

陕西师范大学

硕士学位论文

人脸表情识别研究

姓名：李文

申请学位级别：硕士

专业：计算机软件与理论

指导教师：王映辉

20070501

人脸表情识别研究

李文

摘要：在人与人的交往中，面部表情是一种非常重要的信息传递方式，能够传达很多语言所不能传达的信息。人脸表情识别就是利用计算机对人脸的表情信息进行特征提取并分类的过程，它使得计算机能够根据人的表情信息，推断人的心理状态，从而实现人机之间的智能交互。一般而言，表情识别系统主要有四个基本部分组成：表情图像获取、表情图像预处理、表情特征提取和表情分类识别。本文重点集中在对表情特征提取、表情分类等关键问题的研究上，研究工作可概括为：

(1)概括了人脸表情识别的研究意义和应用前景，回顾了国内外研究现状，对现有的多种表情识别方法进行了综述。

(2)对表情人脸图像进行了预处理。实验数据采用了日本 JAFFE 表情库，在进行表情识别时，首先对表情图像进行亚采样预处理来降低图像的维数，并将图像每个像素灰度值归一化至 $[0, 1]$ 间，利于特征提取和分类；最后对表情图像进行直方图均衡化，经过均衡化后图像的细节更加清楚，直方图各灰度等级的分布更加平均。

(3)采用主成分分析方法(PCA)进行特征提取。在特征提取之后，采用欧氏距离作为模式特征之间的相似性度量，用最近邻分类器进行分类，取得了较好的识别效果，并将核最近邻分类器(Kernel Nearest Neighbor)应用于表情识别，通过实验得出了核最近邻分类效果优于最近邻分类器的结论。

(4)采用支持向量机(SVM)的分类方法，利用 one-against-rest 组合原理，构建了 7 个支持向量机并对之进行组合，然后用组合后的 SVM 进行表情识别。并在此基础上对原有 SVM 进行了改进，对 7 个 SVM 不再共用同一个核参数，而是根据学习样本自适应地估计各个 SVM 的核参数，识别率由原来 94.761905%提高到了 95.238095%。

(5)对基于最近邻分类器的表情识别方法和基于 SVM 表情识别方法进行了分析与实验比较。在要求快速实时分类和分类正确率要求不高的场合，可考虑采用基于最近邻分类器的表情识别；而在对分类准确率要求较高，但对分类速度不作要求或要求较低场合，可考虑采用基于 SVM 的表情识别方法。

关键词：表情识别；主成分分析；最近邻分类器；核；支持向量机

Research of Human Facial Expression Recognition

Li Wen

Abstract : Facial Expression plays a very important role in information transmission when people communicate with each other. It succeeds in transmitting plenty of information people can not express by languages. Human Facial expression recognition is a process of feature extraction and classification, which makes the computer detect the expression state from given expression images and ascertain the subject's specific emotion in order to achieve smarter and more natural inter-action between human beings and computers. Generally, facial recognition system consists of mainly four parts, which are expression image acquisition, expression image preprocessing, expression feature extraction and classification. This paper presents a detailed description of the research on the key issues of expression feature extraction and classification, and the research work can be summarized as:

(1)The significance and application future of research on the Facial Expression Recognition is generalized in this paper. The current domestic and overseas study situation is reviewed and some existing facial recognition methods are summarized.

(2)Facial expression image is preprocessed. The Japanese JAFFE expression database is used as experimental data. Firstly, expression image sub-sampling image preprocessing is done to reduce the dimension, and the image of each pixel gray value belongs to a normalized $[0,1]$ for feature extraction and classification. Finally, the expression image histogram is equalized. After equalization the details of the image get clearer, and the distribution of gray levels of histogram gets evenner.

(3)In the period of feature extraction PCA is adopted, and nearest neighbor classifier is used for classification, using Euclidean distance as the similarity measurement, and we've got a good result. Kernel nearest neighbor is also used for facial recognition and the experimental result show that it is better than nearest neighbor classifier.

(4)Using Support Vector Machine(SVM) classification method and one-against-rest portfolio theory, we construct seven support vector machines, and use the combination of SVM for expression recognition. We improve the original SVM and according to the study sample estimate the nuclear parameters of each SVM adaptively without sharing the same nuclear parameters. Recognition rate improved from

94.761905% to 95.238095%.

(5) Expression recognition method based on the nearest neighbor and expression recognition method based SVM are compared. Expression recognition method Based on the nearest neighbor is adopted when requesting real-time and rapid classification but not high correct classification. Expression recognition method based SVM is adopted when requesting high correct classification but not rapid classification or not caring speed.

Keywords : Facial Recognition; Principal Component Analysis ; Nearest Neighbor Classifier ; Nuclear ; Support Vector Machine

学位论文独创性声明

本人声明所呈交的学位论文是我在导师的指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经注明引用的内容外，论文中不包含其他个人已经发表或撰写过的研究成果，也不包含为获得陕西师范大学或其它教育机构的学位或证书而使用过的材料。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示谢意。

作者签名： 李文 日期： 2007.5.28

学位论文使用授权声明

本人同意研究生在校攻读学位期间论文工作的知识产权单位属陕西师范大学。本人保证毕业离校后，发表本论文或使用本论文成果时署名单位仍为陕西师范大学。学校有权保留学位论文并向国家主管部门或其它指定机构送交论文的电子版和纸质版；有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅；有权将学位论文的内容编入有关数据库进行检索；有权将学位论文的标题和摘要汇编出版。

作者签名： 李文 日期： 2007.5.28

第1章 绪论

心理学家 J.A.Russell 提出,在人们日常的交流中,只有 7%的信息是通过语言来传递的,而 55%的信息则是通过面部表情来传递的^[1],由此可见表情信息在人与人之间交流中的重要性。随着人工智能和模式识别的不断发展,人机智能交互中的一项重要技术——人脸表情识别也越来越受到人们的重视。人脸表情识别就是利用计算机对人脸的表情信息进行特征提取并分类的过程,它使计算机能获知人的表情信息,进而推断人的心理状态,从而实现人机之间的智能交互。表情识别技术是情感计算机研究的内容之一,是生理学、心理学、图像处理、模式识别和机器视觉等领域的一个极富挑战性的交叉课题。

1.1 研究意义及应用前景

人脸面部表情的识别对实现自然的、智能化的人机交互和计算机自动图像理解都有重要的意义。目前人机交互仍停留在基于符号和对我们来说并不自然的行为上。尽管键盘和鼠标对以前的板片式开关和插入式卡片来说,已是不可否认的进步,计算机已不再是瞎子、哑巴或聋子,然而不需要人们进行特殊或深奥训练的、自然、直观的界面仍然是难以捉摸的目标。让计算机具备善解人意的能力和情感识别能力,具有创造性和智能行为,就要使机器具有像人那样的技能,比如识别人脸、表情、语言等,而不是让人具有机器样的技能。

目前,人们研究表情识别的主要目的在于建立和谐而友好的人机交互环境。使得计算机能够看人的脸色行事,从而营造真正和谐的人机环境。

人脸表情识别技术可应用于多个领域:

(1)人机交互:在现今的网络社会里,计算机已经不是仅仅被视为一台用于计算的机器了,它还应该提供各种使人们之间可以相互交流的功能。其中,影像的实时传输就是一个很重要的功能,如果我们和一个朋友在网上聊天的同时还可以看见他的影像,那么交流的效果肯定会更好。但在有些时候,因为流量及速度的限制,影像的传输还是非常缓慢,而且也不是必须的,这时,就可以用摄像机里的表情分析器对用户表情进行分析,然后只传输得到的分析结果,这同样也能够达到很好的交流效果。

(2)安全领域:表情识别可用于核电站的管理和长途汽车司机等着重强调安全的工作岗位。在岗者一旦出现疲劳、瞌睡的征兆,识别系统及时发出警报避免险情发生;还可用于公安机关的办案和反恐中。

(3)智能机器人研制：我国中科院自动化所成功研制能模仿人类表情新一代科普机器人能模仿 40 多种人类的表情动作。在表情的具体实现上，使用了 11 个可以自由控制的节点，分别分布在眉毛、眼皮、眼珠和嘴上。通过 11 个自由节点的自由组合，再通过智能控制，它就能把各种表情展现出来。

(4)医疗领域：表情识别还可用于机器人手术操作和电子护士的护理。可根据患者面部表情变化及时发现其身体状况的变化，避免悲剧发生。

(5)远程教育：有过远程教育经历的老师普遍觉得很难掌握学生对课程的掌握程度，在正常的授课中，老师可以通过学生们的表情来得知他们掌握的怎么样，这就也需要一种表情识别器把学生们的表情规定为对课程的掌握程度并反馈给远程的教师，老师们就可以做出相应的反应了。

(6)电脑游戏：这可能是脸部表情识别和制作最有应用价值的一个领域。如果游戏可以根据游戏者的喜、怒、哀、乐来做出实时的反应，那么这样的游戏肯定比那些传统规定好规则的游戏更加吸引人。

1.2 国内外研究现状

国际上关于面部表情的分析与识别的研究工作可分为基于心理学和计算机两类。早在 1872 年，著名生物学家达尔文^[2]就对人脸表情的相似性和延续性进行了研究，他指出：面部表情不随种族、国家、性别等的不同而不同，并且表情行为可以继承。1971 年美国心理学家 Ekman 和 Friesen^[3]定义了六种基本表情：生气(angry)、厌恶(digest)、害怕(fear)、伤心(sad)、高兴(happy)和吃惊(surprise)，并于 1978 年开发了面部动作编码系统 FACS (Facial Action Coding System) 来检测面部表情的细微变化。系统将人脸划分为若干个运动单元 AU (Action Unit) 来描述面部动作，这些运动单元显示了人脸运动与表情的对应关系。六种基本表情和 FACS 的提出具有里程碑的意义，后来的研究者建立的人脸表情模型大都基于 FACS 系统，绝大多数表情识别系统也都是针对六种表情的识别而设计的。Ekman 和 Roseberg^[4]后来提出的 FACS-AID 系统将每种表情与肌肉的运动对应起来，只需观察肌肉的运动即可判断出表情类别。

计算机方面，1978 年 suwa 和 sugie^[5]根据一段脸部视频动画抽取表情特征与典型模式比较进行表情识别；同年 terzopoulous 和 waters^[6]运用了简化了的 Ekman-Friesen 模型，用计算机产生人脸动画，对人脸视频序列做了表情分析。1991 年 Mase 和 Pentland^[7]利用 8 个方向上跟踪光流特征的变化来检测 FACS 中的运动单元。由于 FACS 中没有时间描述信息，1997 年 Essa^[8]提出基于视频的动态表情描述方式 FACS+，FACS+充分考虑了时间和空间变化特性，可用于动态建模和运

动估计,克服了 FACS 对动态表情特性描述的不足,在进行表情分析时通过光流法来分析视频中的脸部数据。2000 年我国的高文、金辉^[9]提出 FACS' (转换 FACS) 它把运动单元的运动转化为基于物理结构和肌肉模型的运动特征向量序列,对眼部和嘴部分别进行表情编码,相应的运动基于 FACS 规则,同时又克服了它的弱点。鉴于表情识别领域绝大多数是基于二维图像的,2005 年中国科技大学的研究者提出了一种基于三维数据的人脸表情识别方法^[10],给出了基于三维特征的眼角和嘴角新的提取算法。除了理论方法,一些实用性技术也被开发出来,荷兰一座高科技监狱使用表情识别软件监视着罪犯的情绪变化,那些表现良好的囚犯会享受一些特殊优惠待遇;我国东南大学也研制出表情识别系统,它由多个摄像头和处理相关数据的计算机组成,计算机中已存有六种基本表情的相关数据和图片,对同一个人的不同表情识别率可高达 97.65 %。

人脸表情识别经过近几十年的发展目前正逐渐成为一个研究热点。国内外很多机构都在进行这方面的研究,美国的麻省理工学院(MIT)、卡耐基梅隆大学(CMU)机器人研究所、马里兰大学视觉实验室、日本的国际电信技术研究所(ATR)、东京大学等的贡献尤为突出,我国的清华大学、哈尔滨工业大学、中国科技大学、中科院自动化研究所等均作了一定的研究工作。2005 年 10 月在我国召开的首届国际情感计算及智能交互会议,推动了我国表情识别研究的深入研究。

1.3 表情识别的主要方法综述

面部表情是由脸部的肌肉收缩运动引起的,它使眼睛、嘴巴、眉毛等脸部特征发生形变,有时候还会产生皱纹,这种引起人脸暂时形变的特征叫做暂态特征,而处于中性表情状态下的嘴巴、眼睛、鼻子等几何结构、纹理叫做永久特征。人脸表情识别的过程就是将这些暂态特征从永久特征中提取出来,然后进行分析归类。根据处理图像的性质不同,可将表情识别方法分为基于静态图像的表情识别和基于动态图像的表情识别。

从表情识别过程来看,表情识别可分为四部分:表情图像的获取、表情图像预处理、表情特征提取和表情分类识别。



图 1-1 表情识别步骤图

(1)表情图像获取:用于面部表情识别的图像包括静态或者动态图像。因为一个图像序列也描述了表情的时间特征,所以它比一幅静态图像潜在地包含了更多

的信息。

(2)表情图像预处理：图像预处理常常采用信号处理的形式（如去噪、像素位置或者光照变量的标准化），还包括人脸及它的组成的分割、定位或者跟踪。表情的表示对图像中头的平移、尺度变化和旋转是敏感的。为了消除这些不必要的变换的影响，人脸表情图像可以在分类前进行标准化。

(3)表情特征提取：特征提取是将像素数据转化为一个更高水平的人脸及其组成部分的外形、运动、颜色、肌肉和空间结构的表示。提取出的表示被用来进行后续的表情分类。特征提取通常会缩短了输入空间的维数。

(4)表情分类：表情分类由一个分类器执行，该分类器由模式分布模型组成，与判决程序相结合。很多包含参数和非参数的技术的分类器已经被应用于自动表情识别系统中。在面部表情识别中主要的类型标准是由 Ekman 定义的面部表情原型。

由于表情识别与人脸识别在人脸图像的获取和预处理这些环节上是基本一致的，但在提取特征方面有区别，人脸识别提取的特征是不同人脸的个体差异，而表情识别提取的特征是同一人脸在不同表情下的差异。因此本节重点以不同的特征提取方法和分类器的设计方法对表情识别技术进行介绍。

