

学校编码：10384

分类号_____密级_____

学 号：24320071154210

UDC_____

厦门大学

硕士 学位 论文

基于 LBP 特征的人脸表情分析研究

Facial Expression Analysis Based on LBP Feature

黄纯得

指导教师姓名：龙 飞 副教授

专业名称：计算机软件与理论

论文提交日期：2010 年 5 月

论文答辩时间：2010 年 月

学位授予日期：2010 年 月

答辩委员会主席：_____

评 阅 人：_____

2010 年 5 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下, 独立完成的研究成果。

本人在论文写作中参考其他个人或集体已经发表的研究成果, 均在文中以适当方式明确标明, 并符合法律规范和《厦门大学研究生学术活动规范 (试行)》。

另外, 该学位论文为()课题(组)的研究成果, 获得()课题(组)经费或实验室的资助, 在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称, 未有此项声明内容的, 可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

- () 1. 经厦门大学保密委员会审查核定的保密学位论文，于
年 月 日解密，解密后适用上述授权。
() 2. 不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人（签名）：

年 月 日

摘要

近几十年来，人脸表情分析与识别已经成为了一个非常活跃的研究领域，这个领域的包含了众多的理论与技术的研究，并且具有非常广阔的应用潜力。人脸的表情自动识别技术的研究涉及到心理学，解剖学，在计算机方面涉及到模式识别理论与技术，应用方面需要良好的软件工程理论支持。从解剖学角度来说，人脸的表情是由于脸部局部肌肉的收缩与舒张组成的整体变化构成的，这些变化一般持续的时间都很短，最多持续几秒种，一般情况下是非常短暂的，有些肌肉变化仅能用毫秒来衡量。脸部肌肉变化决定着某些特定的表情成分，如眼皮的上下褶皱强度变化，嘴角的拉伸等。通过分析这些局部变化与整体表情的关系，人们就可以找到这些微表情在特定情况下所反应的意义。

在理论研究方面，表情自动识别与人脸的身份识别有很多共同的地方，如分类技术，特征表示方式等，但它们之间也存在一些不同的地方。如：一般来说，人脸身份识别只需要一张人脸识别就可以进行识别，表情识别中单张图片也可以用来识别，但当前研究中已经开如对继续的多帧图片的组合进行识别，即一组表示某表情动态变化的图片，这样的识别也称动态识别。

本文就基于 LBP(Local Binary Pattern)特征表示进行研究，采用了多种模式和不同的尺度来比较识别效果，并且在研究过程中提出了一种新的 LBP，这是本文的主要的新颖点之一。此外，在研究过程中，工程实现的细节与分析使得研究可以从更加细微的角度来分析 LBP 的理论基础与作用效能。文章也对现在有的一些研究方法进行了验证，如用于动态分析表示的空间 LBP (Volume Local Binary Pattern)，并给出了充分的实验数据说明这样的表示并没有总是如人们期待的那样好。除些这外，本文提出了一些新的概念，如 LBP 的零率 (Zero Ratio)，不规则的 LBP 表示 (Irregular Region Local Binary Pattern)，这些概念起源于对 LBP 研究的不断深入，也是源于直觉后大胆的假设，通过实验的方法最后也验证了所有假设，这些结论将在本文中一一阐述。

我们的研究主要集中在对 6 种基本表情的分析上，即：高兴、生气、伤心、害怕、厌恶和惊奇，之所以选取这几个表情作为研究对象是因为这些表情我们每天都可能遇到，从直觉上我们能更好地去判断我们所采取的研究方法和实验结果是否与

我们的生活经验符合。

在研究过程中，我们开发了一系列实验工具，包括数据可视化工具，数据处理工具，各种特征表示生成模块，样本训练分类系统，并开发了一个可用于测试相关算法在实际分析的效果。

关键词： 表情分析; LBP; Zero-Ratio; ILBP

Abstract

Over the last decades, facial expression analysis has been a field attracting more and more attention since it requires board of knowledge and technologies. Automatic facial expression analysis finds promisingly potential application in areas such as more engaging human-computer interface, criminal- track and entertainment etc. Conducting a very elegant and convincing research in facial expression also demands knowledge of physiologic, even that of anatomy. Practically, it requires theory from pattern recognition and software engineering. Anatomically, facial expressions are composed of some action units (AU) caused by muscle deformation in quite short time expansion, usually only several seconds or milliseconds. The deformations determinate how people understand the content of facial expression, such as contractions of eye lids, eye brows and nose. By studying all the possible relationship between muscle contraction and the way we interpretation the meanings of the facial expression, people may be able to understand the presentation of certain expression in given context.

Theoretically, the facial expression analysis and the facial recognition share many in common such as classification technology, feature presentation and so on. However, there should be some difference between them; for instance, the facial recognition requires only one image input when performing recognition, while there should be at least one facial image when performing expression analysis. The dynamic facial expression analysis demands a sequence of video which contains many images and evolves over time form the onset, the apex, to the offset. The similarities of facial identify recognition and facial expression analysis contribute great to researches conducted in this paper, while their differences also inspire the new methods which are the main key innovations of the paper.

In the paper, Local Binary Pattern (LBP) is used for feature presentation. The paper compares the abilities of several types of LBP (Such as Volume Local Binary Pattern^[7-8] Model in several variations with different dimensions, local blocking technologies as well as resolutions. Based on the results of conducted experiments, we proposes a novel LBP Model, we names it Irregular Region Local Binary Pattern (ILBP), besides, and complete

new technology based on Zero-Ration for feature selection is also proposed, experiment and relative definition will detailed in coming chapters.

Our research mainly focus on 6 conventional expressions, namely Happy, Anger, Sadness, Fear, Disgust and Surprise since we will be capable to compare the experimental results to our daily experiences, which would be valuable when time and experimental resource are limited. In the process of our research, we developed a series of tool for data formatting and pre-process, feature generators and classifiers; furthermore, we also developed a system for algorithms evaluation and real-time analysis.

Key words: Facial Expression Analysis; LBP; Zero-Ratio; ILBP

目 录

第一章 绪论	1
1.1 研究背景及选题意义	1
1.2 研究现状及存在的问题	2
1.3 主要的研究内容和特色	5
1.4, 本文的结构安排.....	6
第二章 基于积分图的眼睛定位算法	7
2. 1 积分图 (Integral Image)	7
2. 2 基于积分图的眼睛定位算法.....	8
2. 3 本章小结.....	11
第三章 基于零率的 LBP 特征选取算法	13
3. 1 LBP 特征	13
3. 2 零率与有效特征.....	15
3. 3 本章小结.....	21
第四章 LBP 特征的静态分析与动态分析研究	22
4. 1 相对样本强度.....	22
4. 2 前人的研究.....	24
4. 3 静态表情分析.....	24
4. 3. 1 分块 VS 无分块	25
4. 3. 2 不同参数的精度对比	30
4. 3. 3 本节小结	32
4. 4 动态表情分析研究	32
4. 4. 2 XYT-VLBP	33
4. 4. 3 旋转不变性	35
4. 4. 4 本节小结	43
4. 5 使用 AdaBoost 进行特征选择.....	43
4. 6 一种新的 LBP 分块模型	47
4. 7 不同样本强度下的表情分析研究	52
4. 7. 1 静态分析对样本强度的敏感性研究	52
4. 7. 2 动态分析对样本强度的敏感性研究	58

4.7.3 本节小结	59
4.8 本章小结.....	59
第五章 系统的工程实现.....	60
5.1 开发工具和演示系统	60
5.2 主要研究工具的开发设计	62
5.3 实时识别效果和问题	66
5.4 本章小结.....	69
第六章 总结与展望	70
6.1 工作总结.....	70
6.2 工作展望.....	70
参考文献	71
研究生期间发表的论文	76
致谢	77

