0.0081
一对余	83.2%	0.0092
DAG	86.9%	0.0070
偏态二叉树	85.2%	0.0071
正态二叉树	86.0%	0.0068
CBP+2to2SVM	88.3%	0.0066

从实验结果可以看出，本文算法识别率和识别速度都优于其它多分类方法，这主要是因为采用二对二的方法减少了子分类器的数目，所产生的分类误差与误差积累相对较小，所以本文 2to2SVM 分类算法有着明显的优势。

4.4 小结

本章详细介绍了 CNN 算法和几种常用的多分类 SVM 方法，重点介绍了 CNN 算法和 SVM 中的二对二多分类方法。CNN 不仅能解决小样本问题，而且能减小最近邻特征线的计算量，能提高其训练和检测速度。二对二 SVM 多分类方法以四类别分类为基础，根据每个子分类器的输出判断所属类别，与一对一，一对余方法相比，子分类器的数目少，训练和识别速度快，同时减小了分类误差和误差积累的影响。二对二多分类方法虽然子分类器有四个，但是与 CNN 方法相比，识别一幅图片所用时间短、识别率高，可见二对二多分类方法在人脸表情识别中有较大的优势。

第五章 总结与展望

5.1 总结

人脸表情识别是机器理解人类情感的重要前提,是实现智能机器的关键技术之一,也是人们探索人工智能、理解机器智能的途径之一,虽然人脸表情识别的相关研究在最近几年里取得了较好的成绩,但是要真正达到机器智能化,让机器像人那样理解出人类的表情还需要进一步深入研究。

本文主要研究了人脸表情识别中三个核心问题,即人眼定位、人脸表情的特征提取和表情分类。本文主要做了如下工作:

1. 总结和归纳了近几年国内外人脸表情识别的现状,介绍了表情识别中特征提取和表情分类的常用方法,以及对部分人脸表情数据库做了介绍。

2. 研究了人眼定位问题。本文根据眼睛的结构特征以及眼睛区域在边缘特征提取后的图像特点,提出了基于 Adaboost 算法和多掩模的人眼精确定位算法,实验证明该算法在人眼精确定位中有较好的效果。

3. 深入研究了人脸表情识别中的特征提取问题。由于 LBP 算法存在三大缺陷,本文采用 CBP 算法提取人脸表情特征,实验表明,CBP 算法在提取人脸表情特征方面有较快的速度和较强的抗噪性。

4. 详细分析了中心最近邻分类器和常用的多分类方法,对多分类问题中的二对二多分类方法作了重点介绍。本文对中心最近邻分类器和二对二多分类器都做了尝试,并做了对比实验。实验结果表明,在人脸表情识别中二对二多分类器性能优于中心最近邻分类器和其它多分类 SVM 分类器。

5.2 展望

人脸表情识别是一个跨学科、富有挑战性的前沿课题,涵盖内容十分广泛。本文只是对其若干关键问题进行了研究,现在所做的一些工作仅仅是初步的,还存在许多需要解决的问题:

1. 现有的人脸表情识别算法难以满足现实生活中的应用需求。首先,训练样本不够丰富,对于不同国家,不同民族、不同性别和不同年龄的训练样本的表现不尽人意;其次,一般表情识别中仅将表情划分为六种基本表情,但是在实际中,人脸面部表情相当丰富,除此六种基本表情之外的微表情难以有效地区分。

2. 本文将 Adaboost 和多类模板结合应用进行了一定的考虑,该算法在光照均匀、无遮挡物的正面人脸表情数据库上效果不错,但是在眼睛区域光照变化较大、存在遮挡物的情况下定位效果不佳。因此,如何在眼睛区域光照变化较大、

存在遮挡物的情况下精确定位眼睛非常值得研究。

3. 本文采用分块加权 CBP 算法对人脸表情图像特征提取, 该算法与 LBP 算法相比降低了特征向量的维数、将中心像素点的作用考虑在内并且对噪声不敏感, 但是其特征向量的维数还是对表情分类的分类时间有较大影响。因此, 如何更加有效地减少数据冗余, 从表情特征向量中提取有用信息还需要深入研究。