表情特征的提取根据图像性质的不同可分为：静态图像特征提取和序列图像特征提取。静态图像中提取的是表情的形变特征，即表情的暂态特征。而对于序列图像不仅要提取每一帧的表情形变特征还要提取连续序列的运动特征。形变特征提取必须依赖中性表情或模型，把产生的表情与中性表情做比较从而提取特征，而运动特征的提取则直接依赖于表情产生的面部变化^[11]。表情特征的提取还可以根据特征提取的部位不同分为：基于整体的特征提取法和基于局部的特征提取法。基于整体的特征提取就是把整个人脸看作一个整体进行特征提取，而基于局部的特征提取则把人脸划分为局部块或点来进行研究。

1.3.1 表情特征提取方法

表情特征的提取根据图像性质的不同可分为：静态图像特征提取和序列图像特征提取。静态图像中提取的是表情的形变特征，即表情的暂态特征。而对于序列图像不仅要提取每一帧的表情形变特征还要提取连续序列的运动特征。形变特征提取必须依赖中性表情或模型，把产生的表情与中性表情做比较从而提取特征，而运动特征的提取则直接依赖于表情产生的面部变化。特征选择的依据是：①尽可能多的携带人脸面部表情的特征，即信息量丰富；②尽可能容易提取；③信息相对稳定，受光照变化等外界的影响小。

(1) 常用的静态图像特征提取

形变特征提取的常用方法有：主成分分析法 PCA (Principal Component Analysis)、Gabor 小波法、基于模型的方法等。

PCA基于K-L正交变换，是一种无监督的线性特征提取方法。其基本原理是：假设人脸处于低维线性空间，且不同表情具有可分性，将一组高维训练图像K-L正交变换，生成一组新的正交基，以消除原有向量间的相关性。然后提取出该正交基中对应较大特征值的部分特征向量，从而生成低维的表情特征空间，同时保留原高维空间的主要成分分量，它的优点是最大化的保留了原始数据的差异，缺点是当样本类间离散度增大的同时，样本类内离散度也在增大。Andrew、Calder^[12]等详细的介绍了PCA在面部表情识别方面的应用。PCA现已成功用于人脸识别和表情识别。但是由于PCA只考虑到图像数据当中的二阶统计信息，并未利用高阶统计信息，Bartlett于1998年提出了基于独立分量分析ICA (Independent Component Analysis)算法^[13]。ICA方法是基于信号的高阶统计特性的分析方法，经ICA分解出的各个信号分量是相互独立的。其基本思想是：用一些基函数来表示一系列随机变量，而假设它的各个成分之间是统计独立的或者尽可能独立。徐正光^[14]介绍了ICA在表情识别中的应用，并与PCA进行了对比。除了ICA，在PCA方法基础上改进的算法还有：基于模板的PCA法^[15]将人脸图像分割为小的子图像，对每一个子图像做PCA，识别效果比传统的PCA好。程剑^[16]提出的二维主分量分析(2DPCA)的表情特征提取方法，不仅识别率比较高，而且运算速度也有很大的提高。FLD^[17]也是在PCA基础上的一种改进，它试图建立一个更小维数、只包含最基本元素的空间，其目标是最大化类间离散度，同时最小化类内离散度，对表情识别而言，在FLD 提取得到的特征空间，同一类的人脸图像移得更近，而不同类的人脸图像移得更远，更有利于表情的识别分类。

Gabor小波滤波器是一个由二维高斯函数衍生出的复数域正旋曲线函数，相当于一组带通滤波器，其方向、基频带宽及中心频率均可以调节，不同参数的Gabor滤波器能够捕捉图像中对应于不同的空间频率、空间位置以及方向信息，这些使得其对于亮度和人脸姿态的变化不敏感。叶敬福^[18]提出了基于Gabor小波的弹性图匹配识别，有效地提取了与表情变化有关的特征向量且能屏蔽光照变化及个人特征差异的影响，做到与人无关的表情特征提取。文献^[19]以小波变换后的低频信息作为研究对象，只提取低频信息，不仅减小了计算量、在一定程度上去除了噪声的影响，而且抓住了表情变化的主要特征。文献^[20]采用了包含有5个尺度、8个方向的Gabor滤波器组来进行特征提取，将各尺度的特征结合起来得到更好的识别效果。曹宇佳等^[21]提出了基于多尺度多方向差值模板特征的特征提取方法，利用简

单的灰度值加权求和运算模拟出Gabor变换的特征提取性能，有效的降低了运算量，并达到与Gabor变换特征提取基本相似的识别性能。

PCA和Gabor小波都起着给数据降维的作用。PCA是通过将分散在一组变量上的信息集中到某几个综合指标上来降维，而Gabor小波则是通过把原始图像经过变换，减少子图像的尺寸来降维。

基于模型的方法有：点分布模型 PDM(Point Distribution Model)和活动外观模型 AAM (Active Appearance Models) 等。PDM 是一种参数化的形状描述模型，它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作了约束，从而保证模型的整体形变始终在可接受的范围之内。此方法在模型中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，不过运算量较大。Huang 等^[22]采用 PDM 和嘴巴模板来提取人脸的表情特征，PDM 是利用六种表情训练图像中手工提取的 90 个特征点得到的。AAM 也是一种基于模型的方法，它建立了一种对目标对象变化程度的参数化描述。这种方法首先从训练集中通过手工标定一组点构成形状信息，然后通过灰度插值法得到灰度信息，综合这两种信息并用主成分分析法去除冗余信息得到 AAM 基向量；应用时，对实际图像进行形状定位与灰度信息获取，然后计算这些信息在各个 AAM 基向量上的投影，调整模型参数使得通过 AAM 模型合成图像与真实图像间的差异达到最小，并以满足这个条件时的模型参数构建表情特征向量^[23]。Edwards^[24]提出的活动外观模板 AAM (Active Appearance Model)，在图像中手动标定 122 个点用来描述人脸特征信息。左坤隆^[25]选取 70 幅图像作为 AAM 训练集，在每幅图像标记了 57 个特征点，实验得到的识别率为 93.5%。

(2) 常用的动态图像特征提取

运动特征提取的常用方法有：光流法、特征点跟踪法和差分图像法。

光流是运动特征提取法中的一种，所谓光流是指亮度模式引起的表观运动，理想的情况是这种表观运动反映了实际的运动，基于光流的算法包括：基于梯度的光流计算法，基于区域匹配的光流计算方法，基于频率域的方法。光流场在运动估计、运动分割、物体结构的恢复等领域得到了广泛的应用，也是人脸表情特征提取的一种非常有效的算法。

Horn 和 Schunk^[26]最早提出来基于梯度的光流场计算法，Yoneyama 等利用 Horn 和 Schunck 提出的基于梯度的光流场提取人脸归一化图像中 80 个 20×20 大小的图像块的光流特征，这些光流特征是根据中性表情图像和待检表情图像来计算得到的，并且被量化为垂直方向的三个级别。由于这种方法没有考虑表情特征的水平方向移动，因此无法检测包含水平特征移动的表情。Mase 和 Pentland 利

用光流检测到 FACS 系统中的 44 个运动单元中的 12 个。Yabcoob 和 Davis 使用光流跟踪眉毛、眼睛、嘴巴等区域的运动；Barlet 同时使用 PCA 方法和光流模型识别人脸表情；Otsuka 和 Ohya 对眼睛和嘴部等局部区域估计面部运动，将表情图像进行傅立叶变换，其系数作为光流场，并结合 HMM 识别人脸表情；此外，McCane 提出基准光流算法从复杂的背景和多姿态表情序列中有效的获得运动场信息；余棉水^[27]等利用光流技术对人脸表情图像序列中的特征点进行跟踪，并提取特征向量。用光流特征作为特征向量，对光照均匀程度的鲁棒性较强，只要光照在同一组序列中强度不变，提取的特征就是稳定的，解决了图像处理比较敏感的问题。光流法的缺点是：受光照不均匀和脸部非刚体运动等因素，影响特征提取结果，且计算量比较大，不太适合实时处理。

特征点跟踪法是选择脸部的显著特征点来进行运动估计，为了防止在跟踪过程中丢失特征点，特征点一般都选取在脸部的永久特征上。这种方法只考虑感兴趣的特征点来分析表情，而不必理会背景等无关信息，减少了运算量，但由于只提取了部分特征点，而忽略了其它的一些脸部活动，可能会丢失有用信息。此外，自动标记特征点很难，基本都是人工标记。Tian 等人^[28]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，以根据隐藏的面部特征的变形而获得模型的轮廓。Otsuka 和 Ohya^[29]提出了一种特征点跟踪法，其特征点不是由专家手工标记的，而是在给定的序列图像中的第一帧自动标记的，这是通过从光照分布的局部极值点中获取潜在面部特征点来实现的。

差分图像法的基本思想是将视频序列中相邻的两帧作差分运算，即将对应像素相减，等于零的点代表“不变”区域，而非零点代表“变化”区域。当用于表情分析时，通常对被测帧（含有表情的人脸图像）与标准帧（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时表情对应的关键帧，因为此时的表情信息最为丰富。在后继的表情识别研究中只需要对检测出的关键帧进行分析，而不必分析其他所有帧，节约了大量的计算时间。Ma 等人^[30]将含表情的脸和中性表情的脸相减，然后对差值图像进行离散余弦变换，最后以低阶离散余弦变换系数构建特征向量。对四类表情（高兴、愤怒、悲伤、惊奇）的识别率为 93.75%。

基于静态图像的特征提取主要提取面部表情图像的空域信息，简单快捷，但通常要求所处理的图像包含的表情特征比较夸张、明显，因此此方法鲁棒性较差，识别率偏低；基于视频图像通过提取面部表情变化的时间和空间信息，能揭示更多的表情运动变化信息，因此从图像序列提取表情运动信息可获得更好的表情识别效果，但是这种方法模型复杂，计算量大，实现实时性要求较难。

目前的表情特征提取法的趋势是融合各种方法特征，最大限度的有机利用信息来进行特征提取。刘松^[31,32]将局部特征和整体特征结合起来进行面部表情识别。文献^[33]中用Fisher线性判别提取脸部表情的整体特点，然后通过局部特征和整体特征的融合来提高识别率。

1.3.2 主要的表情分类方法

表情识别的最后一步就是表情的分类，特征提取之后，通过分类器就可以确定给定的对象属于哪一类。基本做法是在样本集的基础上确定判别规则，对于新给定的对象根据已有的判别规则来分类，从而达到识别的目的。一个好的分类器使分类造成的错误率最小，因此，分类器的设计也是表情识别的关键。

一般常用的分类方法有以下几种：

(1)最近邻法(NN)是基于样本间距离的一种分类方法。最初的近邻法是由Cover和Hart于1968年提出的。由于该方法在理论上进行了深入的分析，直至现在仍是模式识别中重要的方法之一。其相似度量以检测样本到模板的距离为判据。距离的测度主要有L1范数，L2范数，马氏距离等。

(2)基于模板的匹配方法。为每一个要识别的表情建立一个模板，将待测表情与每种表情模板进行匹配，匹配度越高，则待测表情与某种表情越相似。对6种典型的基本人脸表情可以建立相应的模板，但对于识别非典型脸部表情比如混合表情时较困难，而且很难用有限的模板来代表无限多的人脸表情的情况。

(3)基于神经网络的方法。神经网络方法的研究具有悠久的历史，它很适合于解决非线性问题。网络的输出节点对应于6种基本表情和中性表情。Padgett等使用反向传播算法训练人工神经网络在Ekman人脸表情库上能够达到86%的识别率。Hara和Kobayashi采用一个 $234 \times 50 \times 6$ 的反向传播神经网络，输入层的神经元对应从输入脸部图像中提取的亮度分布数据，输出神经元对应6种基本表情类别。但是，如果对自然表情类型而不是6种基本表情分类，那么神经网络方法的训练过程的工作量会非常大。另外，神经网络除了可以当作分类器外还可以作为特征提取器。

(4)基于概率模型的方法。估计表情图像的参数分布模型，分别计算被测表情属于每个类的概率，取最大概率的类别为识别结果。如Bays决策，HMM等。隐马尔可夫模型是一个统计模型，具有十分健壮的数学结构，适用于动态过程时间序列建模并具有强大的时序模式分类能力，理论上可以处理任意长度的时序。1995年Sakaguchi首先将HMM应用到图像序列，实现人脸表情的识别。尹星云等^[34]用隐马尔可夫模型设计人脸表情识别系统。

(5) 基于支持向量机 SVM(Support Vector Machines)的方法。SVM^[35]是一种二类分类器,其基本思想是将训练样本经非线性变换映射到另一高维空间中,在变换后的高维空间内寻找一个最优的分界面,在类间空隙最大的情况下将样本无错误或接近无错误分开。由于表情识别一般是多类别的分类问题,因此需要将 SVM 方法扩展到多类别的情形。主要有两种策略,一种是将多类别问题分解成一系列二类别问题来求解;另外一种是将包含一类不同样本的多类别问题分解成一个二类别问题,每次任选 C 类样本作为二类别问题的一类,剩下的 C-1 类样本作为另一类。

1.4 本文的主要工作及结构安排

本文从模式识别和统计学习的角度对人脸表情识别进行研究。实验数据采用日本的 JAFFE 表情数据库,首先对表情图像进行亚采样预处理,去掉了头发、颈部和背景等与表情无关信息,同时将图像降低为 64×64 ,并将每个像素灰度值都归一化至 $[0, 1]$ 间,利于特征提取。然后对表情图像进行直方图均衡,使得图像的细节更加清楚。采用 PCA 进行特征提取,用欧氏距离作为模式特征之间的相似性度量,用最近邻分类器进行分类,取得了较好的识别效果,最后将核最近邻分类器应用于表情识别,利用 Mercer 核把输入空间样本映射到特征空间,通过非线性映射后较好地加大样本间的区别,从而提高了分类的准确率。采用了基于支持向量机(SVM)的分类方法,利用 one-against-rest 组合原理,构建了 7 个 SVM 并对之进行组合,然后用组合后的 SVM 进行表情识别,取得了较好的效果。在原有 SVM 的基础上进行了改进,提高了识别率。最后对最近邻分类方法和支持向量机分类方法进行了简单的比较和结合,识别率得以进一步的提高,并得出下面的结论,即:在有些应用场合中,要求快速实时分类,而对分类正确率要求不高,此时可考虑使用基于最近邻的表情识别方法;而在另外一些应用场合中,要求分类准确率较高,而对分类速度不作要求或要求降低时,可考虑使用基于 SVM 的表情识别方法。

本文的结构安排如下:

第一章是论文的绪论部分。对人脸表情识别进行了简单介绍,概述了人脸表情识别的研究意义和应用前景,分析了国内外的研究现状,并对表情识别的研究方法进行了综述,最后介绍了本文的主要研究工作和结构安排。