Contents

Charpter 1 Introduction 1

1.1 Research BackGround and Objective 1

1.2Research Status and Difficulty 2

1.3Main Contents and Key Points of the Paper 5

1.4 Structure of the Paper 6

Charpter 2 An Eye Localization Based On Integral Image 8

2.1Integral Image 8

2.2 Eye Localization Using Integral Image 9

2.3 Charpter Summary 12

Charpter 3 Feature Selection Based On Zero-Ratio Approach 13

3.1 LBP Feature Representation 13

3.2 Zero-Ration and Discriminat Sub-Feature 15

3.3 Charpter Summary 21

Charpter 4 Static and Dynamic Analysis Using LBP 22

4.1 Relative Sample Intesity 22

4.2 Some Research 24

4.3 Static Facial Expression Analysis 24

4.3.1 Researches of Diffirent Block Sizes 25

4.3.2 Researches with Different Parements Configuration 30

4.3.3 Summary 31

4.4 Dynamic Facial Expression Analysis 32.

4.4.2 XYT-VLBP 33

4.4.3 Rotation Invariation 35

4.4.4 Summary 42

4.5 Feature Selection Using AdaBoost 42

4.6 A Novel Block-Based Approach	47
4.7 Researches of Different Sample Intensity	51
 4.7.1 Static Analysis Sensitivity to Sample Intensity	51
 4.7.2 Dynamic Analysis Sensitivity to Sample Intensity	52
 4.7.3 Summary	56
4.8 Chapter Summary	57
Chapter 5 Engineering Development	58
 5.1 System Development	59
 5.2 Research Tools Development	60
 5.3 Real-Time Recognition and Problems	64
Chapter 6 Conclusion and Future Work	68
 6.1 Conclusion	68
 6.2 Future Work	68
References	69
Publications	72
Acknowledgments	73

第一章 绪论

表情识别^[1-8]是最近几十年来非常热门的一个研究领域，其是人工智能研究的一个方向，也是我们人类非常渴望实现的一个愿望。LBP(Local Binary Pattern)^[9]作为一种常用的特征表示手段，在应用上也已经取得非常好的效果。本章将对表情识别的研究现状以及存在的问题等进行阐述，并且在最后对本文研究内容及研究过程进行一些总体的概述。

1.1 研究背景及选题意义

随着计算机理论与技术的不断发展，人类对于实现人工智能的梦想也变得更现实。几十年来，情感感知一直是我们对于计算机的良好期望，人们希望有朝一日能创造出可以像我们一样人感受到我们表情，甚至是内心情感状态及其变化的智能机器。并对其寄托着我们美好的心愿，那就是我们可以利用这样的机器来代替自身去做一些我们有时不愿意做或者没办法做到的事，如 24 小时照顾一些需要一直关注的病人，并及时满足他们的需要；此外，我们还希望创造出具有情感的机器人，它们就可以和我们的小孩子一起玩耍，帮助我们照顾孩子老人。设想这样的机器人将会给我们的生活带来多少便利。尽管科幻电影中不断向我们描述这样的技术革命可能会给我们人类带到灾难，但这些都无法让我们放弃对人工智能的追求。

此外，人脸表情是人类一种最有效，最自然和最直接情感交流方式。而自动的人脸表情分析（Automatic Facial Expression Analysis）是一个非常有趣但同样充满挑战的问题。表情分析具有非常广阔的应用前景，如人机交互（Human—Computer Interaction），数据驱动动画（Data-Driven Animation）等。

由于表情识别的现实意义和应用潜力，使之成为目前非常热的一个研究领域。又由于作为人工智能领域非常重要的一个分支，在表情分析方面的研究方法和理论成果也能够为其它智能领域的研究能提供很好参考。LBP 作为一种非常有效的特征表示方式，已经在非常多的研究中应用，如用 LBP 可以非常有效地表示出区域内的，如在人脸的特征定位 ASM (Active Shape Model) 和 AAM (Active Appearance Model)

研究中，就可以用 LBP 来描述脸部特征的统计信息；在人脸身份识别，LBP 也是一种很有用的表示特征。此外，由于在我们研究过程中，基于 LBP 的研究占了比重比较大，因此本文中也采用这种表示特征来进行表情分析的研究。

我们之前的研究主要集中的基于人脸的身份识别，这个研究也为我们开展表情识别的研究打下了很好的基础，尤其是在特征表示，特征提取和分类技术方面积累了良好的理论认识和工程经验。

1.2 研究现状及存在的问题

表情识别在现有的理论基础上，主要是采取对基本表情进行分类识别。从本质上来说就是分类的问题，也就是如何让计算机准确地区分一个表情所包含的成分，并通过动态地跟踪，判断出一个变化的表情所表示的情感内容。当然，目前的工作基本上是集中在表情的识别中，不论是静态的识别的研究，还是动态的识别的研究，其目的都是尽可能地提高识别的精度。由于表情变化还涉及到人物内心情感的变化，目前在这方面的研究报告还比较少，一个重要的原因是对于这种动态变体的解释需要涉及到文化背景，年龄或个人经历等非技术类问题。

由于表情分析的本质上是分类的问题，因此和很多分类问题一样，表情识别的研究也主要集中在以几个方面，即，特征表示、特征选择、分类等。当然，有些研究方法也可以不遵循这样的过程，如一些使用 PCA (Principal Component Analysis) ^[10] 的分析方法的。一个完全的识别过程大致可以用以下的过程图上来描述：

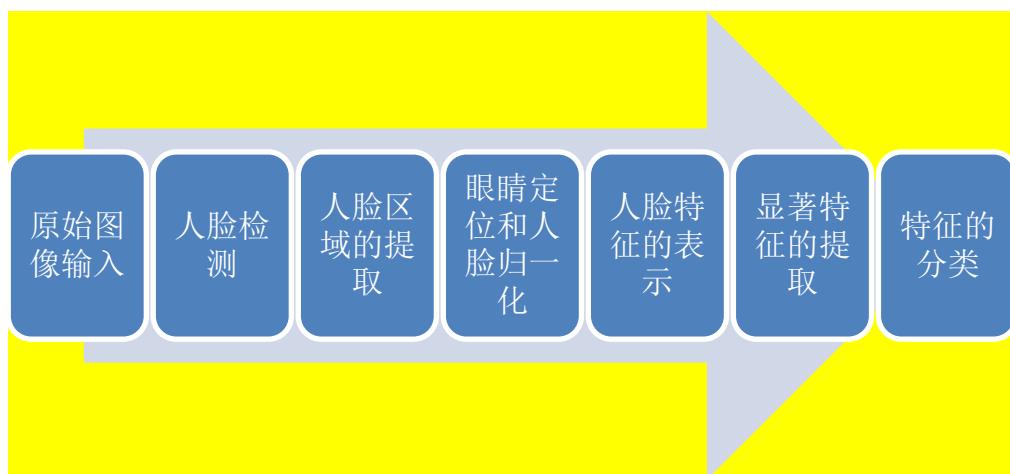


图 1.1：一个广泛的表达分析框架

在特征表示方面的研究也非常的多，有单一的表示，也有多种方法复合使用。如有人只用 Gabor 小波，通过选择显著的特征就可以进行很好的分析和识别；也有的只用 LBP 来分析表情；近些年来也出现了用 Gabor 小波和 LBP 结合^[11-12]的表示方式，同样取得了比较满意的效果。近十几年的早些时候，表情的研究主要是针对静态的表情研究，即一般分析过程只涉及到一帧表情图像。随着近几年的发展，已经有很多研究把表示方法扩展到动态的表情分析中^[13-17]，即对一组连续的动态变化的表情进行分析与识别。在表示方法上也出现了变体，如有人为了继续用 LBP 来表示这样的动态特征，对其做了扩展，于是出现了 Volume Local Binary Pattern(VLBP)^[18-19]，即将 LBP 的表示在时间轴上也进行了变换。此外，对于某种表示方法，也会有众多的变种，如 LBP 有人就采用了多分辨率（Multiple Resolution）^[19]的方法，即用不同半径的 LBP 作用在同一图像上以获得更多的信息。总之，目前的每种特征表示方法及其变种的目的都是为了获取更多显示的特征以达到提取分析精度的目的。

特征提取的方法也有多种，常用的有用 AdaBoost^[20]来选择出足够多的显著特征，这是一种非常有效而广泛应用的方法；也有用 PCA 来进行降维的；FLD (Fisher Linear Discriminant)^[21-23]同样能够用来选择特征和设置权重；本文中我们根据 LBP 表示的一些特点，采用了一种甚于零率（Zero-Ratio）大小的降维方法，这样的可以非常方便快速地降维。这样的方法可以极大地减少每个样本间的线性相关和非显著性特征的数量，我们会在后面的章节中详细介绍这种方法的有效性和处理方式。