4. 本文利用二对二多分类 SVM 方法对六种表情进行分类, 与一对一方法、一对余方法相比, 该方法减少了子分类器数量, 缩短了训练和分类时间, 但是当要求对更多类、更细微表情进行分类时, 如何根据表情特征的特点, 设计出合理的分类器, 减少积累误差, 提高分类器速度, 是一项十分值得研究的内容。

致谢

值此论文完稿之际，首先衷心感谢我的导师卢朝阳教授。在卢老师的严格要求和指导下，我愉快而充实度过了两年半的研究生生活。卢老师广博的学识、严谨的治学作风、敏锐的学科洞察力以及诚恳待人、无私奉献的高尚风格对我的影响十分深远，这将使我受益终身。在此，谨向卢老师表示诚挚的敬意，祝愿卢老师身体健康，工作顺利！

还要感谢李静老师，李老师平易近人、治学严谨，不厌其烦的指出我研究中的缺失，对此表示衷心的感谢。

感谢实验室李远征博士在工作和生活上对我的帮助，同时感谢刘卫、姜维、李益红、姚超、刘晓佩、崔玲玲、许亚美、温浩等博士的关心和支持，能和他们在一起工作是我的荣幸。

感谢我的同学汪文芳、谭福秀、孙华凯、王炜、柴进、郁映卓、张宇、王晓明、李乔等，感谢严新平师妹帮我整理实验数据，感谢王新民师弟的关心和支持，与他们一起学习和生活的日子，将是我人生中最美好的回忆；同时感谢魏茜、魏韡、张晶等人，她们既是我的挚友也是我的亲友，在我考研时期她们给我以极大的支持和鼓励，谢谢你们！

感谢我的父母一直以来给予我默默地关怀和支持。

特别感谢我的女朋友李芳，远方的她在我背后默默给予的支持是我前进的动力，是她的包容和体谅，才使我得以顺利完成学业。

参考文献

- [1] Picard R. W.. Affective computing. MIT Press. 1997.
- [2] Suwa M., Sugie N., and Fujimora K.. A preliminary note on pattern recognition of human emotional expression. Proceedings of the Fourth International Joint Conference on Pattern Recognition. 1978. 408-410.
- [3] Fasel B., Luetttin J.. Automatic facial expression analysis: a survey. Pattern Recognition. 2003. 36(1): 259-275.
- [4] Ekman P., Friesen W. V.. UNMASKING THE FACE: A guide to recognizing emotions from facial clues. Prentice-Hall, Inc. 1975.
- [5] Jin Hui, Gao Wen. The human facial combined expression recognition system. Chinese Journal of Computer, 2000. 23(6): 602—608.
- [6] Gianluca Donato, Marian Stewart Bartlett, Joseph C. Hager, et al. Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999. 21(10): 974-989.
- [7] Fasel B., Luetttin J.. Automatic facial expression analysis: a survey. Pattern Recognition. 2003. 36(1): 259-275.
- [8] Pantic M., Rothkrantz L. J. M.. Automatic analysis of facial expression: the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000. 22(12): 1424-1446.
- [9] 薛雨丽, 毛峡, 郭叶等. 人机交互中的人脸表情识别研究进展. 计算机工程与应用. 2009. 14(5): 764-772.
- [10] Andrew J., Calder A., Burton M., et al. A principal component analysis of facial expression. Vision research. 2001. 1(9): 1179-1208.
- [11] 程剑, 应自炉. 基于二维主分量分析的脸部表情识别. 计算机工程与应用. 2006. 42(5): 32-33.
- [12] Bartlett M. S., Hades H. M., Sejnowski T. J.. Independent component representations for face recognition. Human Vision and Electronic Imaging. 1998. 528-539.
- [13] Ahonen T, Hadid A, Pietikäinen M. Face description with local binary patterns: Application to face recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence. 2006. 28(12): 2037-2041.
- [14] Shan C., Gong S., McOwan P. W. Robust facial expression recognition using