第二章介绍了常用的表情数据库和表情图像预处理方法。在进行表情特征提取和分类前对表情图像进行亚采样和归一化预处理,经过亚采样抽取的像素组成的网格,只保留了眼睛、嘴角及眉毛等能够反映表情变化的面部区域,去掉了一

些干扰信息，并将图像降低为 64×64 ，以利于后续研究。最后对表情图像进行直方图均衡，使得图像的细节更加清楚。

第三章介绍了特征提取的一些基本概念及PCA的基本原理。并将PCA方法应用于表情特征提取中，给出了一种基于PCA的特征提取方法。将PCA方法用于人脸表情特征提取时，由高维图像空间经K-L变换后得到一组新的正交基，对这些正交基作一定的取舍，保留其中的一部分生成低维的人脸空间，也即人脸表情的特征子空间。

第四章给出了一种基于最近邻分类器的表情识别方法。利用PCA进行特征提取后，用最近邻分类器对表情进行分类。分类正确率随着主成分个数的变化而变化，并且分类速度随着亚采样间隔的变化也在变化。最后将核最近邻分类器引入到表情识别领域，利用Mercer核把输入空间样本映射到特征空间，通过非线性映射后较好地加大样本间的区别，从而实现更为准确的分类。最后对这两种方法进行了比较。实验数据采用JAFPE表情库中的210幅图像。

第五章给出了一种基于 SVM 的表情识别方法。介绍了最优分类面的概念，讨论了多类 SVM 分类器的两种构造方法。本文利用 one-against-rest 组合原理，构建了 7 个 SVM 并对之进行组合，然后用组合后的 SVM 进行表情识别。同时在此基础上对原有 SVM 进行了改进，使 7 个 SVM 不再共用同一个核参数，而是根据学习样本自适应地估计各 SVM 的核参数，取得了较好的识别效果。最后对最近邻分类方法和 SVM 分类方法进行了分析与比较。

第六章总结了本文所做的研究，指出了需要进一步的研究的内容。

第2章 表情图像的预处理

2.1 表情数据库

表情识别研究的一个重要前提是建立表情数据库,它是整个表情识别过程的基础。

最早的人脸表情图像库是Ekman和Friesen建立的数千幅不同人脸表情图像库。目前最为常用的人脸表情数据库是日本女性表情数据库JAFFE (Japanese Female Facial Expression) 和CMU的Cohn—Kanade人脸表情数据库,前者由日本ART 媒体信息科学实验室的Lyons博士提供研究用的,此数据库包含了10名日本年轻女性的213幅图像,每人包含6种基本表情和中性表情灰度图像,对每个人来说,这些图像没有光照、头部角度、穿戴和发型的影响,便于运算处理。库中图像的原始尺寸为256×256像素大小,256级灰度。后者是CMU机器人研究所和心理系于2000年共同建立的人脸表情数据库,它包含了200个由18到50岁的210个成年人近2000张不同表情灰度图像序列,其中69%是女性,31%是男性,81%是欧美人,13%是黑人,6%的其他人种。这是个基于动作单元(Action Unit)编码的人脸表情数据库,它根据人脸动作编码系统(Facial Action Coding system),把脸部动作分为46种AU,每一种表情均由一系列的动态图像构成,每幅图像大小为640×490或640×480,均为8位的灰度图像或24位的彩色图像,采集过程中,部分表情图像人为地引入光照变化。



图 2-1 JAFFE 人脸表情数据库中的表情图

除了JAFFE表情库和Cohn—Kanad表情库外,目前还有其它一些表情库。例如CMU的PIE数据库^[36]在CMU的3D Room中录制完成。该数据库包含13个姿态,43种光照条件,收集了人脸常见的4种表情:中性表情、微笑、眨眼、说话,规模为68人。MIT人脸库^[37]有16人,每人27幅图片,有光照、尺度和头部偏转变化。UMIST人脸库有20人564幅图像,包含从正面到侧面的各个角度。Kodak人脸库为彩色图像,有尺寸、姿态和光照变化。但是,这些数据库的表情分类都不够丰富,包含的人数较少,大部分都没有图像序列或者不是彩色图像。而且,这些数据库没有专门针对说话类表情的细致分类,更没有专门用于中文视位研究的说话类表情。

我国学者吴丹、林学闯^[38]建立了一个大型人脸表情视频数据库，该数据库从3个不同的视角记录图像数据，包括了70个人的1000段脸部表情视频，它是目前我国人脸表情研究较全面的基础资源库。东北大学的张庆凯^[39]等建立了一个小型人脸表情视频数据库，并制定了一套人脸表情视频数据库技术规范。该数据库采集时采用正面充分光照，从三个不同的视角记录图像数据，内容包括了7个人的100段脸部表情视频，涵盖了常见的8类情感类表情和中文语音发音的说话类表情，并对表情序列采用了FACS(脸部动作编码系统)评价。

由于JAFPE表情库图像没有光照、头部角度、穿戴和发型的影响，规范化程度较高，故本文选取它作为实验样本。

2.2 表情图像的预处理

一般情况下，成像系统获取的图像(即原始图像)由于受到各种条件的限制和随机干扰，往往不能直接使用，必须在图像处理的早期阶段对原始图像进行灰度校正、噪声过滤等预处理。所用的图像预处理方法并不考虑图像降质原因，只将图像中感兴趣的特征有选择地突出，衰减其不需要的特征。故预处理后的输出图像并不需要去逼近原图像。

本论文所用的表情图像来自日本的JAFPE人脸表情库。对表情识别来说，与人脸相连的颈部、头发等只是背景，对识别没有帮助，因此在实验中将原始图像进行了切割，只保留了面部的主要区域。最后对图像进行了直方图均衡处理。

2.2.1 表情图像的亚采样和归一化

亚采样就是将图像当作二维信号，并对之进行抽样，目的是降低图像的维数，便于下一步用PCA进行特征提取。亚采样的原理如图2-2所示，设亚采样间隔为 Δ ，现每隔 Δ 个图像在图像上抽取一个像素，亚采样后的图像则由这些抽取的像素组成。所有抽取像素可在原图像上构成一个亚采样网格。从图中可见该网格主要覆盖了面部区域，只保留了眼睛、嘴角及眉毛等能够反映表情变化的部分，而没有覆盖头发、颈部以下区域和背景区域，这是因为颈部以下和背景区域对表情识别不起作用，这样便减少了无用信息。

本文的亚采样由函数 `subsample (aaa, row, col, intv, shift)` 完成，其中 `row` 为图像经亚采样后的行像素数，`col` 为图像经亚采样后的列像素数，`intv` 为亚采样间隔 Δ ，参数 `shift` 控制着亚采样网格的水平位移的大小。

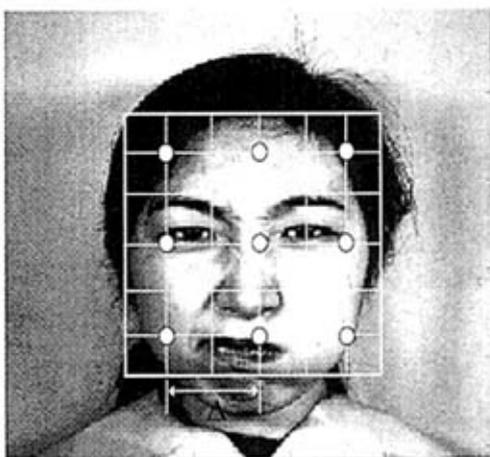
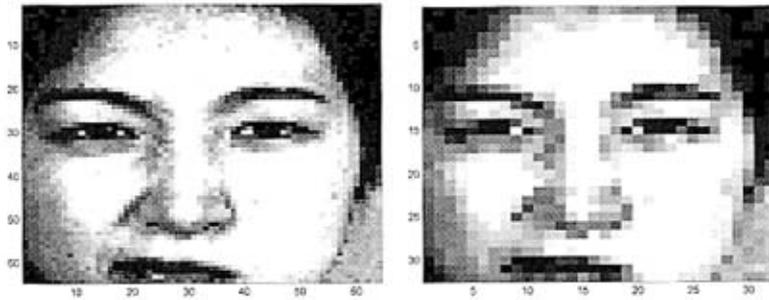


图 2-2 图像亚采样原理示意图(图像大小 256×256 像素)

图 2-3 显示了图 2-2 中的表情图像经亚采样后的结果，其中图(a)的亚采样间隔 $\Delta=2$ ，图像大小为 64×64 像素；图(b)的亚采样间隔 $\Delta=4$ ，图像大小为 32×32 像素。降低像素主要是为了降低运算量，因为我们需要多次计算一个矩阵的逆矩阵、特征向量矩阵等，而维数较高的矩阵运算往往很费时，因此在不影响图像信息的前提下，尽可能缩小图像。

对比图 2-3(a)和图 2-3(b)可以看出，当 Δ 增大时，图像的像素变少，但图像的分辨率也随之降低了。

亚采样间隔 Δ 的取值合适与否对分类正确率与分类速度有很大影响。 Δ 过小，尽管可以保持较好的分类正确率，但由于经亚采样后的图像像素数仍很多，因此分类速度很慢； Δ 过大，尽管分类速度较快，但由于在亚采样时图像丢失了很多对表情识别有用的细节，使得分类正确率下降。 Δ 的合理取值通过实验确定，下文还会专门讨论分类正确率随亚采样间隔变化而变化的情形。



(a) 亚采样间隔为 2 (b) 亚采样间隔为 4

图 2-3 图像经亚采样后的结果

本文采用的 JAFFE 表情图像库中的每幅图像均为灰度图像，每个像素的亮度值在 $[0, 255]$ 间。在后续的运算（例如特征提取）中会涉及像素亮度值的连乘积，因此很可能会出现很大的数值。而过大的数值与较小的数值并存会影响矩阵求逆的精度，为避免此情况发生，特对图像进行归一化处理。

归一化意味着将图像的每个像素亮度值都转换到 $[0, 1]$ 间。设某图像需归一化，现首先将之转化成一维向量 x ，然后对 x 中的每个元素 $x(i)$ 进行如下处理：

$$\frac{x(i) - \min(x(i))}{[\max(x(i)) - \min(x(i))]}$$

上式中 $\min(x(i))$ 表示从向量 x 找出其最小的元素值。

2.2.2 表情图像的灰度值的归一化

由于在图像采集中光照的改变容易导致图像呈现不同的明暗程度，因此需要对表情图像进行灰度归一化处理。灰度归一化也称为灰度均衡化，其作用是增强表情图像的整体对比度，并使灰度分布均匀，以消除光照变化的影响，此外还可以消除不同人种的肤色差异。进行灰度均衡化处理可优化特征提取结果。通过对表情图像灰度直方图的修正从而调整图像的均值和方差来完成图像的均衡化处理，因此灰度均衡又叫直方图均衡。

灰度直方图是数字图像处理中一个实用的工具，它可有效描述一幅图像的灰度级内容，尽管它不能表示出有某灰度级的像素在什么位置，更不能直接显示图像内容，但是具有统计特性的直方图却能描述该图像的灰度分布特性，使人们从中得到的诸如总体明亮程度、对比度等灰度概貌，成为一些处理方法的重要依据。直方图的横坐标表示像素的灰度级别，对于 256 色的灰度图像，该坐标值范围为 0 到 255；其纵坐标是相应的灰度值在整幅图像中出现的频率，即像素的个数。图像的明暗程度可由均值表示，图像亮度越高均值越大，亮度越低，均值越小。

直方图均衡化是一种灰度增强的算法。一幅对比度小的图像，其直方图分布

一定集中在某一比较小的范围内，经过均衡化处理的图像，其所有灰度级出现的概率相同，此时图像的熵最大，图像所包含的信息最大。理论上来说，直方图均衡化是通过变换函数将原始图像的直方图调整为平坦的直方图，然后用次均匀直方图校正图像。

图 2-4 是 JAFFE 表情库中一幅表情图像直方图均衡化后的图像，经过均衡化后，图像的细节更加清楚，直方图各灰度等级的分布更加平均。

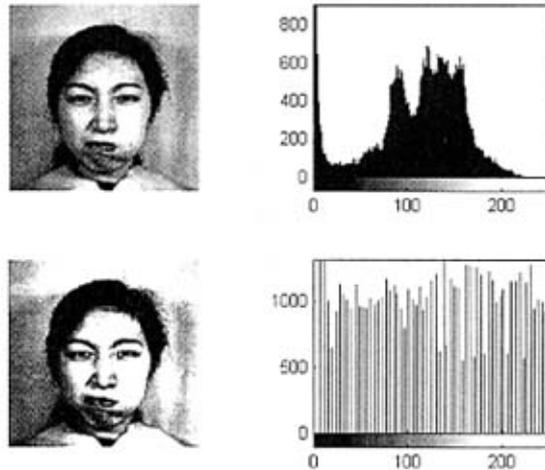


图 2-4 直方图均衡化后的一幅表情图像

2.3 本章小结

本章介绍了常用的一些表情数据库。并对表情图像进行亚采样、归一化和灰度均衡化预处理。通过对表情图像的亚采样，只保留了眼睛、嘴角及眉毛等能够反映表情变化的面部区域，将像素降低为 64×64 ，并将图像的每个像素亮度值都转换到 $[0, 1]$ 间，利于后续的特征提取。然后对表情图像进行直方图均衡化处理，经过均衡化后，图像的细节更加清楚，直方图各灰度等级的分布更加平均。论文后面章节所进行的实验都是在本章预处理的基础上进行的。

第3章 基于主成分分析的特征提取

3.1 引言

特征提取是表情识别中的关键，其目的是获取特征数目少而分类错误概率小的特征向量。

特征的选择和提取的基本任务是从样本的许多特征中找出那些最有效的特征。我们知道对于人脸表情识别问题，其原始特征为 $m \times n$ 维的矩阵，其维数比较高，相比而言样本数目则较少，如果不对特征进行选择而直接进行分类器设计，无论从计算的复杂程度还是分类器性能来看都是不适宜的。因此研究如何把高维特征空间压缩到低维特征空间以便有效的设计分类器就成为重要的课题。