训练分类器的方法也有许多。如有人用训练离散 HMM^[24-26]模型来分析各种表情；上面提到的 AdaBoost 其实也是一种有效的分类器，我们人脸识别的研究中也使用过这样的方法；还有使用神经网络来分类，但目前的情况来看神经网络的训练经常无法有效收敛；目前比较常用和有效的方法是使用 SVM((Support Vector Machine))^[27-29]来分类，我们在研究也采用了 SVM 作为分类器，因为在众多方法中，SVM 已经被很多实验证明是非常有效的方法，从理论上，SVM 的分类能力也是非常强大，而且有很多可以选的核和类型，可以通过实验找出最好的类型。

虽然表情分析在很大程度上已经取得令人兴奋的进展，并且也开发出一些实验原型演示系统。但由于表情分析的复杂性，还存在着一些没有很好解决的问题，如在特征表示方面，虽然人们已经从实验上证明 LBP 特征在描述脸部信息上有很好的效果，但实验上也出现了非常多不同的变种，不同种类之间所取得的效果也是不同

的，如分块方法，目前的分块一般都是采用简单的矩形子区域（如图 1.2），即把每一张人脸图片通过水平和垂直的分割使之成为一组大小差不多的子块，然后在每个子块上应用 LBP 来描述，最后在用某种方式将表示后的结果组合起来。这样的简单的分块与人脸部的非简单分布存在着明显的不协调关系，根据本文的研究结果和实验过程积累的认识，本文提出了一种新的分块模型（ILBP）。这个模型能非常有效地表示出人脸的复杂区域关系，而不是简单地将其分割成矩形。

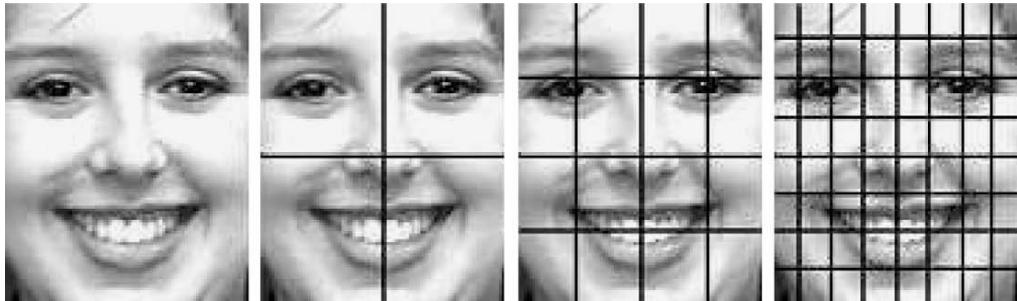


图 1.2：一组人脸分块方式

除此之外，目前引起非常多关注的动态表情分析也存在着技术上的难题，由于动态表情是人脸部肌肉在非常短暂的时间内连续变化引起的，而一个完全的表情过程应该是由一个起始点（Onset）渐渐过渡到表情强度最大点（Apex），然后回来与起始点一致的结束点（Offset），这整个过程够成一个表情序列（Sequence）。实际应用中普遍采用的并不是这样的一完整的序列，而是仅使用从 Onset 到 Apex 的部分。这样的分析技术在静态的样本分析上虽然是可以取得很好的效果，但在在实际的应用中，很难简单地判断在视频中什么时候是 Onset，什么时候达到 Apex，这使得在实验上的实验方法难以有效的适应现实系统。

最后，由于训练和分析用的样本一般都是采用样本中完整的序列，这些样本集都是从 Onset 到 Apex 的，即从表情强度为中性到最大值的序列，而现实的应用中，人们很少做出如此完整的表情过程，而且经常人们的表情只是非常小的副度变化，采用完全强度的样本来训练会在一定程度上降低实际应用的鲁棒性（Robustness），因此如何在训练过程中选择样本的强度也是一个值得深入研究的方向。



图 1.3: 从训练样本中取出的一些样本序列

1.3 主要的研究内容和特色

本文的研究内容并不是一开始就非常确定的，而是随着研究的深入不断地发现问题，通过反复的实验论证不断地提出新的问题，再对新问题提出相应的解决和论证方法。因此本文的研究是一个反复发现问题，分析问题和解决问题的过程，这是本文研究的一个特点之一。本文的主要研究内容如下：

首先，本文提出了一种基于积分图的全新眼睛定位算法，这是由于在对现在有的很多算法进行实验后我们发现这些算法的实时性让人很不满意，我们提出的算法在图像相对清晰的情况下能以非常高的精度来定位眼睛，而且在速度上有了极大的提高。由于眼睛定位能为人脸标准化的提供有效的参考点，标准化后的人脸可以大大降低由于人脸尺度的大小差异和旋转方向不同所带来的模式识别难度。

第二，在特征选取方式上，本文同样在对样本特征数据分析和统计后，提出了基于零率的特征选择方法，并在随后的实验中验证这样的算法是有效可行的。从理论上来说，这样的选择算法可以做为其它特征选取算法的前端，同样可以用来降低特征选择的难度。

第三，针对对目前动态分析方法和静态分析方法的差别，本文进行了大量的实验比对。对不同的表情上进行系统的分析，给出两种分析方法在不同表情上的相对

优劣结果，并进行相关讨论。

第四，本文在对动态和静态分析方法的比较后发现 LBP 算法对于区域的分割非常敏感。在此基础上提出了一种通用的，能够描述任何一种分割方式的区域分块模型，并能过实验给出更好的结果。

第五，本文就不同的样本强度对在实验上的影响和在实际中的应用也做了深入的研究，并且在实验上进行了验证。

最后，本文就实验原型系统的工程实现问题进行了一定的研究，对于工程上如何实现实时的识别系统做出了有意义探索，并最终开发出可以用于演示的实时系统。

1.4，本文的结构安排

本文共分成七章，系统地研究了基于 LBP 特征的表情分析方法中涉及到的各种细节和存在的问题，并简要地介绍了工程实现经验和方法。

第一章 绪论首先介绍本文的研究背景，选题意义和当前研究的现状，并对本课题的研究方法、研究内容和研究特色及文章结构进行了概述。

第二章 首先就现有的眼睛定位算法的优劣做了比对和分析，然后给出本文提出的新眼睛定位算法，最后在实验上论证所提出的算法的可行性，优劣以及适用条件。

第三章 介绍对 LBP 特征分析，提出 LBP 表示的特征的特点，并在理论上给出了相关分析，在分析的基本上提出基于零率的特征选择方法，最后在实验上验证算法的可行性。

第四章 重点研究目前的动态分析方法与静态分析。通过大量的实验数据给出两种分析方法的差别和优劣。研究在不同样本强度下实验效果与现实应用的效果的差别，并提出另一种训练思路。

第五章 介绍本文提出的新的 LBP 区域分块模型，阐述模型所包含的分析思想。并通过实验方法给出新模型的有效性。

第六章 探讨系统开发的工程问题和解决方法。

第七章 总结与展望，对本文的研究结果进行了归纳和总结，并在此基础上提出了未来的研究方向。

第二章 基于积分图的眼睛定位算法

我们知道，无论是在人脸识别还是表情分析，在进行特征提取，通过预处理的方法消去由于光照、人脸方位的差异将有助于后期的分析。而眼睛定位是人脸标准化过程中经常使用的一个途径，这是因为人的双眼是人脸非常显著的特征，且位置在人脸的相对位置固定，因此可以通过双眼的位置来计算人脸的标准位置，通过定位双眼，就可以消去因此人脸的左右转动和旋转等造成的人脸方位差异。从技术角度来说，眼睛定位好坏直接决定着后期分析的好坏；此外，眼睛定位的执行速度在很大程度上影响实时系统的性能，因此，定位的速度也是非常重要的，本章就一种全新的基于积分图的眼睛定位算法进行介绍，并通过实验与已有的成熟算法进行优劣比较。

2.1 积分图 (Integral Image)