- local binary patters. IEEE International Conference on Image Processing. 2005. 370-373.
- [15] 施兴华. 自动人脸表情识别. 西安电子科技大学硕士论文. 2009. 21-44.
- [16] 付晓峰, 韦巍. 基于多尺度中心化二值模式的人脸表情识别. 控制理论与应用. 2009. 26(6):629-633.
- [17] Lee T.. Image representation using 2D Gabor wavelets. IEEE Transaction Pattern Analysis and Machine Intelligence. 1996. 18(10): 959-971.
- [18] Sun Xinghua, Xu Hongxia, Zhao Chunxia. Facial expression recognition based on histogram sequence of local Gabor binary patterns. IEEE International Conference on Computational Intelligence and Security. 2008. 158-163.
- [19] Cootes T., Edwards G., Taylor C.. Active appearance models. The 5th European Conference on Computer Vision. 1998. 615-625.
- [20] Horn B. K. P., Schunck B. G.. Determining Optical Flow. Artificial Intelligence.1981:185-203.
- [21] Wildes R. P., Michael J., et al. Recovering estimates fluid flow from image sequence data. Computer Vision and Image Understanding. 2008. 246-266.
- [22] Béréziat D., Herlin I., Younes L.. A generalized optical flow constraint and its physical interpretation. Proceedings of Computer Vision and Pattern Recognition. 2000. 487-492.
- [23] Tian Y., Kanade T., Cohn J. F.. Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001. 23(2): 97-115.
- [24] Ma L., Khorasani K. Facial Expression recognition using constructive feedforward neural networks. IEEE Transaction on Systems Man and Cybernetics: Part B. 2004. 34(3): 1588-1595.
- [25] Cover T., Hart P.. Nearest neighbor pattern classification. IEEE Transactions in Information Theory. 1967. 13(1): 21-27.
- [26] Corinna Cortes, Vladimir Vapnik. Support vector networks. Machine Learning. 1995. 20(3): 273-297.
- [27] Tsai, Hung Hsu, Lai Yenshou, Zhang Yicheng. Using SVM to design facial expression recognition for shape and texture features. International Conference on Machine Learning and Cybernetics. 2010. 5: 2697-2704.
- [28] Gunes T., Polat E.. Feature selection for multi-SVM classifiers in facial expression classification. International Symposium on Computer and Information Sciences. 2008. 1-5.

- [29] Zhao Yisu, Shen Xiaojun, Georganas N.D.. Facial expression recognition by applying multi-step integral projection and SVMs. IEEE Instrumentation and Measurement Technology Conference. 2009. 686-691.
- [30] Baum L. E., Petrie T.. Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*. 1966. 37: 1554-1563.
- [31] Baum L. E., Eagon J. A.. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for recology. *Bull Amer. Math. Stat.* 1967. 37(2): 360-363.
- [32] Baum L. E., An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov process. *Inequalities*. 1972. 3(1): 1-8.
- [33] Sakaguchi T., Morishima S.. Face feature extraction from spatial frequency for dynamic expression recognition. *The International Conference on Pattern Recognition*. 1996: 451-455.
- [34] 杨国亮, 于仲安. 基于改进光流算法和 HMM 的面部表情分析. *微计算机信息*. 2008. 1: 260, 284-285.
- [35] Viola P., Jones M.. Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*. 2001: 511-518.
- [36] 章毓晋. 图像工程(上册)-图像处理和分析. 清华大学出版社. 1999. 65-66.
- [37] Pageorgiou C., Oren M., Poggio T.. A general framework for object detection. *International Conference on Computer Vision*. 1998. 555-562.
- [38] Oren M., Pageorgiou C., Poggio T.. Example-based object detection in images by components. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 2001. 23(4): 349-361.
- [39] Freund Y., Schapire R. E.. A decision-theoretic generalization of online learning and an application to boosting. *The Second European Conference on Computational Learning Theory*. 1995: 23-37.
- [40] Viola P., Jones M.. Robust real-time object detection. *The Second International Workshop on Statistical and Computational Theories of Vision: Modeling, Learning, Computing, and Sampling*. 2001.
- [41] Reinders M. J. T., Koch R. W. C., Gerbrands J. J.. Location facial features in image sequences using neural networks. *The Second International Conference on Automatic Face and Gesture Recognition*. 1996. 16: 230-235.
- [42] Huang J., Harry.. Eye location using genetic algorithms. *The Second International*