特征的选择和提取一般分成三个步骤^[40]：特征的形成、特征的提取和特征的选择。

特征的形成：根据被识别的对象产生出一组基本特征，当识别对象是人脸表情图像时，原始量就是各点灰度值。

特征的提取：从一组特征中挑选出一些最有效的特征以达到降低维数的目的，这个过程叫特征提取。原始特征的数量可能很大，或者说样本处于一个高维空间中，通过映射(或变换)的方法可以用低维空间来表示样本，这个过程叫特征提取。映射后的特征叫二次特征，他们是原始特征的某种组合(通常是线性组合)。所谓特征提取在广义上就是指一种变换。若 Y 是原始空间， X 是特征空间，则变换 $A: Y \rightarrow X$ 就叫做特征提取器。特征选择和提取的任务是求出一组对分类最有效的特征，因此我们需要一个定量的准则(或称判据)来衡量特征对分类的有效性。具体的说来，把一个高维空间变换为低维空间的映射是很多的，哪种映射对分类最有利，需要一个比较标准。

特征选择：从一组特征中挑选出一些最有效的特征以达到降低维数的目的，这个过程叫特征选择。特征选择的标准为：可区别性、可靠性、独立性、数量少。

下面对PCA特征提取算法作详细的介绍。

3.2 基于主成分分析的表情特征提取

3.2.1 主成分分析方法的基本原理

主成分分析方法是应用最广泛的一种特征提取方法之一，它是一类重要的多元统计分析方法，在信号处理、模式识别、数字图像处理等领域已经得到了广泛的应用。

主成分分析法是模式识别中最常用的一种线性映射方法之一，该方法是根据样本点在高维模式空间的位置分布，以样本点在空间中变化最大方向，即方差最大的方向，作为判别矢量来实现数据的特征提取与数据压缩。从概率统计观点可知，一个随机变量的方差越大，该随机变量所包含的信息就越多，当一个变量的方差为零时，该变量为一常数，不含任何信息。所谓主成分就是原始数据的 m 个变量经线性组合(或映射)后得到的变量，该变化使得其变换后的变量方差为最大(第一主成分)的部分。各个主成分之间是相互线性无关的(正交的)，从第一主成分往后，主成分按方差大小的顺序排列(对应特征值按大小顺序排列)。对于特征值为 λ_i 的主成分， λ_i 也是该主成分的方差，该值表示样本点在该主成分方向上的离散程度，主成分 λ_i 的贡献率 η_i 可用下式表示

$$\eta_i = \lambda_i / (\lambda_1 + \lambda_2 + \dots + \lambda_p) \quad (3-1)$$

主成分中方差较小或 λ_i 较小的主成分被认为包含的是噪声，在分析时不使这些变量引入模型，这样使分析的主成分减少，以达到降维的目的。

主成分中任意两个主向量可构成判别分析平面，因此可实现高维空间向两维平面及其他维平面映射的目的。一般取方差较大的几个主成分构成判别分析平面。

主成分分析法的基础是Karhunen-Loeve(简称K-L变换)，它是一种常用的正交变换。

假设 X 为 n 维的随机变量， X 可用 n 个基向量的加权和来表示：

$$X = \sum_{i=1}^n \alpha_i \phi_i \quad (3-2)$$

式中 α_i 一加权系数， ϕ_i 一基向量，此式还可以用矩阵的形式表示：

$$X = (\phi_1, \phi_2, \dots, \phi_n)(\alpha_1, \alpha_2, \dots, \alpha_n)^T = \Phi \alpha \quad (3-3)$$

其中

$$\Phi = (\phi_1, \phi_2, \dots, \phi_n), \alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$$

我们取基向量为正交向量，即

$$\Phi^T \Phi_j = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases} \quad (3-4)$$

Φ 由正交向量构成，所以 Φ 是正交矩阵，即

$$\Phi^T \Phi = I \quad (3-5)$$

将公式(3-1)两边左乘 Φ^T ，并考虑到 Φ 为正交矩阵，得

$$\alpha = \Phi^T X$$

即

$$\alpha_i = \Phi_i^T X \quad (3-6)$$

为了保证 α 的各个分量互不相关,取决于选取什么样的正交向量集 $\{\Phi_j\}$
 设随机向量的总体自相关矩阵为:

$$\begin{aligned} R &= E[X^T X] \\ R &= E[X^T X] = E[\Phi \alpha \alpha^T \Phi^T] = \Phi E[\alpha \alpha^T] \Phi^T \end{aligned} \quad (3-7)$$

如果要求向量 α 的各个分量间互不相关,即满足下列关系:

$$E[\alpha \alpha^T] = \begin{cases} \lambda_j, & j = k \\ 0, & j \neq k \end{cases} \quad (3-8)$$

写成矩阵的形式是:

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \Lambda$$

则

$$R = \Phi \Lambda \Phi^T \quad (3-9)$$

将上式两边右乘 Φ ,得

$$R\Phi = \Phi \Lambda \Phi^T \Phi$$

由于 Φ 是正交矩阵,故

$$R\Phi = \Phi \Lambda$$

即

$$R\Phi_j = \lambda_j \Phi_j, (j=1,2,\dots,n) \quad (3-10)$$

可以看出, λ_j 是 X 的自相关矩阵 R 的本征值, Φ_j 是对应本征向量。因为 R 是实对称矩阵,其不同本征值对应的本征向量应正交。

K-L变换的实质是建立了一个新的坐标系,将一个物体主轴沿特征向量对齐的旋转变换,这个变换解除了原有数据向量的各个分量之间相关性,从而有可能去掉那些带有较少信息的坐标系以达到降低特征空间维数的目的。

在主成分分析法中主要的是要求出使得方差最大的转化方向,其具体的求解步骤如下所示:

(1) 构建关联矩阵, $C_x = E[xx^T]$, $C_x \in R^{n \times n}$ 。

在实际应用中,由于原始数据的数学期望不容易求解,我们可以利用下式来近似构造关联矩阵:

$$C_x = \frac{1}{N} [x_1 x_1^T + x_2 x_2^T + \dots + x_N x_N^T]$$

(其中 x_1, x_2, \dots, x_N 是各个原始灰度图像所有像素点对应的向量, N 是原始图像的个

数)

(2) 先计算出 C_x 的各个特征值。

(3) 把特征值按大小排序

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

(4) 计算出前 m 个特征值对应正交的特征向量构成 w 。

(5) 将原始数据在特征向量 w 上投影, 即可获得原始图像的主特征数据^[41]。

主成分分析的两大优点是: 消除了模式样本之间的相关性; 实现了模式样本的维数压缩, 换言之, 主成分分析给出了高维数据的一种简约的表示。

主成分分析方法基本思想是提取出空间原始数据中的主要特征(主元), 减少数据冗余, 使得数据在一个低维的特征空间里被处理, 同时保持原始数据的绝大部分信息, 从而解决数据空间维数过高的瓶颈问题。它在保证信息损失最少的前提下, 对高维变量空间进行降维处理。

主成分分析通过求解训练样本散布矩阵的特征值问题, 给出一组数量远远小于样本空间维数的正交基来表示训练样本张成的子空间。从线性重建的角度而言, 这组基的优点在于能够最充分地表征样本。因此K-L变换具有降维、去相关和集中能量的特性。

3.2.2 主成分分析方法提取表情特征

2001年, Andrew J. Calder等用主成分分析算法在面部表情识别方面做了详尽细致的工作, 证明了PCA在表情识别算法上的可行性和有效性。作者认为PCA是表情种类和识别图像维数之间的一个桥梁, 达到了最佳的识别率和识别速度的整体性能。

将PCA方法用于人脸表情特征提取, 其实是假设所有的人脸都处于一个低维线性空间, 而且不同的人脸表情在这个空间中具有可分性。其具体做法是由高维图像空间经K-L变换后得到一组新的正交基, 对这些正交基作一定的取舍, 保留其中的一部分生成低维的人脸空间, 也即人脸表情的特征子空间。

将主成分分析方法应用于表情识别, 可以按照以下步骤进行:

使用大量的表情图像进行训练, 得到一个用主成分描述的多维坐标系(即特征空间)。然后把表情模板投影到该特征空间中, 得到表情模板在该空间的投影系数, 作为该模板的特征。对于一幅待识别的表情图像, 将其投影到该特征空间中, 与表情模板进行匹配, 从而对该幅表情图像进行分类。

表情图像的特征空间的训练过程如下:

假设一幅图像由 M 个像素组成, 那么可以将它表示成一个 M 维的矢量 x , 设

$\{x_i | i=1, K, N\}$ 是表情图像的一个训练集合。则这N幅表情图像的均值向量（所有图像的平均值，即平均表情）为：

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (3-11)$$

每幅图像 X_i 与均值向量图像之间的差为：

$$\phi_i = x_i - \bar{x} (i=1, 2, K, N) \quad (3-12)$$

将相减得到的表情图像作为列向量，创建数据矩阵：

$$A = [\phi_1, \phi_2, K, \phi_N] \quad (3-13)$$

将数据矩阵与其转置相乘，得到训练图像集的协方差矩阵：

$$C = AA^T = \frac{1}{N} \sum_{i=1}^N \phi_i \phi_i^T \quad (3-14)$$

根据K-L变换原理，特征子空间的坐标系即由矩阵的非零特征值所对应的特征向量组成。

求解 AA^T 的特征值和特征向量来获得 AA^T 的特征值和特征向量。

令 $\lambda_i (i=1, 2, K, k)$ 为矩阵 AA^T 的K个非零特征值， u_i 为 AA^T 对应于 λ_i 的特征向量，则 AA^T 的正交归一特征向量 v_i 为

$$v_i = \frac{1}{\sqrt{\lambda_i}} Au_i (i=1, 2, K, k) \quad (3-15)$$

这就是协方差矩阵的特征向量。将其特征值从大到小进行排序，同时特征值所对应的特征向量顺序也作相应调整。然后取前d个较大特征值 ($\lambda_1 \geq \lambda_2 \geq K \geq \lambda_d$) 的特征向量 (u_1, u_2, K, u_d) 构造特征子空间，即 $X = (u_1, u_2, K, u_d)$ 。

把训练图像和测试图像投影到该子空间中获得一组坐标系，由于这组系数可表明该图像在子空间的位置，因此原表情图像的识别问题就可转化为依据子空间中的训练样本点进行分类的问题。

最后通过一定的距离准则，计算待识别的表情投影矢量与表情模板之间的距离，得到与各个表情模板的相似度，最后按照一定的分类准则，对该表情进行分类。

该算法的意义就是通过对大量样本进行主成分分析得到表征表情子空间的一组正交基（在人脸识别中，该特征基被称为特征脸），所提取的特征就是人脸表情图像在这个子空间中的投影系数向量，再根据表情图像在该空间中的位置采用分类器分类。

3.3 本章小结

本章介绍了特征提取的一些基本概念，以及表情识别过程中的特征提取问题。将主成分分析方法应用于表情特征提取时，首先对表情图像进行训练，得到一个用主成分描述的多维坐标系。然后把表情模板投影到该特征空间中，得到表情模板在该空间的投影系数，作为该模板的特征。对于一幅待识别的表情图像，将其投影到该特征空间中，与表情模板进行匹配，从而对该幅表情图像进行分类。

第4章 基于最近邻方法的表情识别

最近邻法 NN (Nearest Neighborhood Classifier) 是基于样本间距离的一种分类方法。最初的近邻法是由 Cover 和 Hart 于 1968 年提出的。由于该方法在理论上进行了深入的分析，直到现在仍是模式识别中重要的方法之一。

4.1 最近邻分类器的基本原理

最近邻方法是一种经典的机器学习方法，主要用来解决分类问题。NN 的常见形式有 1-NN 和 k -NN 两种，其中前者可看作后者的特例。本章不细分二者，而是笼统地称之为 NN 方法。

正如其名所示，NN 在分类时运用如下“最近邻分类规则”对测试样本 x 进行分类：在 x 的周围找出离其最近的 k 个近邻样本，该 k 个样本中最频繁出现的那一类样本所对应的类别就是 x 所属的类别。以图 4-1 为例，图中存在两类样本，第一类记为 c_1 并用黑点表示，第二类记为 c_2 并用方框表示。现欲用 NN 判决 x 是属于 c_1 类还是 c_2 类。假设取 $k=5$ ，则可找出 x 的 5 个近邻样本并用圆圈围住，圆圈中有 3 个样本属于 c_1 类（目标值 $t_n=1$ ， $n=1,2,3$ ），2 个样本属于 c_2 类（目标值 $t_n=-1$ ， $n=4,5$ ），那么 x 根据规则应属于 c_1 类。

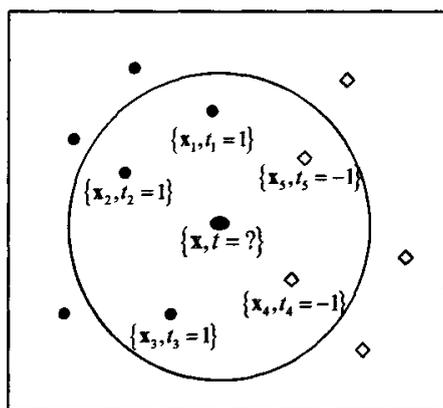


图 4-1 NN 的分类原理示意图

NN 的历史最早可以追溯到 1951 年 Fix 和 Hodges 未发表的一篇研究报告^[42]，但直到 1962 年 Cover 的经典论文^[43]发表，该方法才逐步受到重视。NN 原理简单，便于从理论角度对其分类错误率 p_e 进行深入研究，也便于与其它分类器进行比较，因此它已成为衡量诸多分类器性能的标尺。

设 N 个样本可被分为 c_1 与 c_2 两类（为简单起见只讨论两类样本的情形），且设

样本 x 的最近邻为 x' ，则在有 N 个样本的前提下，NN 对 x 的分类错误率为

$$\begin{aligned} p_N(e|x) &= P(c_1|x)P(c_2|x') + P(c_2|x)P(c_1|x') \\ &= P(c_1|x)[1 - P(c_1|x')] + P(c_1|x')P(c_2|x) \\ &= 2p(c_1|x)p(c_2|x) + [p(c_1|x) - p(c_2|x)] \cdot [p(c_1|x) - p(c_1|x')] \end{aligned} \quad (4-1)$$

其中 $p(c_1|x)$ 表示 x 属于 c_1 类的概率。上式意义很直观：根据 NN 分类规则， x 应与其最近邻点 x' 划归为同一类别，若二者没被划归为同一类，则是对 x 的误判。当 $N \rightarrow \infty$ 时，NN 对 x 的分类错误率变为

$$p(e|x) = \lim_{N \rightarrow \infty} p_N(e|x) = 2p(c_1|x)p(c_2|x) \quad (4-2)$$