积分图是由 Paul Viola 和 Michael Jones 在^[30]中引进的，它是一种新的图像表示方式。积分图的引进可以极大地提高在对象检测过程中所需要的区域灰度值和计算，所谓的区域灰度和计算就是在图像中的某一区域内进行对该区域的所有像素的灰度值进行相加。在积分图表示中，积分图点 x, y 即是以原灰度图片的左上部分的所有像素的灰度值的总和，即：

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (2.1)$$

其中 $ii(x, y)$ 表示积分图坐标为 x, y 上的点的数值， $i(x', y')$ 为原灰度图片在坐标为 x', y' 的灰度值（如图 2.1 所示）。

基于积分图的眼睛定位算法

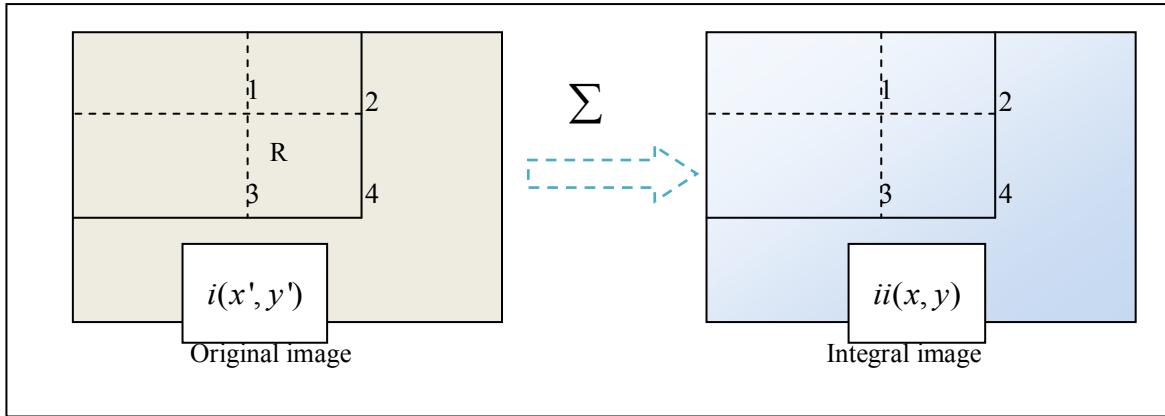


图 2.1：积分图的计算过程

原图像中的任一矩形区域内的像素的灰度值各可以由下式快速计算得到：

$$sum(R) = ii(4) + ii(1) - ii(2) - ii(3) \quad (2.2)$$

其中 $sum(R)$ 为图 2.1 中左边图中中标有 R 的矩形内的所有像素的有灰度值的总和, $ii(1)$

$\sim ii(4)$ 分别是与 R 四个顶点相对的, 具有相同坐标, 在积分图上的四个点的数值。从 (2.2) 式中可以看出不论任何大小的矩形, 利用积分图对其灰度值求和所用的计算量都是一样的, 而且非常快速。

2.2 基于积分图的眼睛定位算法

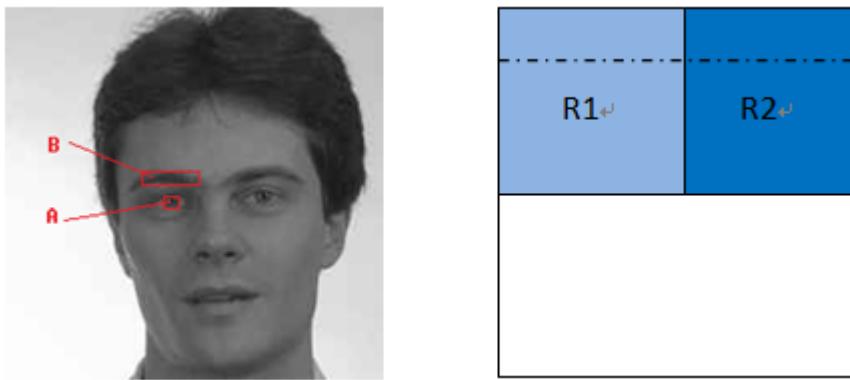


图 2.2：左边部标出了眼球和眉毛的主要区域;右边部分标出了左右眼的分布区域和眉毛的相对位置（虚线）

我们知道, 在一般情况下, 人的双眼眼球和眉毛区域相对于其周边区域的灰度和都要小, 即使在光照强度不同的情况下, 也依然是这样的 (如图 2.2)。本文提出

基于积分图的眼睛定位算法

的基于积分图的眼睛定位算法（Integral Image Approach Based Eye Detection Algorithm）^[31],简称 IIA，也就是基于这样的简单的原理来研究的。

在眼睛定位过程中，IIA 就是通过分别搜索 R1 和 R2（见图 2.2）两个区域定出左右眼的。然而我们在实验过程中发现，IIA 如此简单的处理并无法区分出眼球与眉毛的区域，因为两个区域都有很小的区域和。为了解决这个不足，本文采用了另一种思路：经过分析和观察，我们发现，眉毛的长度要远大于眼球的直径。为此，IIA 的实现思路上直接搜索眼球变成了先搜索眉毛再搜索眼睛，具体的过程如算法 1。

算法 1：

- 1)：根据生物特征，人的双眼是在人脸的上半区域内，因此首先计算待搜索人脸图片的上半部分的积分图 IM。
- 2)：利用得到的积分图 IM，用一个宽度比眼球直径大 4 倍左右，高为其一倍的矩形窗口（如图 2.2 左边部分用红色字母 B 指明的矩形）在 R1 区域内搜索出左眉毛的位置，设为 p3（如图 2.3 (a) 所示）。
- 3)：利用 IM 在 R2 区域搜索出右眉毛的位置，设为 p4，见图 2.3。
- 4)：利用 IM 在 p3 以下的 R1 其区域用差不多眼睛大小的方形窗口（如图 2.2 左边部分用红色字母 A 指明的矩形），搜索出积分和最小的区域，其中心点即为左眼球的位置，记为 p1。
- 5)：用和 4) 一样的方法在 p4 以下的 R2 区域搜索出右眼球的位置，记为 p2。
- 6)：进行相关校正工作。



图 2.3：左边的图片标示出理想的搜索结果；右边的标示出一种可能的实际结果

在图 2.3 中，当 IIA 返回的结果和图(b)一样时，我们需要在算法中进行校正。通

基于积分图的眼睛定位算法

过比较 p_3 与 p_1, p_4 与 p_2 的纵坐标。当左（右）眉毛的纵坐标与对对应的眼球的纵坐标之差小于一个眼睛的直径，我们认为它们是重合的，需要在当前眼球的位置之下再进行一次搜索，以得到正确的位置。

为了确定 IIA 的实验效果，本文分别从实验精度和速度上进行了测试。本文所采用的数据库是 FERET 数据库，本文选取这个数据库中的三个子集分别进行测试，其中 bg 集是向左旋转 22.5° 的，每个子集有 200 个样本。

表 2.1: IIA 在不同分辨率下定位眼睛所需要的时间

图像大小(像素 × 像素)	平均搜索时间 (毫秒)
256 × 256(100%)	1.574
192 × 192(75%)	0.868
128 × 128(50%)	0.450

表 2.2: IIA 在不同数据集上的实验精度。

数据集	FERET(ba)	FERET(bj)	FERET(bg)
样本数量	200	200	200
实验精度	94.9%	90.2%	93.4%

为了与现有的算法比较优劣，本文选用了由山世光等人提出的 Gabor-EYE 算法进行比较。Gabor-EYE^[32,33]是利用 Gabor 小波核在人脸区域在通过滤波强化眼睛区域，然后再定位出双眼的一种算法，由于需要用到小波变换，因此此算法的计算量会比较大，所需要的时间也比较长。由于在其文章中作者并没有给出算法的运算时间，本文仅就实验精度与其进行了对比。Gabor-EYE 的实验精度如下表所示（无 bj 的实验精度）：

表 2.3: Gabor-EYE 的实验精度

数据集	FERET(ba)	FERET(bg)
样本数量	200	200
实验精度	95%	91%

对比表 2.3 与 2.2，我们不难发现，Gabor-EYE 在 ba 集上的精度略高于 IIA，而在 bg 集上则不如 IIA。图 2.4 给出了一组在不同旋转角度与光照条件下的实验截图。