- Conference on Audio and Video-Based Biometric Person Authentication. 1996. 16: 248-252.
- [43] Feng G. C., Yuen P. C.. Variance projection function and its application to eye detection for human face recognition. *Pattern Recognition Letters*. 1998. 19(9): 899-906.
- [44] 戴景文, 刘丹, 苏剑波. 基于投影峰的眼睛快速定位方法. *模式识别与人工智能*. 2009. 22(4): 605-609.
- [45] Kith V., El-Sharkawy M., Bergeson-Dana T., et al. A feature and appearance based method for eye detection on gray intensity face images. *International Conference on Computer Engineering and Systems*. 2008. 25: 41-47.
- [46] Cheng Ziguo, Liu Yuncai. Efficient technique for ellipse detection using restricted randomized Hough transform. *Proceedings of the International Conference on Information Technology: Coding and Computing*. 2004. 2: 714-718.
- [47] 陈海峰, 雷华, 孔燕波等. 基于最小二乘法的改进的随机椭圆检测算法. *浙江大学学报(工学报)*. 2008. 42(8): 1360-1364.
- [48] Andrew Fitzgibbon, Maurizio Pilu, Robert B. Fisher.. Direct least square fitting of ellipses. *Tran Analysis and Machine Intelligence*. 1999. 21(5): 476-480.
- [49] 王磊, 邹北骥, 彭小宁等. 一种改进的提取人脸面部特征点的 AAM 拟合算法. *电子学报*. 2006. 34(8): 1424-1427.
- [50] Ojala T., Pietikiäinen T.. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002. 24(7): 971-987.
- [51] Ahonen T., Hadid A., Pietikiäinen M. Face description with local binary pattern: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006. 28(12): 2037-2041.
- [52] Fu Xiaofeng, Wei wei. Centralized binary patterns embedded with image Euclidean distance for facial expression recognition. *The Fourth International Conference on Natural Computation*. 2008. 4: 115-119.
- [53] Li S. Z., Lu J.. Face recognition using the nearest feature line method. *IEEE Transaction on Neural Networks* 1999. 10: 439-443.
- [54] Gao Qingbin, Wang Zhengzhi. Center-based nearest neighbor classifier. *Pattern Recognition*. 2007. 40(1): 346-349.
- [55] Hu Zhonghui, Cai Yunze, He Xing, et al. Fusion of multi-class support vector machines for fault diagnosis. *The American Control Conference*. 2005: 1941-1945.

- [56] Krebel U.. Pairwise classification and support vector machines. *Advances in Kernel Methods: Support Vector Learning*. 1999: 255-268.
- [57] Platt J. C., Cristianini N., Shawe-Taylor J.. Large margin DAGs for multi-class classification. *Advances in Neural Information Processing Systems*. 2000. 12: 541-553.
- [58] 孟媛媛, 刘希玉. 一种新的基于二叉树的 SVM 多分类方法. *计算应用*. 2005. 25(11): 2653-2657.
- [59] 于清, 赵晖. 一种 2_a_2 支持向量机多类分类新方法. *计算机工程与应用*. 2008. 44(25): 186-188.
- [60] Chang C. C., Lin C. J.. LIBSVM--A Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>. Available at 2010-12-24.
- [61] Shinohara Y, Otsu N. Facial expression recognition using fisher weight maps. *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*. 2004. 499-504.
- [62] Feng Xiaoyi. Facial Expression Recognition Based on Local Binary Patterns and Coarse-to-Fine Classification. *International Conference on Computer and Information Technology (CIT'04)*. 2004.
- [63] He Lianghua, Zou Cairong. An Enhanced LBP Feature Based on Facial Expression Recognition. *The 27th International Conference of the IEEE Engineer in Medicine and Biology Society*. 2006.3300-3303.
- [64] 张一鸣. 人脸表情识别. 大连理工大学硕士论文. 2005. 56-77.