而根据对分类错误率的定义，在 N 个样本的前提下，NN 的分类错误率为：

$$p_{Ne} = \int p_N(e|x)p(x)dx \quad (4-3)$$

因此当 $N \rightarrow \infty$ 时，最终 NN 的分类错误率为

$$p_e = \lim_{N \rightarrow \infty} p_{Ne} = \int p(e|x)p(x)dx \quad (4-4)$$

从上式可以看出，该式与式(2-7)对分类错误率的定义是一致的。

已经证明

$$p_{Bayes} \leq p_e \leq p_{Bayes} \left(2 - \frac{c}{c-1} p_{Bayes} \right) \quad (4-5)$$

上式中 p_{Bayes} 是贝叶斯决策器的分类错误率。式(4-5)说明 NN 的分类错误率介于 p_{Bayes} 与两倍 p_{Bayes} 间。正是 NN 的这种优良性质，使其成为经典的学习方法之一^[44]。

距离分类准则是以特征空间中各点间的距离作为样本的相似性度量，并以各类训练样本点的集合所构成的区域表示各决策区，空间中两点距离越近，则认为两样本越相似。当采用距离方式对特征空间中的样本进行分类时，最“自然”的一种距离就是欧几里德距离(欧式距离)。在一些分类问题中，欧式距离对于衡量样本间的相似性是比较有效和方便的。

设待识别的模式向量为 x ，类别 $\omega_i (i=1,2,K,C)$ 由典型模式向量 x_i 表示，则 x 与 x_i 的欧氏距离为

$$d(x, x_i) = [(x - x_i)^T (x - x_i)]^{1/2} \quad (4-6)$$

采用欧式距离的平方定义欧氏距离判决准则如下：

$$\text{若} \quad \min_i [d^2(x, x_i)] = d^2(x, x_i) \quad (4-7)$$

$$\text{则} \quad x \in \omega_k$$

4.2 核最近邻分类器

核最近邻分类(Kernel nearest neighbor)方法通过利用Mercer核,把输入空间的样本映射到高维特征空间,并且在特征空间中用最近邻法分类。核最近邻分类方法在性能上比最近邻分类方法有较大的改进,它通过非线性映射能够较好地加大样本间的间隔,从而实现更为准确的分类。

核最近邻法的基本思想是利用Mercer核把输入空间样本映射到特征空间,使得映射后同类样本接近,异类样本变远。假设输入空间样本已被映射到特征空间变为 $\Phi(x_1), \Phi(x_2), \dots, \Phi(x_l)$,则特征空间中欧几里德(Euclidean)距离可表示为:

$$d^H(x, y) = \sqrt{\|\Phi(x) - \Phi(y)\|^2} = \sqrt{\Phi(x) \cdot \Phi(x) - 2\Phi(x) \cdot \Phi(y) + \Phi(y) \cdot \Phi(y)} \quad (4-8)$$

即

$$d^H(x, y) = \sqrt{K(x, x) - 2K(x, y) + K(y, y)} \quad (4-9)$$

已知分别属于 C 个类 $\{\omega_i, i=1, 2, \dots, C\}$ 的 l 个样本 $\{x_1, x_2, \dots, x_l\}$,对于未知类别样本 x 的判决过程为,计算未知样本到已知样本的核距离,找到其最小距离对应的样本 x_i 。

判决决策规则为:

$$\text{若 } x_i \in \omega_j, j=1, 2, \dots, C \text{ 则 } x \in \omega_j \quad (4-10)$$

该决策准则的几何解释是,对于未知样本 x ,比较 x 与 l 个已知类别的样本之间的核距离,判决 x 与其最近的样本同类。

4.3 实验过程及分析

本文实验图像来自日本的JAFFE人脸表情静态图像库,由10位日本年轻女性的213幅图像组成,每人展示七种表情(生气AN、厌恶DI、恐惧FE、高兴HA、中性NE、悲伤SA、惊讶SU)各2-4幅图像,均为 256×256 的8位灰度图像。为实验方便,本文选取每个人各3幅图像,有缺失的按顺序补全,合计为210幅图像。实验前先对图像进行亚采样等预处理,得到大小为 64×64 的规范化图像。图4-2为10位日本年轻女性组成的图像,其中每人含有七种表情,共70幅图像。



图 4-2 分别含有 7 种表情的 10 个日本人图像

以下所有实验均在 DELL P4:3.00GHz 主频 CPU, 512M 内存计算机上用 Matlab6.5 实现。

由于是采用最近邻方法进行表情识别，其实质是任给一副测试图像，计算它与图中每副图像的欧氏距离。若测试图像与图像 X 距离最近，则将测试图像判别为 X 所属的类别。最近邻方法对七类表情的分类正确率如表 4-1 所示，其中某类表情的分类正确率定义为：

某类表情的分类正确率 = $\frac{\text{NN 对属于某类表情的测试样本进行判别, 判别正确的样本数}}{\text{属于某类表情的测试样本数目}}$

该指标用于度量最近邻方法对某类表情的识别能力。

另外定义“平均分类正确率”指标，它并不是 7 类表情的分类正确率的简单平均，其定义为：

平均分类正确率 = $\frac{\text{NN 对所有测试样本进行判别, 判别正确的样本数}}{\text{所有测试样本数目}}$

该指标用于度量最近邻方法对 7 类表情的综合识别能力。

表 4-1 最近邻方法对七类表情的分类正确率表

序号	表情类型	分类正确率(%)
1	生气(AN)	90.000000
2	厌恶(DI)	90.000000
3	恐惧(FE)	76.666667
4	高兴(HA)	90.000000
5	中性(NE)	93.333333
6	悲伤(SA)	76.666667
7	惊讶(SU)	83.333333
平均分类正确率		85.714286

从表 4-1 看出，最近邻方法对“中性(NE)”表情的分类正确率最高，对“恐惧(FE)”和“悲伤(SA)”表情的分类正确率最低，另外对“生气(AN)”、“厌恶(DI)”和“高兴(HA)”也取得了不错的分类结果。根据平均分类正确率“85.714286%”来看，分类结果令人满意。之所以对“中性(NE)”表情的分类正确率最高，是因为从图 4-2 可以看出，该类表情与其它类表情差异较大。而对“恐惧(FE)”和“悲伤(SA)”表情的分类正确率最低，是因为这两类表情与其它类表情差异较小，而且表情本身的特征不明显。

下面分析“提取的特征数目(numpc)”对平均分类正确率的影响。让 numpc 分别取 5-40，记录 numpc 每取一个值时对应的平均分类正确率的关系如图 4-3 所示。

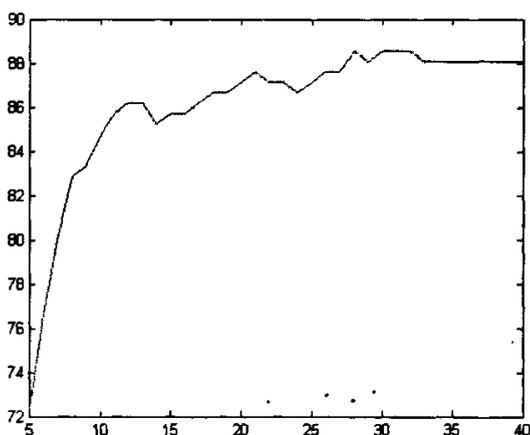


图 4-3 特征提取数目与平均分类正确率的关系图

从图 4-3 可以看出，当提取的特征数目逐渐增多时，平均分类正确率逐渐提高。但当特征数目增多至某一阈值时（在图中阈值为 33），平均分类正确率不

再提高。这意味着在表情识别中不需要提取过多的特征数目。

下面分析“亚采样间隔”大小对平均分类正确率和分类速度（即分类所花时间的）影响。让亚采样间隔分别取 2、4、6，记录当采样间隔每取一个值时对应的平均分类正确率和分类速度，如表 4-2 所示：

表 4-2 亚采样间隔与平均分类正确率、分类时间的关系表

亚采样间隔	平均分类正确率	分类所花时间（秒）
2	85.714286	68.1
4	84.761905	7.5
6	81.428571	5.0

从表 4-2 可看出，亚采样间隔的大小对平均分类正确率的影响不大，但对分类速度影响较大：当亚采样间隔从 2 增至 6 时，分类所花时间明显变短（从 68.1 秒缩短至 5.0 秒）。

在常规最近邻方法中，用欧氏距离度量样本间的距离

$$r(\mathbf{x}, \mathbf{x}_i) = \sqrt{\|\mathbf{x} - \mathbf{x}_i\|^2} \quad (4-11)$$

而在核最近邻方法中，用以下核距离

$$r_k(\mathbf{x}, \mathbf{x}_i) = \sqrt{K(\mathbf{x}, \mathbf{x}) + K(\mathbf{x}_i, \mathbf{x}_i) - 2K(\mathbf{x}, \mathbf{x}_i)} \quad (4-12)$$

度量样本间的距离。上式中 $K(\cdot, \cdot)$ 是核函数。本文采用的核函数为多项式核：

$$K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^5 \quad (4-13)$$

当采用核最近分类方法时，由表 4-3 可知道，分类正确率得到进一步的提高。

表 4-3 最近邻法和核最近邻法对七类表情的分类正确率比较

序号	表情 类型	最近邻 分类正确率 (%)	核最近邻 分类正确率 (%)
1	生气(AN)	90.000000	91.356879
2	厌恶(DI)	90.000000	91.777788
3	恐惧(FE)	76.666667	77.000000
4	高兴(HA)	90.000000	92.333335
5	中性(NE)	93.333333	94.000000
6	悲伤(SA)	76.666667	78.111425
7	惊讶(SU)	83.333333	85.333333
平均分类正确率		85.714286	87.130394

4.4 本章小结

本章利用最近邻分类器对表情进行分类:选用日本的JAFFE表情库中的210个样本来做实验,取得了较好的识别率。分类正确率随着主成分个数和亚采样间隔的变化而变化,当主成分个数增加时,分类正确率逐渐提高;随着亚采样间隔的增加,对分类正确率的影响不大,但对分类速度影响较大,分类所花时间明显变短。然后将核最近邻方法应用于表情识别,利用Mercer核把输入空间样本映射到特征空间,通过非线性映射后较好地加大样本间的区别,从而实现更为准确的分类。

第5章 基于支持向量机的表情识别

传统统计学研究的是样本数目趋于无穷大时的渐近理论，现有学习方法也多是基于此假设。但在实际问题中，样本数往往是有限的，因此一些理论上很优秀的学习方法在实际应用中却可能不尽人意。与传统统计学相比，统计学习理论(Statistical Learning Theory或SLT)是一种专门研究小样本情况下机器学习规律的理论。V.Vapnik等人从六七十年代便开始致力于此方面的研究。

支持向量机就是在统计学理论的基础上发展起来的一种新的学习方法，该方法具有良好的分类能力和较高的计算效率，不需要对原始数据进行预处理就可达到满意的效果，适合于多类分类。由于采用了结构风险最小化原则代替经验风险最小化原则，较好的解决了小样本学习的问题；又由于采用了核函数思想，把非线性空间的问题转换到线性空间，降低了算法的复杂度。正是因为其完备的理论基础和出色的学习性能，该技术已成为机器学习界的研究热点，并在很多领域都得到了成功的应用。

也正是因为这些特点，我们在表情分类的时候考虑利用支持向量机分类器。下面首先介绍一些与SVM相关的基本概念。

5.1 最优分类面

5.1.1 最优分类面

如图5-1所示，欲在样本线性可分情形下求取最优分类线(optimal hyperplane) H_2 ，该问题是下一节分析SVM基本原理的基础。 N 个独立样本 (x_1, t_1) , (x_2, t_2) , ..., $(x_N, t_N) \in R \times T$ 分属于两类，在图中分别用圆圈和方块代表。 R 为原始输入空间， $T = \{-1, 1\}$ 代表类别标号集合，并事先约定圆圈的类别标号 $t_i = -1$ ，方块的类别标号 $t_i = 1$ 。图中 H_1 和 H_3 分别为过各类样本中离 H_2 最近的点且平行于 H_2 的直线， H_1 和 H_3 间的距离被称为分类间隔(margin)。所谓最优分类线就是要求 H_2 不但能将两类样本无错误地分开，而且使分类间隔最大。推广到高维空间，最优分类线就成为了最优分类面，为叙述方便下面以最优分类面统称之。

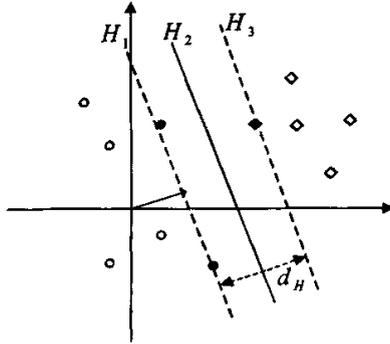


图 5-1 最优分类面示意图

若设 H_2 的方程为 $\tilde{w} \cdot x + \tilde{b} = 0$, 则 H_1 与 H_3 的方程分别为 $H_1: \tilde{w} \cdot x + \tilde{b} = -k_1$, $H_3: \tilde{w} \cdot x + \tilde{b} = k_1$. 由平行直线间的距离公式, H_1 与 H_3 间的距离为 $d_H = 2k_1 / \|\tilde{w}\|_2$. 由于 d_H 的表达式中含有两个变量, 不利于下一步优化, 因此将三个方程两端同除 k_1 , 并令 $w = \tilde{w}/k_1$, $b = \tilde{b}/k_1$, 得到

$$H_1: w \cdot x + b = -1 \quad (5-1)$$

$$H_2: w \cdot x + b = 0 \quad (5-2)$$

$$H_3: w \cdot x + b = 1 \quad (5-3)$$

此时 H_1 与 H_3 间的距离变为

$$d_H = 2 / \|w\|_2 \quad (5-4)$$

方程两端同除 k_1 起的是归一化作用, 它只改变直线方程的形式, 并不改变线的形态。

现设与 H_2 的方程 (5-2) 对应的线性判别函数表达式为

$$f(x) = w \cdot x + b \quad (5-5)$$

欲使 H_2 成为最优分类面, 必使 (5-4) 最大化, 也即使 $\frac{1}{2} \|w\|_2^2$ 或者 $\frac{1}{2} \|w\|_2^2$ 最小;

另外从图 5-1 可看出, 圆圈类和方块类样本分别位于 H_1 和 H_3 的外侧, 两条直线间不能含有样本, 这实际上是要求当 $t_i = -1$ 时, $w \cdot x_i + b \leq -1$; 当 $t_i = 1$ 时 $w \cdot x_i + b \geq 1$. 因此 (5-5) 中 w 与 b 的求取可归结为求解下面的约束优化问题

$$\min \frac{1}{2} \|w\|_2^2 \quad (5-6)$$

$$s.t. \quad t_i [w \cdot x_i + b] \geq 1 \quad i=1, 2, \dots, N$$

(5-6) 的解由下面的拉格朗日 (Lagrange) 函数的鞍点给出

$$G(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i \{t_i [(w \cdot x) + b] - 1\} \quad (5-7)$$