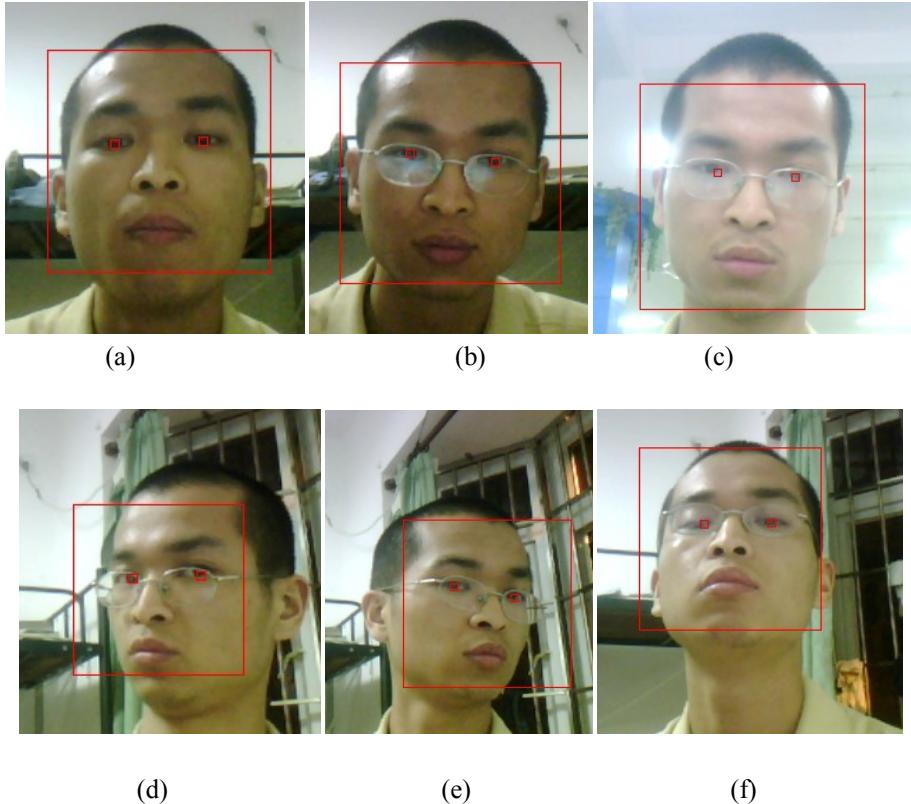


图 2.4：一组实际实验结果截图（不同的旋转角度和光照情况）

从理论和工程上我们知道，IIA 在搜索人的双眼时所需要进行的运算并不涉及任何浮点运算，仅仅用了加法，减法与比较操作，借助积分图，更是可以快速地计算出区域和。由于 CPU 在执行浮点运算时，尤其是进行浮点乘法或除法时，所需要的时钟周期是整数加减少的数十倍基于百倍。而在 Gabor-EYE 中，小波变换需要大量浮点乘法，因此其运算时间必然要远远大于 IIA 所需要的搜索时间。从表 2.1 的实验数据也可以看出，即使对人脸大小为 200*200 的图片进行眼睛定位时，IIA 也需要 1ms 左右的。也就是可以在一秒钟处理 1000 张这样的图片，而一般在实验中，人脸定位后的脸部区域也不大于这个大小，如果是表 2.1 中的第三行中的大小，就可以一秒内完成超过 2000 张图片的处理。

2.3 本章小结

本章介绍了本文在进行表情分析时，为了消除由于在样本图片在人脸定位后由于人脸的旋转和尺寸等不确定因素引起的差异对对于后期分析的影响，提出的基于积分图的全新的眼睛定位算法 IIA。通过实验在通用的数据库上分析其执行的精度和执行速度，与现有的复杂的眼睛定位算法进行比较后，IIA 的定位精度与 Gabor-EYE

基于积分图的眼睛定位算法

的差别并不大，然而其执行的速度要远远快于算法 Gabor-EYE。最后的实验结果与实时环境的测试也证明了 IIA 的可行性和有效性。

第三章 基于零率的 LBP 特征选取算法

在本章中，将介绍 LBP 特征表示的基本原理，然后通过可视化分析结果定量地分析 LBP 特征的特点，并在此基础上给出相应零率法特征选取的数学分析，最后通过实验数据的对比来验证基于零率降维的有效性。

3.1 LBP 特征

LBP 特征是指利用 LBP 算符操作输入数据的一种特征表示方式。它是一种用来描述局部数据区域的统计工具，其在一般纹理¹描述中具有很强的抗多变性，即局部内的数据微小的变化并不会对表示的结果有太大的影响。LBP 操作符最早是由 Ojala et al.^[9]在纹理识别研究时引进的。LBP 操作符的运算过程是这样的：在灰度图片上，对于每个 3×3 大小的区域，用中间的那个灰度值 V_c 来对其周围的 8 个值进行阀处理，过程如图 3.1 所示：

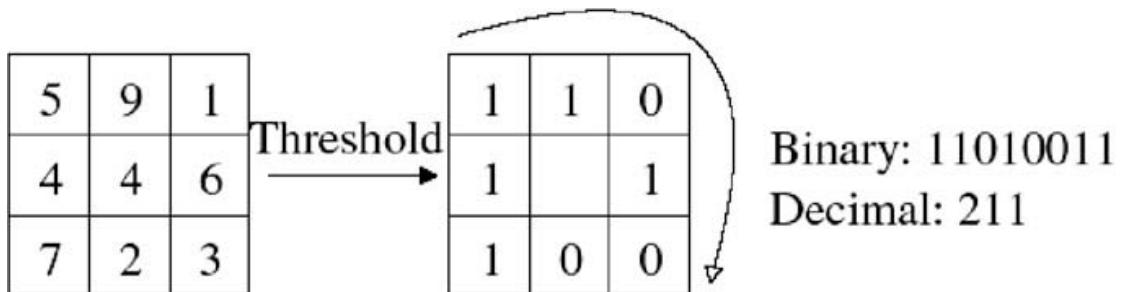


图 3.1：LBP 操作符的运算过程

在图 3.1 中，假设在前面的是一灰度中一个 3×3 的子区域，里面的灰度值如第一个格里的数字所示，那么依据 LBP 操作符，中间的值就是 $V_c=4$ 。接下去以 V_c 为参考，对其周边 8 个数进行二值化，二值化的操作方式由下面这个原则来决定：

$$V_i = \begin{cases} 0, & V_i < V_c \\ 1, & V_i \geq V_c \end{cases}, \text{ 其中 } V_i \text{ 即为 8 个周围点的灰度二值化后的值, } i=1, 2, \dots, 8.$$

¹这里所说的一般纹理不仅是指图片中的灰度阶所表示的色差变化，也可以通指一般数据的分布情况，如一个由浮点数组成的数据串也可以通过某种操作对应到常规的纹理中来。

通过这样的二值化阀处理后，第一个方格中的数据就可以变成第二个方格的数据。接下去按顺时间的方向把这 8 个 0 或 1 串起来，就得到方格中心点对应的 LBP 值的二进制形式（如图 3.1）即二进制表示值为 11010011，写出其对应的十进制数值表示，即 211。通过这样的操作，就可以得到一灰度图片中每一个点对应的 LBP 特征值。

由于这样的 LBP 操作符对应的周围点数有 8 个，因此，其对应的特征值就有 2 的 8 次方种可能，即 256 种不同的取值，每个可能取值我们称之为一个箱子（Bin）。一般地说，设环绕同心的点有 N 个点，这样的 LBP 特征值就有 2 的 N 次方个可能，是指数关系，因此实际应用中 N 不可能取得很大，一般取 16 个点就已经非常多了。在此强调一个，上面介绍的周边点是形成一个正方形的，实际应用中也经常使用以中心点以圆心的，半径为 R 的圆上均匀分布的 N 个点作为 LBP 操作符的输入，有的甚至可以使用椭圆（如图 3.2，和图 3.3 所示）。当然，原则上可以使用任何形状的上的点，但这样太过复杂的描述方式显得多余而没有必要。