上式中 $\alpha_i \geq 0$, $i=1,2,K,N$ 为拉格朗日乘子。为得到鞍点令

$$\frac{\partial G}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i t_i x_i \quad (5-8)$$

$$\frac{\partial G}{\partial b} = 0 \Rightarrow \sum_{i=1}^N t_i \alpha_i = 0 \quad (5-9)$$

将 (5-8) 和 (5-9) 一起代入到 (5-7) 中可得 (5-6) 的对偶规划

$$\begin{aligned} \min G(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j (x_i, x_j) \\ \text{s.t. } &\sum_{i=1}^N t_i \alpha_i = 0 \quad \alpha_i \geq 0 \end{aligned} \quad (5-10)$$

设 $\alpha = (\alpha_1, \alpha_2, K, \alpha_i, K, \alpha_N)^T$, $d = (1, 1, K, 1)^T$, $t_N = (t_1, t_2, K, t_N)^T$, 矩阵 Q 的第 i 行第 j 列为 $Q_{ij} = t_i t_j (x_i, x_j)$, 则 (5-10) 可写成如下矩阵形式:

$$\begin{aligned} \min G(\alpha) &= d^T \alpha - \frac{1}{2} \alpha^T Q \alpha \\ \text{s.t. } &t_N^T \alpha = 0 \quad \alpha \geq 0 \end{aligned} \quad (5-11)$$

解 (5-10) 或 (5-11) 可得 α_i 的最优值, 并将 (5-8) 代入到 (5-5) 中, 得到判别函数的表达式

$$f(x) = \sum_{i=1}^N \alpha_i t_i \cdot (x, x_i) + b \quad (5-12)$$

(5-10) 或 (5-11) 是一个二次规划 (quadratic program, QP), 它存在唯一解。并且根据 Kuhn-Tucker 条件应满足

$$\alpha_i \{t_i [w \cdot x_i + b] - 1\} = 0 \quad (5-13)$$

大多数学习样本满足 $t_i [w \cdot x_i + b] > 1$, 因此根据 (5-13) 知 $\alpha_i = 0$; 只有少数样本满足 $t_i [w \cdot x_i + b] = 1$ 而使得 $\alpha_i \neq 0$, 这些样本被称为支持向量, 它们位于 H_1 或 H_2 上。在图 5-1 中用实心圆圈或方块表示支持向量。

5.1.2 广义最优分类面

上述最优分类面是在线性可分的前提下得到的, 在线性不可分的情况下上述

做法不再成立。若图中圆圈和方块交杂在一起，无法用一条直线分开，此时可仿照上述做法，仍用 H_1 、 H_2 和 H_3 来划分样本，但允许 H_1 和 H_3 间含有样本。为此给每个样本引入一个大于 0 的松弛因子 ξ_i ，并将(5-6)改写成

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (5-14)$$

$$s.t. \quad t_i [w \cdot x_i + b] \geq 1 - \xi_i, \quad \xi_i \geq 0$$

上式中 C 为某个指定的常数。

下面分析 ξ_i 的物理意义。结合 (5-14) 与 (5-5) 可推知

$$\xi_i = [1 - t_i f(x_i)]_+ \quad (5-15)$$

并在图 5-2 中绘出 ξ_i 随 $f(x_i)$ 的变化曲线。由于与 t_i 有关，因此 ξ_i 曲线有两条。当 $t_i = 1$ 时方块类样本与 H_3 的距离为 $\xi_i / \|w\|$ ，因此 ξ_i 是方块类样本与 H_3 距离远近的度量。比如图中 A 处的样本 $\xi_i = 1$ ，因此它与 H_3 的距离为 $1 / \|w\|_2$ 。类似地，当 $t_i = -1$ 时 ξ_i 是圆圈类样本与 H_1 距离远近的度量。另外从 (5-14) 可看出 ξ_i 还充当惩罚因子的角色：当方块类样本出现在 H_1 和 H_3 之间时，它距 H_3 越远，对其惩罚的力度越大；当圆圈类样本出现在 H_1 和 H_3 之间时，它距 H_1 越远，对其惩罚的力度越大。

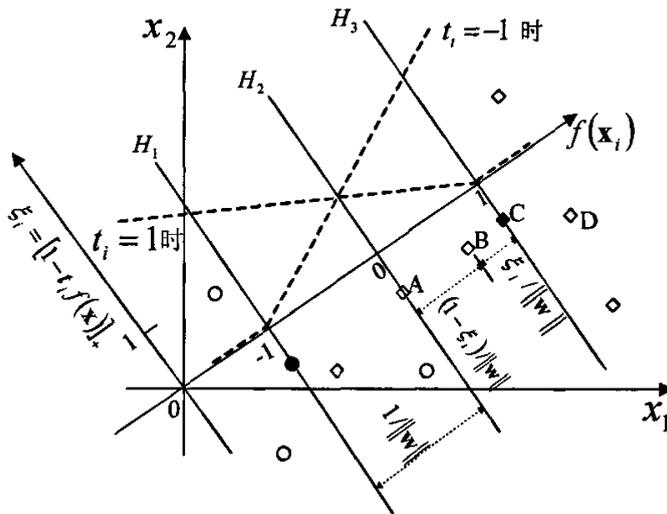


图 5-2 广义最优分类面中松弛因子的物理意义示意图

(5-14) 的解由下面的拉格朗日函数的鞍点给出

$$G(w, b, a, \tau) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i \{t_i [(w \cdot x_i) + b] - 1 + \xi_i\} - \sum_{i=1}^N \tau_i \xi_i \quad (5-16)$$

上式中 $\alpha_i \geq 0$, $\tau_i \geq 0$, $i = 1, 2, \dots, N$ 为拉格朗日乘子。分别对上式中的 w 、 b 与 ξ_i 求

偏导可得 (5-8)、(5-9) 与下式

$$C - \alpha_i - \tau_i = 0 \quad (5-17)$$

将 (5-8)、(5-9) 和 (5-17) 一起代入到 (5-16) 中同样得到如下二次规划

$$\begin{aligned} \min G(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j (x_i, x_j) \\ \text{s.t. } \sum_{i=1}^N t_i \alpha_i &= 0 \quad C \geq \alpha_i \geq 0 \end{aligned} \quad (5-18)$$

上式与 (5-10) 相比只不过是约束条件中的 $\alpha_i \geq 0$ 变成了 $C \geq \alpha_i \geq 0$ 。

5.2 支持向量机的基本原理

支持向量机是一种泛化能力很强的分类器，它在解决小样本问题方面表现出了许多特有的优势，已成为国际上模式识别领域新的研究热点^[45, 46]。

支持向量机是一种基于统计学习理论的机器学习算法。其主要的内容是：当问题是线性可分时，给出一个求解最大间隔解的方法；当问题不是线性可分时，提出利用核函数将样本集映射到某一高维空间，使得样本集在高维空间的像是线性可分的。

如图 5-3 (a) 所示，如果 N 个独立样本 $(x_1, t_1), (x_2, t_2), \dots, (x_N, t_N) \in R \times T$ 不是线性可分的，这时可以通过一个未知的映射函数 $\varphi: R \rightarrow F$ 将样本映射到某特征空间 F 中，得到 $(\varphi(x_1), t_1), (\varphi(x_2), t_2), \dots, (\varphi(x_N), t_N) \in F \times T$ ，然后在 F 中构造线性最优分类面（图 (b)），从输入空间 R 的角度看，所得最优分类面是非线性的（图 (c)），SVC 正是基于上述思路。

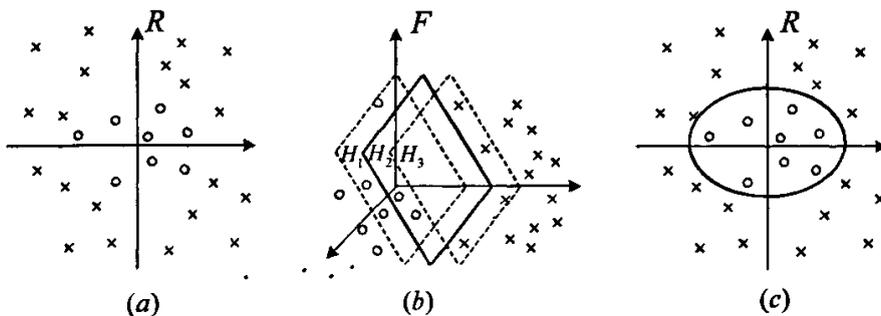


图 5-3 SVM 基本原理示意图

(a) 输入空间 R 中的 N 个待分类样本 (b) 样本被映射到特征空间 F 中，并构建最优分类面 (c) 从输入空间 R 的角度看，分类面是非线性的

在 F 中构造线性最优分类面的方式与上节在输入空间 R 中构造最优分类面的方式是一样的。与 (5-5) 类似, 设与 H_2 对应的线性判别函数表达式为

$$f(x) = w \cdot \varphi(x) + b \quad (5-19)$$

上式中求取 w 与 b 可归结为求解下面的约束优化问题

$$\min \frac{1}{2} \|w\|_2^2 \quad (5-20)$$

$$s.t. \quad t_i [w \cdot \varphi(x_i) + b] \geq 1 \quad i = 1, 2, \dots, K, N$$

(5-20) 与 (5-6) 很相似, 它的解由下面的拉格朗日函数的鞍点给出

$$G(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i \{t_i [(w \cdot \varphi(x_i)) + b] - 1\} \quad (5-21)$$

上式中 $\alpha_i \geq 0$, $i = 1, 2, \dots, K, N$ 为拉格朗日乘子。为得到鞍点令

$$\frac{\partial G}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i t_i \varphi(x_i) \quad (5-22)$$

$$\frac{\partial G}{\partial b} = 0 \Rightarrow \sum_{i=1}^N t_i \alpha_i = 0 \quad (5-23)$$

将 (5-22) 和 (5-23) 一起代入到 (5-21) 中可得 (5-20) 的对偶规划

$$\max -G(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j \varphi(x_i) \varphi(x_j) \quad (5-24)$$

$$s.t. \quad \sum_{i=1}^N t_i \alpha_i = 0 \quad \alpha_i \geq 0$$

上式与 (5-10) 形式相同, 只是含有映射函数 φ 。

若在 F 中定义内积

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) \quad (5-25)$$

将之代入到 (5-24) 后有

$$\max -G(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j K(x_i, x_j) \quad (5-26)$$

$$s.t. \quad \sum_{i=1}^N t_i \alpha_i = 0 \quad \alpha_i \geq 0$$

又设 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$, $d = (1, 1, \dots, 1)^T$, $t_N = (t_1, t_2, \dots, t_N)^T$, 矩阵 Q 的第 i 行第 j 列为 $Q_{ij} = t_i t_j K(x_i, x_j)$, 则 (5-26) 仍可写成如 (5-11) 所示的矩阵形式。解 (5-26) 可得 α_i 的最优值, 并将 (5-22) 代入到 (5-19) 中可得 F 中的线性判别函数表达式

$$f(x) = \sum_{i=1}^N \alpha_i t_i K(x, x_i) + b \quad (5-27)$$

式(5-25)中的 $K(x, x_i)$ 被称为核函数，它欲成为 F 中的内积必须满足 Mercer 条件：使得积分算子 $(T_k f)(x) = \int K(x, x_j) f(x_j) dx_j$ 为正。如果 F 的维数很高，(5-25)等号右边的计算量会很大，甚至会陷入“维数灾难”而使得计算不可行；但通过上述代换， F 中的内积可基于 R 中的变量通过给定的核函数直接计算得到，即使 F 的维数非常高，SVM 本身也并没有增加多少计算复杂度。特别是对某些映射函数而言， F 的维数是无限的，此时内积必须用积分来计算，这种代换的作用就更为明显。另外从以上推导过程可看出，在特征空间中构造最优分类面，并不需要以显式形式来考虑特征空间，也没有必要知道映射函数的形式，只需通过(5-25)计算内积即可。

值得指出的是，以上推导的假设前提是 N 个样本被映射到 F 中以后能够线性可分，如果该前提不满足，则必须在 F 中构造广义最优分类面。相应地(5-20)要改成

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (5-28)$$

$$s.t. \quad t_i [w \cdot \varphi(x_i) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0$$

上式与(5-14)很相似，只是式中含有 φ 。

在(5-27)中，假设输入空间 R 内的向量为 $x = (x^{(1)}, x^{(2)}, \dots, x^{(d)})^T$ ，另有 s 个向量 x_1, x_2, \dots, x_s 对应的 α_i 不为零，它们被称为支持向量，则 SVC 的拓扑结构如图 5-4 所示，从图中看出 SVM 的计算复杂度取决于支持向量的数目。

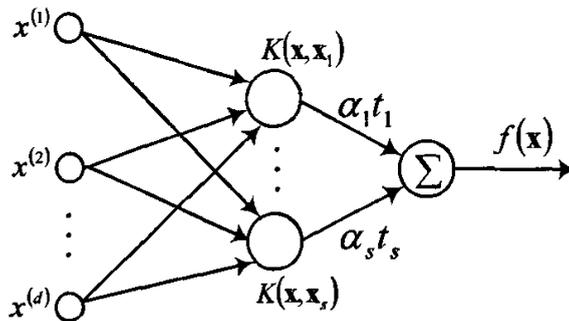


图 5-4 SVC 拓扑结构图

最初，SVM方法是针对二类别分类提出来的，而表情识别一般是多类别($K>2$)的分类问题，因此需要将SVM方法扩展到多类别的情形。主要有两种策略来求解多类别问题，一种是称为“一对一(one-against-one)”的方法，另外一种解决多类别分类SVM算法是所谓的“一对多(one-against-the-rest)”方法。

5.3 组合型多值分类支持向量机

多类别样本分类问题和两类别样本分类问题存在一定的对应关系，如果一个分类问题 K 类可分，则这 K 类中任何两类一定可分；反之在一个 K 类可分问题中，如果已知其任意两两可分，则可通过一定的组合法则，由多个两两可分最终实现 K 类可分。所谓组合型就是将多个 SVM 通过一定的组合法则组合在一起解决多类别样本分类问题^[47]。常见的组合型有以下两种：