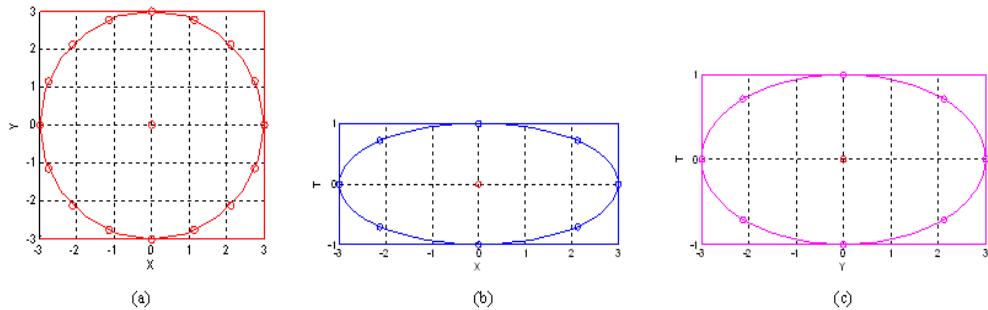


图 3.2：另一些常用的 LBP 周边点的采集模式

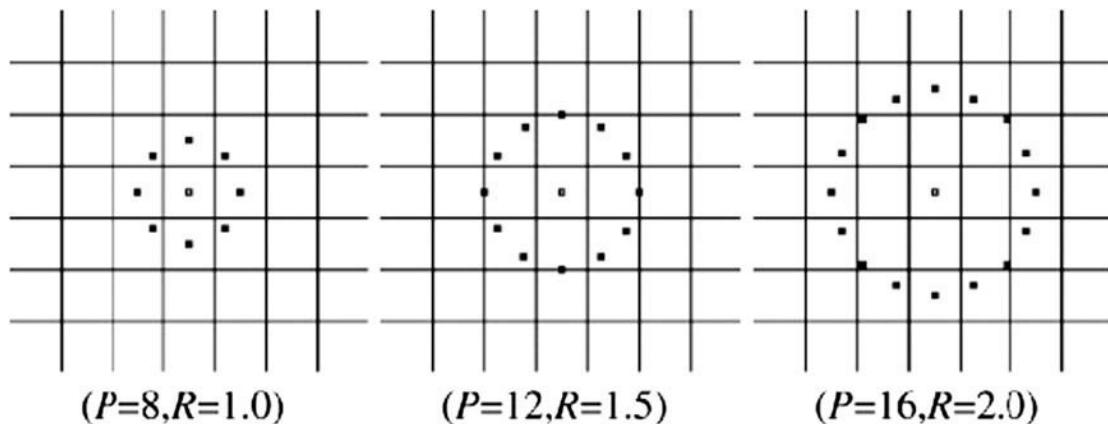


图 3.3：三种不同的取点方式，其中 P 是指所取点数，R 表示所用半径。

当选好一定的 LBP 描述模式后，就可以给出每个图片的 LBP 值统计直方图。即对一图片所有可能的 LBP 值的数量进行统计，这样的统计结果就形成了其最后的 LBP 特征表示。用以下的式子来描述：

$$H_i = \sum I(f_i(x, y) = i), i = 0, \dots, n - 1 \quad (3.1)$$

$$I(A) = \begin{cases} 1, & A = \text{True} \\ 0, & A = \text{False} \end{cases} \quad (3.2)$$

其中 n 为所有可能箱子个数，也就是 2 的 N 次方。

通过这样的统计后，就得到最后的 LBP 特征表示 H_i 表示 LBP 特征值为 i 的像素点的数量，所有的这些值按 i （箱子的编号，因此每个 i 会包括不同数量的像素数量，就如现实的箱子一样可以装东西，所以可以叫做箱子）的从小到大排列后，可视化的效果就是直方图（如图 3.4 所示）。我们称这样的特征为相应图片的 Histogram。

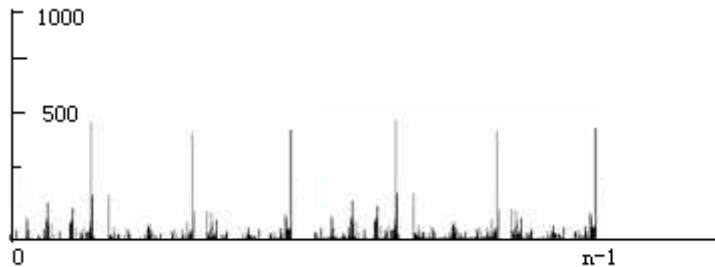


图 3.4：LBP 特征的直方图显示

3.2 零率与有效特征

除了一般情况下对整个图片进行 LBP 表示外，人们在实验中还经常把一整个图片分割成一些小块（如图 1.2 所示），然后单独对每个子块做 LBP 特征变换，再把变换后的 LBP 特征（Histogram）连接起来组成一个特征表示，这样就成倍地增加整个图片的 LBP 的 Histogram 表示的长度，如把图片分割成 4×4 的，那整个 Histogram 变成原来的 16 倍，因为每个子块都是单独进行特征变换的。

接下来我们将分析分块后的 Histogram 的特点，以标准人脸图片的大小为 96×96 为例，分成 3×3 ，采用圆形模式，在中心位置的周围均匀地取 12 点。这样一张

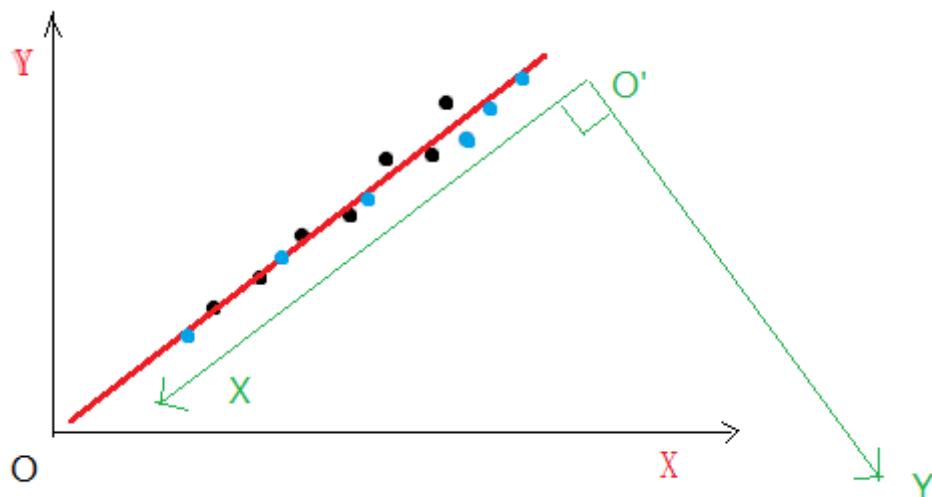
图片的 Histogram 长度为:

$$3 \times 3 \times 2^{12} = 36864$$

即有 36864 个箱子，那么每个箱子能分到的点数为 $96 \times 96 / 36864 = 0.25$ ，远远小于 1，因此，很可能出现非常多的空箱子。为了知道这些箱中的点的分布情况，我们进行了实验分析。我们记 H_{ij} 为第 i 个样本的第 j 个箱子所包括的点数，并对所有的样本进行统计，统计第 j 个箱子中的点数分布情况，得到 P_{kj} ， P_{kj} 是由下面的式子计算得到的：

$$P_{kj} = \sum_i I(H_{ij} = k) \quad (3.3)$$

上式中 $I(A)$ 为式子(3.2)。于是 P_{kj} 即表示在第 j 个箱子包含有 k 个点的样本数量，我们做这样的分析是为了确定一个 Histogram 的第 k 子特征是否存在显著的类间差别²，如果没有，第 k 个对于所有类别没有大的区别，也就无法用来区分两个或多个不同的类，即这个子特征不显著。从线性代数角度来理解，当所有的样本在第 j 个子特征的数值表示基本一样的，我们就可以用一条直线来连接这些数值，如下图所示：



图：3.5 坐标变换可以让某些存在线性相关的子特征在另一坐标系中有相似的数值，在 O' 坐标系中具有相近数值的点无法用来区分不同的类（蓝色的点和黑色的点），但在 O 坐标系中是可能构造出一微弱分类能图片的弱分类器。如 O' 的点没有通过坐标变换变成 O 中的点，我们就可以认为这些点无法区分。