5.3.1 one-against-rest方法

最早实现 SVM 对多类别进行分类的方法就是 one-against-rest。对 k 个类别的样本数据 one-against-rest 方法就构造 k 个 SVM 二值子分类器。

在构造第 i 个 SVM 子分类器时，将属于第 i 类别的样本数据标记为正类，其它不属于 i 类别的样本数据标记为负类。因此，第 i 个 SVM 子分类器所需解决的问题就是：

$$\begin{aligned} \min_{w', b', \xi'} & \frac{1}{2} (w')^T w' + C \sum_{j=1}^l \xi_j' \\ \text{s.t.} & (w')^T \phi(x_j) + b' \geq 1 - \xi_j' \quad \text{if } y_j = i, \\ & (w')^T \phi(x_j) + b' \leq -1 + \xi_j' \quad \text{if } y_j \neq i, \\ & \xi_j' \geq 0, i = 1, K, l \end{aligned} \tag{5-29}$$

此处训练数据 x_j ，由函数 ϕ 映射到高维特征空间， C 是惩罚系数。求解 (5-29) 的对偶问题可得到 k 个判别函数：

$$\begin{aligned} & (w^1)^T \phi(x) + b^1 \\ & \quad \cdot \\ & \quad \cdot \\ & \quad \cdot \\ & (w^k)^T \phi(x) + b^k \end{aligned} \tag{5-30}$$

测试时，对测试数据分别计算各判别函数值，并选取判别函数值最大所对应的类别为测试数据的类别。

$$\text{class of } x = \arg \max_{i=1, K, k} ((w^i)^T \phi(x) + b^i) \tag{5-31}$$

但此方法的泛化错误(the generalization error)是无界的；并且由于各个

样本都要参与所有 SVM 子分类器的训练，所以训练时间关于 k 线性增长。

5.3.2 one-against-one方法

one-against-one 方法首先由 Knerr 提出。分别选取两个不同类别构成一个 SVM 子分类器，这样共有 $k(k-1)/2$ 个 SVM 子分类器。在构造类别 i 和类别 j 的 SVM 子分类器时，从样本数据集中选取属于类别 i 、类别 j 的样本数据作为训练样本，并将属于类别 i 的样本标记为正，将属于类别 j 的样本标记为负，对应的二值分类问题为：

$$\min_{w^y, b^y, \xi^y} \frac{1}{2} (w^y)^T w^y + C \sum \xi_i^y \quad (5-32)$$

$$(w^y)^T \phi(x_i) + b^y \geq 1 - \xi_i^y, \text{ if } y_i = i$$

$$(w^y)^T \phi(x_i) + b^y \geq -1 + \xi_i^y, \text{ if } y_i = j$$

$$\xi_k^y \geq 0$$

对于 one-against-one 测试方法有多种，Knerr 使用“与”门把所有的二值分类器组合起来测试样本。使用最多的是 Friedman 提出的投票策略(voting strategy)：将测试数据 x 对 $k(k-1)/2$ 个 SVM 子分类器分别进行测试，若由 $sign((w^y)^T \phi(x) + b^y)$ 得到 x 属于第 i 类，则第 i 类投票数加 1，属于第 j 类，则第 j 类投票数加 1。累计各类别的得分，选择得分最高者所对应的类别为测试数据的类别。若存在多个最高得分的类别，从而出现不可区分区域，此时可选择小序号类作为测试数据的类别。

One-against-one 方法最明显的缺点就是，每个子分类器必须都要非常仔细地调整，如果某个子分类器不规范化，则整个 k 类分类系统将趋于过学习。并且“与”门组合和投票策略方法对于系统的推广误差都是无界的。再者，所需构造和测试的二值分类器的数量关于 k 成超线性增长。

5.4 基于SVM的表情识别

5.4.1实验过程

由于采用 one-against-rest 组合方式解决 7 类表情识别问题，根据 one-against-rest 组合原则，需要构建 7 个支持向量机，将其进行组合。第 1 个

支持向量机将属于“生气(AN)类”表情的样本与属于“非生气(AN)类”表情的样本分开，其中属于“生气(AN)类”表情的样本被称为“正(positive)”学习样本，属于“非生气(AN)类”表情的样本被称为“反(negative)”学习样本。以此类推，第 2 个支持向量机将属于“厌恶(DI)类”表情的样本与属于“非厌恶(DI)类”表情的样本分开，第 3 个支持向量机将属于“恐惧(FE)类”表情的样本与属于“非恐惧(FE)类”表情的样本分开，依次类推。

考虑到支持向量机涉及到使用正反两类学习样本，因此对每个支持向量机而言，设置其“正”学习样本数为 20，“反”学习样本数为 60。例如对第 1 个支持向量机而言，其 20 个“正”学习样本如图 5-5 所示，60 个“反”学习样本如图 5-6 所示。

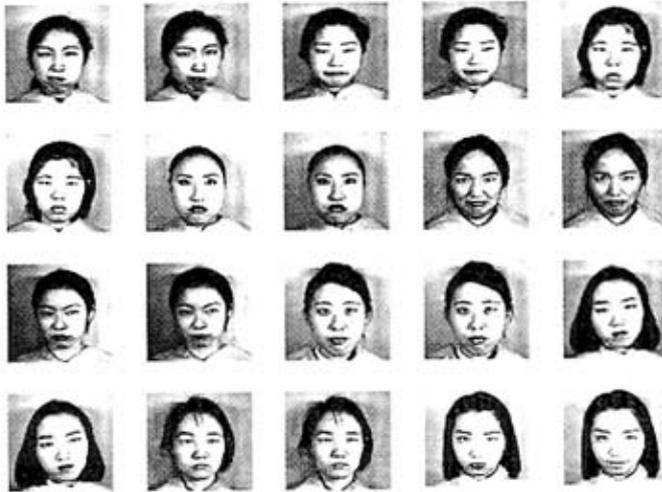


图 5-5 第 1 个支持向量机的“正”学习样本

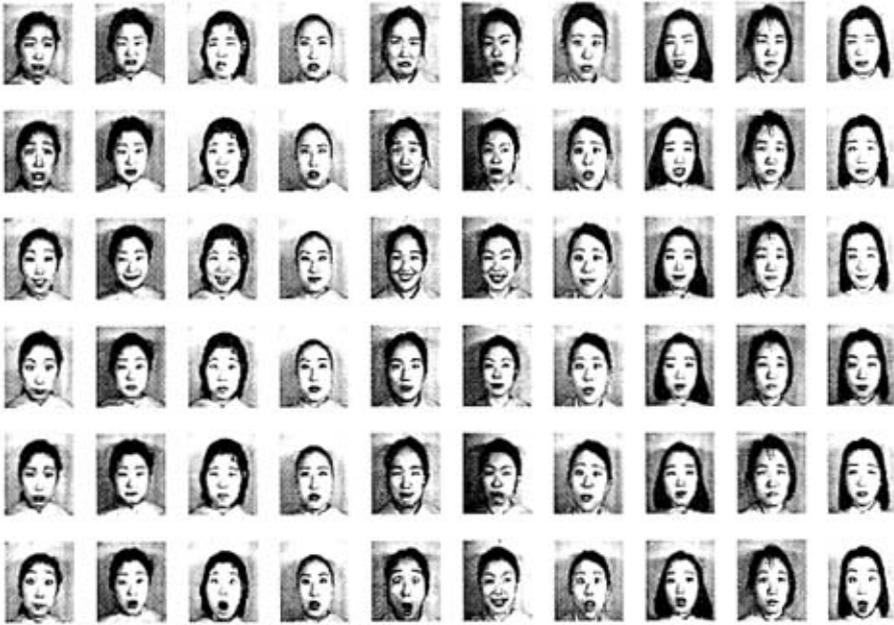


图 5-6 第 1 个支持向量机的“反”学习样本

当用 SVM 进行表情识别时，其测试样本的组成方式与最近邻方法一样。即考虑到 7 种表情分别显示在 10 位日本年轻女性脸上，每位女性有 3 至 4 幅表情图。现每位女性取 3 幅图作为测试样本，因此测试样本总数为 210。

5.4.2 实验结果及分析

支持向量机对七类表情的分类正确率如表 5-1 所示，从该表看出，支持向量机对“中性 (NE)”表情的分类正确率最高，对“恐惧 (FE)”和“悲伤 (SA)”表情的分类正确率最低，这一点与最近邻方法的分类结果一致。对“中性 (NE)”表情的分类正确率最高，是因为该类表情与其它类表情差异较大。而对“恐惧 (FE)”和“悲伤 (SA)”表情的分类正确率最低，是因为这两类表情与其它类表情差异较小，而且表情本身的特征不明显。另外根据平均分类正确率“94.761905%”来看，支持向量机的分类结果要优于最近邻方法。

表 5-1 支持向量机对七类表情的分类正确率表

序号	表情类型	分类正确率 (%)
1	生气(AN)	93.333333
2	厌恶(DI)	93.333333
3	恐惧(FE)	90.000000
4	高兴(HA)	96.666667
5	中性(NE)	100.000000
6	悲伤(SA)	93.333333
7	惊讶(SU)	96.666667
平均分类正确率		94.761905

5.5 改进的基于SVM的表情识别

5.5.1 改进原理

在用 7 个支持向量机进行组合来解决 7 类表情识别问题时，我们将 7 个支持向量机共用同一个核参数 σ ，这样就不能保证 σ 的取值对所有支持向量机均合适。而不合适的核参数会降低分类正确率，因此有必要对 σ 的取值方式进行改进。

一种可行的思路是使各 SVM 的核参数能随学习样本自适应变化，即根据学习样本自适应地估计各 SVM 的核参数。具体做法是，对第 i 个 SVM 的核参数 σ_i ，其估计值 $\hat{\sigma}_i$ 由下式给出：

$$\hat{\sigma}_i = \sqrt{\frac{\|x_i - \bar{x}_i\|^2}{M}} \quad (5-33)$$

上式中 x_i 为第 i 个 SVM 对应的学习样本（已做过特征提取）， \bar{x}_i 为 x_i 的均值， M 为 x_i 的维数。

经过上述改进后，支持向量机的核参数可以自适应变化，这样每个支持向量机均采用一个最合适的核参数，程序灵活性大为提高，同时可以预计分类正确率也会提高。

5.5.2 实验结果及分析

为公平起见，支持向量机用于表情识别时，在方法改进前后程序的参数设置保持不变。现将改进前后的分类结果列于表 5-2 中，通过对比两个分类结果可以得出如下结论：支持向量机经改进后，平均分类正确率得以提高。

表 5-2 支持向量机改进前后对七类表情的分类正确率

序号	表情类型	改进前 分类正确率 (%)	改进后 分类正确率 (%)
1	生气(AN)	93.333333	100.000000
2	厌恶(DI)	93.333333	100.000000
3	恐惧(FE)	90.000000	86.666667
4	高兴(HA)	96.666667	90.000000
5	中性(NE)	100.000000	100.000000
6	悲伤(SA)	93.333333	93.333333
7	惊讶(SU)	96.666667	96.666667
平均分类正确率		94.761905	95.238095

5.6 两种识别方法的简要比较

本文已经分别用最近邻与支持向量机等两种方法进行表情识别，现对两种方法进行比较。“有比较才有鉴别”，通过比较，很容易就知道两种方法的优劣以及适用场合。

为公平起见，在比较两种分类方法时二者参数设置相同：

取测试样本的个数均为 210 个，提取的特征数目为 11，亚采样间隔取 4，亚采样网格的水平位移取 70，图像经亚采样后的大小为 32×32；

但是，两种分类方法的学习样本数目没法统一，因为支持向量机还涉及到使用“反”学习样本。因此本文设置最近邻的学习样本数为 70；单个支持向量机的学习样本数分为：“正”学习样本数为 20：“反”学习样本数为 60。

现将两种分类方法对七类表情的分类结果列于表 5-3 中。

表 5-3 两种分类方法对七类表情的分类正确率比较

序号	表情类型	最近邻方法 分类正确率 (%)	支持向量机 分类正确率 (%)
1	生气 (AN)	90.000000	93.333333
2	厌恶 (DI)	86.666667	93.333333
3	恐惧 (FE)	80.000000	90.000000
4	高兴 (HA)	83.333333	96.666667
5	中性 (NE)	93.333333	100.000000
6	悲伤 (SA)	73.333333	93.333333
7	惊讶 (SU)	86.666667	96.666667
平均分类正确率		84.761905	94.761905

从表 5-3 中可以看出，最近邻与支持向量机各有千秋。最近邻方法的优点是分类速度快，劣势是分类正确率相对较低；而支持向量机刚好与之相反，其优点是分类正确率高，劣势是分类速度相对较慢。

在有些应用场合中，要求快速实时分类，而对分类正确率要求不高，此时可考虑使用最近邻方法；而在另外一些应用场合中，要求分类很准确，而对分类速度不作要求，此时可考虑使用支持向量机。

将支持向量机理论用于表情的识别，是很有实际意义的，既可以解决非线性分类的问题又适合小样本、高维的模式分类问题，是非常值得研究和重视的。

5.7 本章小结

本章介绍了最优分类面、广义最优分类面的概念，由此引出支持向量机并讨论了多类 SVM 分类器的两种构造方法。本章构造了 7 个支持向量机，并基于 one-against-rest 原则对 7 个支持向量机进行组合，然后用于人脸表情识别。另外在此基础上对原有 SVM 进行了改进，对 7 个支持向量机不再共同一个核参数，而是根据学习样本自适应地估计各 SVM 的核参数，实验结果表明，SVM 分类器有较强的推广能力，对于小样本训练集分类问题有自身的优势。

第6章 总结与展望

人脸表情是人体语言的一部分，表达了人的丰富情感，在人们的日常交流中起着至关重要的作用。人类肉眼识别人脸表情信息时几乎没有什么困难，但对计算机视觉系统来说则非常不容易。用计算机来分析面部表情是一个非常复杂的问题，它关键在于建立一个人类情绪的模型，和把它们同人脸面部特征与表情的变化联系起来。但由于人脸是个柔性体，而不是刚体，因此很难用模型来精确描述。

在自然化的人机交互过程中，计算机不能没有理解和表达人类情感的能力。这种能力对促进计算机视觉系统建模和数据库的发展都有直接的作用：在语言学中，可辅助唇读；在行为学中，能帮助人们研究和建立交流中的可信度；在商业应用、可视电话和电视会议方面以及国际间商业政治的交流方面都有着重要的应用价值。如果能够让机器很好地理解人脸表情的变化及其对应的心理活动，将极大地改变人机之间的关系，使得机器能够更好地为人类服务。因此，用计算机模仿人工智能来识别人脸表情是一个极富意义的课题。

近年来，人脸表情识别逐渐成为模式识别、认知科学、计算机图形学等学科的研究热点。人脸表情识别是一个具有挑战性的课题，现在还处于研究阶段，各种算法还不很十分成熟。人脸表情识别大致可分为四部分：人脸图像的获取、图像预处理、表情特征提取和表情分类。其中特征的提取和分类是最重要的工作。

6.1 研究内容总结

本文提出了亚采样的表情预处理方法，实验选取日本的 JAFFE 表情数据库，经过亚采样处理，去掉了头发、颈部和背景等与表情无关信息，同时降低了图像的维数，将像素降低为 64×64 ，并将图像的每个像素亮度值都转换到 $[0, 1]$ 间，利于后续的特征提取。接着对表情图像进行直方图均衡化，经过均衡化后，图像的细节更加清楚，直方图各灰度等级的分布更加平均。然后采用主成分分析进行特征提取，采用欧氏距离作为模式特征之间的相似性度量，用最近邻分类器进行分类，取得了较好的识别效果，并将核最近邻应用于表情识别，利用 Mercer 核把输入空间样本映射到特征空间，通过非线性映射后较好地加大样本间的区别，从而实现更为准确的分类。

然后将机器学习领域中的支持向量机理论应用于人脸表情识别。其实质是通过非线性变换将输入变量变换到一个高维空间，然后在这个新空间中求取最优线性分类面而这种非线性变换是通过定义适当的内积函数实现的。我们将两类问题的 SVM 扩展到多类分类，用 one-against-rest 方法对人脸表情采用的方法进行识

别，得到了很好的识别效果，

本文将 PCA 与 SVM 相结合，充分利用了 PCA 在特征提取方面的有效性以及 SVM 在处理小样本问题和泛化能力强等方面的优势，并通过改进 SVM，使其核参数可以自适应变化，这样每个支持向量机均采用一个最合适的核参数，程序灵活性大为提高，所提出的算法具有较高的识别率，实验结果也验证了本算法的有效性。但由于部分训练样本集也用于测试集，故识别率较好。

有些应用场合中，要求快速实时分类，而对分类正确率要求不高，此时可考虑使用最近邻方法；而在另外一些应用场合中，要求分类很准确，而对分类速度不作要求，此时可考虑使用支持向量机。

图像处理和识别本身是一个实践性很强的研究工作，很多理论的应用效果需要大量实验的证明，特别是参数的选择，往往不能完全靠理论推导。当然，由于时间与本人水平有限，在论文研究中还存在着许多不足之处，需要在今后的进一步研究中不断完善和弥补。

总之，人脸表情识别研究是一个新兴的且难度很大的研究课题，它的内容相当丰富和广博，国内外很多知名的研究机构都在从事这方面的研究工作。本人从事人脸表情识别研究时间不长，论文所提出的方法也只是一种尝试。而且因为 JAFFE 表情库的规范性比较好，该识别方案中的预处理方法是针对这个特定的表情库设计的，所以实验效果比较好。对于其他有干扰信息存在或光照变化的图像，识别效果肯定要受影响，因此提高人脸表情识别的鲁棒性也是将来要研究的方向之一。

6.2 未来展望

曲波是最新的多尺度几何分析工具。曲波变换直接把边缘作为基本的表示元素，是各向异性的，具有很强的方向性。它是一个多分辨率、带通和方向函数的分析方法，对更有效地描绘图像的边缘和曲线特征是有用的。

对于人脸表情图像，我们更多的关心表达人脸表情主要特征的直线和曲线。按照这个观点，曲波变换首先会突出人脸表情识别中图像的曲线特征，因而它能更好的刻画人脸表情图像。曲波方法提供了一个更好的解决方案来描述曲线特征。根据曲波变换能更好的处理曲线特征和包含曲线的人脸表情。

可以通过使用基于第二代曲波的快速离散曲波变换(FDCT)来提取脸部表情特征。由于时间紧迫，未能实现之，但我预测，曲波应用于表情识别中应该有良好的前景。

人脸表情识别是一个具有挑战性的课题，尽管已经取得了许多重大成果，但

目前还处于研究阶段，依然存在很多问题，比如：人脸是一个柔性体而不是刚体，很难建立精确的模型，受光照，姿态等因素的影响，尽管对图像进行了预处理，但是要精确提取表情特征还是有一定的难度。而且表情识别是人类视觉独特的过程，要实现计算机识别需要结合各门学科的知识为之服务，目前这些问题都是表情识别领域亟待解决的问题。

(1)细微表情提取的研究。人脸表情变化细微并且复杂，细节是区分人脸表情的关键。如何精确提取表情的细微变化，是表情特征提取的关键，也是今后的一个研究重点。

(2)三维表情识别研究。提取面部表情时要尽量分离头部的刚性运动和脸部的柔性运动，为了解决这个问题可引入三维表情信息，三维图像能提供较二维图像更能提供完整而真实的内容，且减少了光照和姿态的变化影响，如何有效利用三维信息进行表情识别将是一个具有挑战性的研究课题。

(3)多特征方法信息融合的表情识别研究。任何一种表情识别的方法都有其局限性，不能适用于任意环境下的表情识别。因此最大限度的有机结合各种信息方法（语音识别、姿态识别等）来提高了表情识别率。

(4)表情数据库的建立。人脸表情识别技术的进步离不开表情库的建立，它可以为各种算法的测试提供方便。但各种文献中提出的算法并不都是在同一个表情数据库当中测试的，因此无法比较各种算法之间的优劣，故建立一个通用的表情数据库成为以后要解决的问题之一。

(5)综合各学科的研究。人类从生理、心理、计算机等方面来研究人脸面部表情，对其进行识别时，应该结合各类相关学科，充分发挥计算机的强大功能和人类的智慧。

(6)表情识别的鲁棒性。目前的研究多是在特定条件下来进行识别，改善识别算法，使之能应用于任意条件下的图像中。

参考文献

- [1]M. Pantic, L. J. M. Rothkranz. Expert system for automatic analysis of facial expression. Image and computing[J], 2000, 18:881-905.
- [2]C. Darwin. The Expression of the Emotions in Man and Animals, J. Murray, London, 1872.
- [3]Ekman P, Friesen W V. Facial action coding system(FACS):manual[M]. Consulting Psychologists Press, 1978.
- [4]Ekman P, Rosenberg E, Hager J, Facial action coding system affect interpretation database (FACSAID). <http://www.yahoo.com>.
- [5]Suwa M, Sugie N, Fujimora K. A preliminary note on pattern recognition of human emotional expression[C]. In: Proc 4th Int Joint Conference on Pattern Recognition, 1978, 408-410.
- [6]王志良, 陈锋军, 薛为民. 人脸表情识别方法综述[J]. 计算机应用与软件, 2003, 20(12):63-66.
- [7]K. Mase, A. Pentland. Recognition of facial expression from optical flow[J]. IEICE Trans, 1991, E74(10):3474-3483.
- [8]Essa Irfan A. Coding, Analysis, Interpretation, and recognition of Facial Expressions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997-07, 19(7):757-763.
- [9]金辉, 高文. 人脸面部混合表情识别系统[J]. 计算机学报, 2000, 23(6):602-608.
- [10]文沁, 汪增福. 基于三维数据的人脸表情识别[J]. 计算机仿真, 2005, 22(7):99-103.
- [11]Fasel B, Luetttin J. Automatic Facial Expression Analysis:a Survey[J]. Pattern Recognition, 2003, 36:259-275.
- [12] Andrew J, Calder A, Burton M, et. al. A Principal Component Analysis of Facial Expressions[J]. Vision research, 2001, 1(9): 1179-1208.
- [13]Bartlett M S, Lades H M, and Sejnowski T J. Independent component representations for face recognition[J]. In: Proceedings of SPIE 1998:528-539.
- [14]徐正光, 闫恒川, 张利欣, 独立成分分析在表情识别中的应用[J], 微计算机信息, 2006, 22(17):287-289.
- [15]Rajkiran Gottumukkal, Vijayan K. Asari. An improved face recognition

technique based on modular PCA approach[J]. Pattern Recognition Letters, 2004, 25(4):429-436.

[16]程剑, 应自炉. 基于二维主分量分析的面部表情识别[J]. 计算机工程与应用, 2006(5):32-33.

[17]唐京海, 张有为. 基于 FLD 特征提取的 SVM 人脸表情识别方法[J]. 计算机工程与应用, 2006(11):10-12.

[18]叶敬福, 詹永照. 基于 Gabor 小波变换的人脸表情特征提取[J]. 计算机工程. 2005, 31(15):172-174.

[19]徐正光, 沈晓冬, 张利欣. 基于小波变换和神经网络集成的人脸表情识别[J]. 微电子学与计算机, 2006, 23(7):144-147.

[20]朱健翔, 苏光大, 李迎春. 结合 Gabor 特征与 Adaboost 的人脸表情识别[J]. 光电子·激光, 2006, 17(8):993-998.

[21]曹宇佳, 郑文明等. 基于差值模板特征的表情识别方法[A]. 第十二届全国图形图象学学术会议, 2005:263-267.

[22]Huang C L, Huang Y M. Facial expression recognition using model-based feature extraction and action parameters classification[J]. Journal of Visual Communication and Image Representation, 1997, 8(3):278-290.

[23]章品正, 王征, 赵宏玉. 面部表情特征抽取的研究进展[J]. 计算机工程与应用. 2006(9):38-41.

[24]G. Edwards, T. Cootes, C. Taylor. Face recognition using active appearance models[C]. Proceedings of Fifth European Conference on Computer Vision(ECCV), University of Freiburg, Germany, 1998:581-695.

[25]左坤隆, 刘文耀. 基于活动外观模型的人脸表情分析与识别[J]. 光电子·激光. 2004, 15(7):853-857.

[26]Horn B, Schunck B. Determining optical flow. Artificial Intelligence, 1981, 17(1-3):185-203.

[27]余棉水, 黎绍发. 基于光流的动态人脸表情识别[J]. 微电子学与计算机, 2005, 22(7):113-115.

[28]Y. Tian, T. Kanade, J.F. Cohn. Recognizing action units for facial expression analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2):97-115.

[29]T. Otsuka, J. Ohya. Extracting facial motion parameters by tracking feature points[C]. Proceedings of First International Conference on

- Advanced Multimedia Content Processing, Osaka, Japan, 1998:442-453.
- [30]Ma, Khorasani.Facial Expression Recognition Using Constructive Feedforward Neural Networks[J]. IEEE Transactions on Systems, Man and Cybernetics, PartB, 2004, 34(3):1588-1595.
- [31]刘松, 陈茜. 基于信息融合的面部表情识别[J]. 计算机工程与应用, 2006(12):60-62.
- [32]刘松, 应自炉. 基于局部特征和整体特征融合的面部表情识别[J]. 电子技术应用, 2005, 3:4-6.
- [33]Spiros V Ioannou, Amaryllis T Raouzaïou, Vasilis A Tzouvaras, Theofilos PMAïlis, Kostas C Karpouzis, Stefanos D Kollias. Emotion recognition through facial expression analysis based on a neurofuzzy network[J]. Neural Networks, 2005, 18(4):423-435.
- [34]尹星云, 王洵, 董兰芳. 用隐马尔可夫模型设计人脸表情识别系统[J]. 电子科技大学学报, 2003, 32(6):725-728.
- [35]张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1):32-42.
- [36]Sim T, BakerS, BsatM. Proceedings of Automatic Face and Gesture Recognition[J]. The CMU Pose, Illumination, and Expression(PIE) database, ashington, DC, USA, 2002, 9(6):53-58.
- [37]Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1):71-86.
- [38]吴丹, 林学阁. 人脸表情视频数据库的设计与实现[J]. 计算机工程与应用, 2004, 5:177-180
- [39]张庆凯. 人脸表情视频数据库系统的实现及相关算法研究[D], 沈阳: 东北大学, 2005.
- [40]边肇祺, 张学工等. 模式识别[M]. 第二版. 北京: 清华大学出版社, 2000.
- [41]马晓普, 关于人脸识别系统中的 PCA 算法[D]. 成都: 电子科技大学, 2004.
- [42]E. Fix and J.L. Hodges. Discriminatory analysis: nonparametric discrimination: small sample performance[R]. USAF School of Aviation Medicine, 1952, 11:280-322.
- [43]T.M. Cover and P.E. Hart. Nearest neighbor pattern classification [J]. IEEE Trans. Information Theory, 1967, IT-13(1):21-27.
- [44]周亚同. 基于再生核和贝叶斯推理的机器学习方法研究[D]. 西安: 西安交通大

学, 2006.

[45] 张小云. 关于分类问题的支持向量机[D]. 西安: 西安交通大学, 2001.

[46] C. J. C. Burges. A tutorial on support vector machines for pattern recognition[J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121-167.

[47] 邓乃扬, 田英杰. 数据挖掘中的新方法——支持向量机[M]. 北京: 科学出版社, 2004.

致 谢

在陕西师范大学攻读硕士学位的学习生活即将结束，在做论文的这段时间里，我的老师、同学、朋友和家人给予我的关心，指导和帮助，借此硕士论文完成之际，谨向他们致以我最真诚的感谢！

首先衷心感谢我的导师王映辉教授。王老师学识渊博、治学严谨、平易近人。他那精湛的专业知识、求实的科研精神、严谨的治学态度、诚信的为人风尚、踏实的工作作风，无论是治学还是为人，都是我终生学习的榜样。在三年的硕士研究生学习过程中，他在生活上、学习上给予我极大的帮助，再次感谢导师给予的悉心教诲和无微不至的关怀。

特别感谢西安交通大学的李小和博士、周亚同博士，感谢他们给我耐心的指导和不断的鼓励，是他们的无私帮助，给我的学习和研究带来了诸多的便捷。

感谢赵志博、李锋、王郑耀在我写论文期间给予的热情帮助！

感谢叶飞、聂小颖、马慧、赵燕妮、王丽萍、张莹、刘丁、田晓辉在我最艰难的时刻给予的鼓励和默默的支持，向他们表示最衷心的感谢！

感谢单位的领导和同事在我脱产学习期间的支持和帮助。

最后我要感谢我的父母和家人，感谢他们这么多年来一直给我的关爱、理解和支持！

攻读学位期间的研究成果

[1] 李文. 人脸表情识别方法[J]. 电子科技, 2007, (6) :63-68.

人脸表情识别研究

作者:

李文

学位授予单位:

陕西师范大学

本文链接: http://d.g.wanfangdata.com.cn/Thesis_Y1167028.aspx