² 类间差别在这里是指用来训练分类器的样本中不同类之间的差异性

基于零率的 LBP 特征选取算法

由于我们通过一次坐标转换把所有子特征的点的类间差别尽可能拉大，就会可能出现有些子特征无法在当前坐标空间中没有区分能力。因此，我们就可以将这子特征去除，而不会对整个特征的分辨能力差异造成影响，从而可能降低特征的维度。

下面我们将就给出 LBP 特征的某些子特征的 P_{kj} 分布情况，如图 3.6 所示。

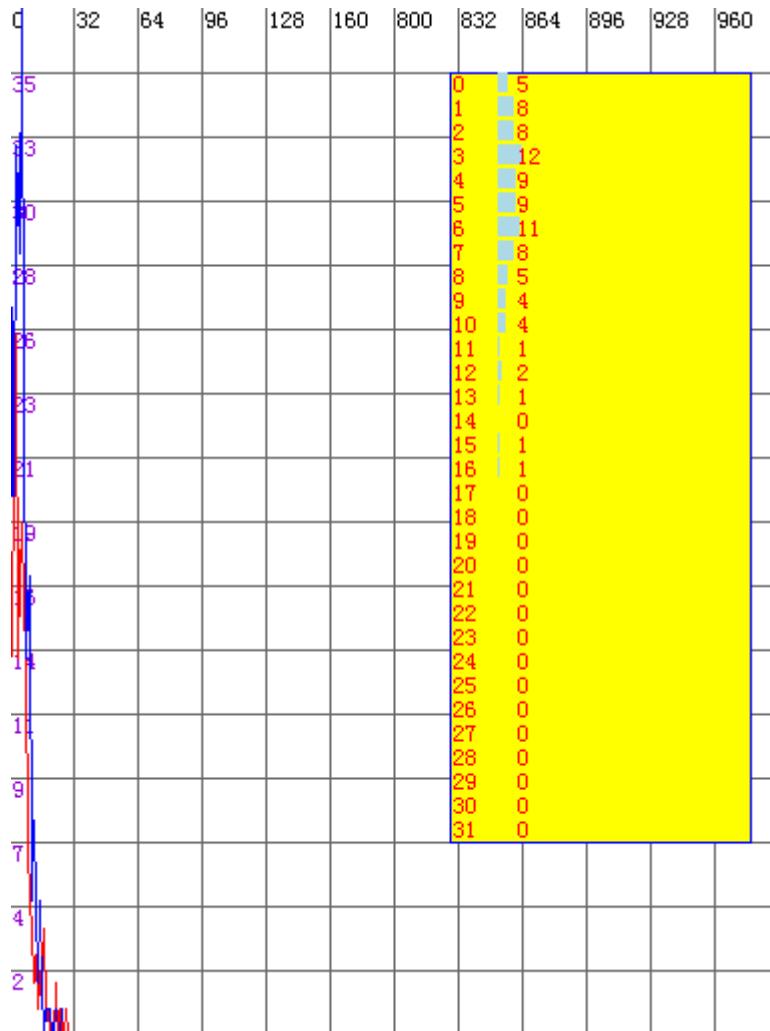


图 3.6：样本的 P_{kj} 分布情况，竖轴为 P_{kj} ，水平轴为 k 的数值，红色为正样本的分布，蓝色对应负样本的分布。可以看到这个子特征的 P_{kj} 基本集中在 0-16 之间。黄色区域显示是 k 为 0-31 的样本数占总样本的比例。

从图 3.6 我们可以知道，大部分样本的第 j 个特征的数值都集中在 1-16 之前，进一步的分析我们发现，在众多的子特征中，有很大一部分的数值为 0，有的甚至全部是 0，如图 3.7 所示。从图 3.7 中可以看到 0 的比例如此之大，基于有些子特征全是 0。以下我们给出零率的定义。

定义 3.1：在一个训练样本集中，所有样本的特征中的某个子特征中，值为 0 的

样本数占所有样本的比例称为零率 (Zero-Ratio)。

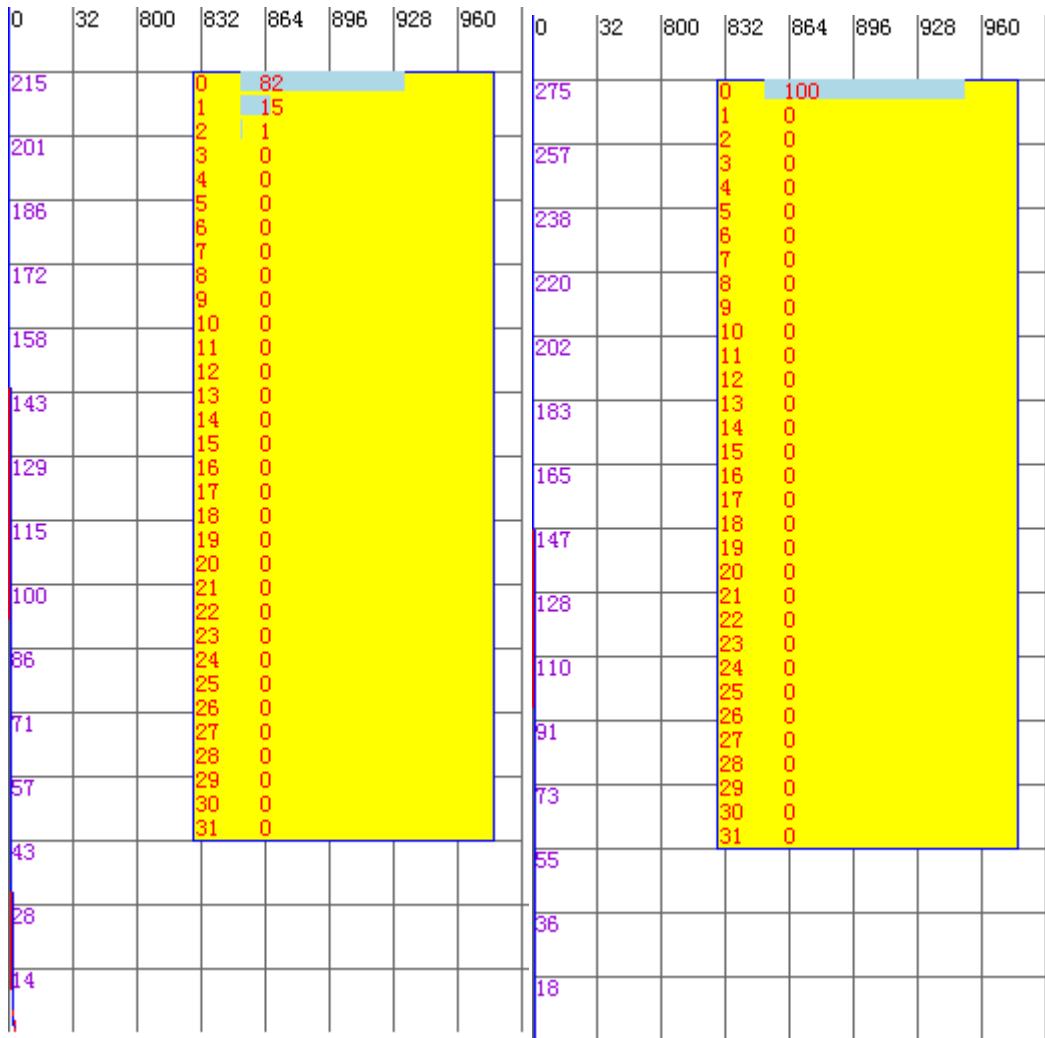


图 3.7：左边 0 占的比例高达 82%; 右边的则达到 100%，也就是所有样本特征中的相应子特征全为 0，显然这样的子特征不具有有类间差异性。

在零率的基础上，我们就可以根据零率大小来计算有效的特征数量 V_{zr} ， V_{zr} 由以下式子决定：

$$V_{zr} = \sum_j I(ZR(j) < zr) \quad (3.4)$$

$ZR(j)$ 表示第 j 个子特征的零率， zr 为 0 到 1 之间的数， $I()$ 为式 (3.2)。

图 3.8 为 V_{zr} 与 zr 的关系图。从图中可以看到，当 zr 小于 1 时， V_{zr} 都要远小于 LBP Histogram 特征表示的长度。从图中可知 LBP 特征的维度约为 69000，当 zr 取 0.5 时，有效的特征值的长度还不到 2000 维；而当 zr 取 0.8 时，有效特征也不足 4000 维，都远远小于 69000 维。 zr 的取值我们也叫有效特征截取值。

基于零率的 LBP 特征选取算法

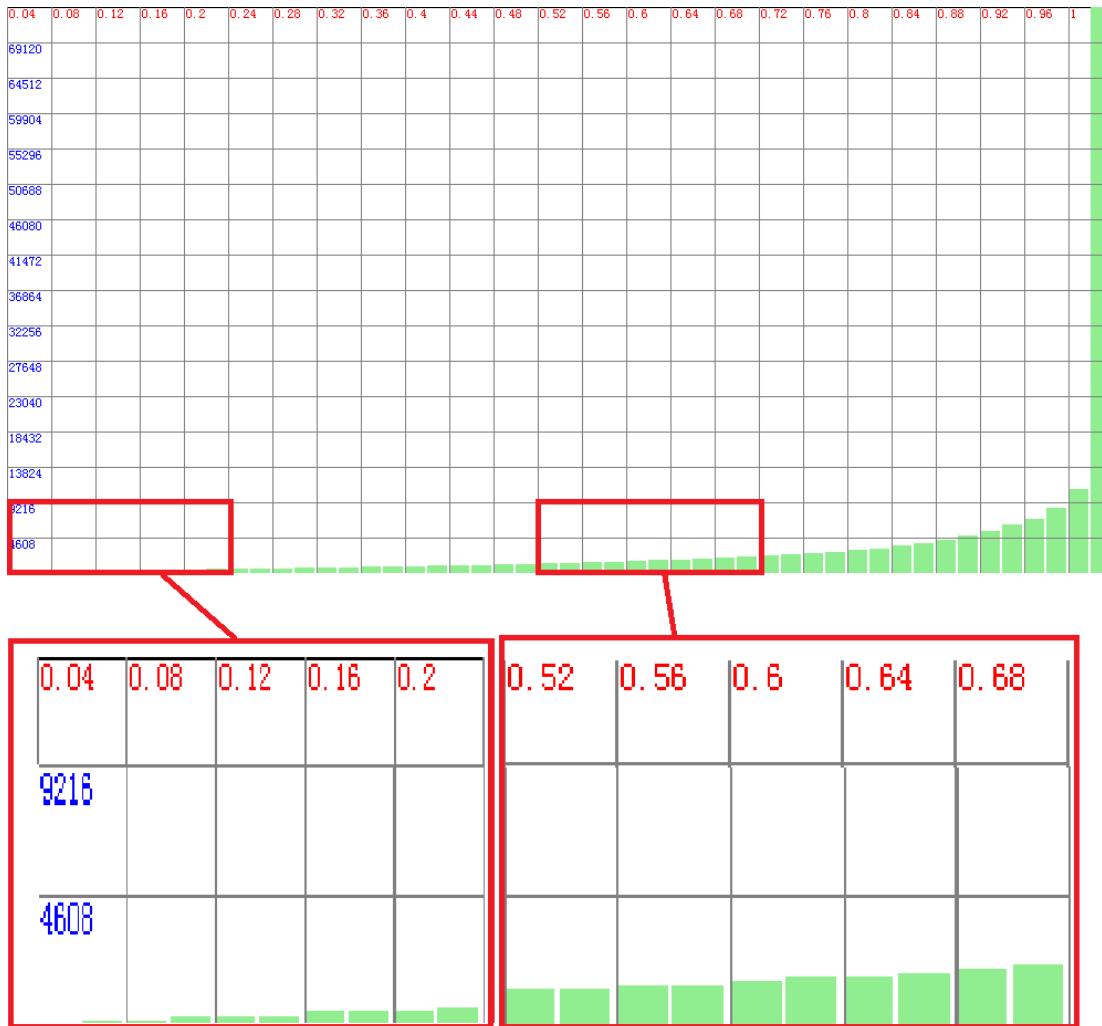


图 3.8: V_{zr} 与 zr 的关系图。上面水平的红色为 zr , 而竖轴的蓝色数值为 V_{zr} 。

只有当 $zr=1.0$ 时, V_{zr} 才会与整个 LBP 特征的长度一样, 也就是选取所有子特征。随后的实验证明, 当 zr 取 0.5, 0.8 和 0.90 三个值时, 实验精度并没有太大的变化, 而当 zr 取 0.2, 0.3 和 0.4 时, 实验精度就如 zr 取 0.5 或 0.5 以上的精度 (如表格 3.1 和图表 3.1 所示)。从表中我们可以看出, 随着 zr 的增大, 有特征越来越接近完整 LBP 表示特征, 训练精度也总体上也随着提高, 但并不是很明显, 对于某个表情基本没有太大的变化。以上实验是用 SVM 来进行分类的, LBP 类型为平面 LBP (Plane Local Binary Patter), 训练样本来自 CKD 数据库中的一个集合, 这个集合是由 UCSD 的 Machine Perception Laboratory 标注的, 包含 6 种基本表情 (Happy 86 个、 Surprise 56 个、 Fear 48 个、 Sadness 67 个、 Disgust 27 个和 33 个 Anger, 以后我们的实验都是在这个样本集合上进行的)。以上分类精度是通过分叉数 (Folder) 为 40

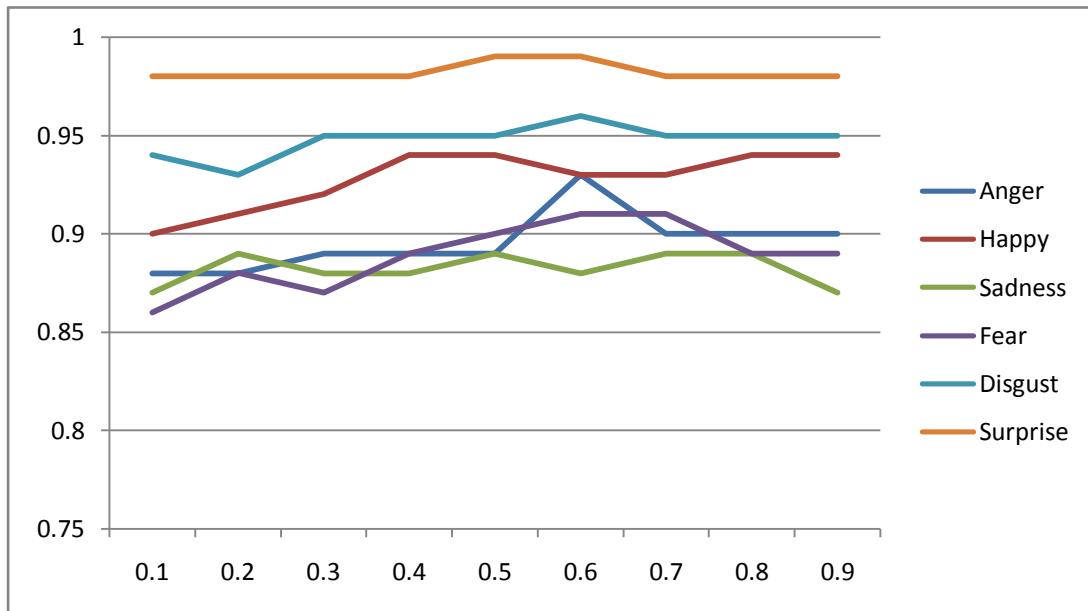
基于零率的 LBP 特征选取算法

的交叉验证（Cross-Validation）得到的，SVM 的内核类型为线性。

表 3.1：对于特定的 LBP 特征，有效特征截取值与训练后的精度的关系

零率	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Anger	0.88	0.88	0.89	0.89	0.89	0.93	0.90	0.90	0.90
Happy	0.90	0.91	0.92	0.94	0.94	0.93	0.93	0.94	0.94
Sadness	0.87	0.89	0.88	0.88	0.89	0.88	0.89	0.89	0.87
Fear	0.86	0.88	0.87	0.89	0.90	0.91	0.91	0.89	0.89
Disgust	0.94	0.93	0.95	0.95	0.95	0.96	0.95	0.95	0.95
Surprise	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.98	0.98

图表 3.1：表 1 的图形化



在此，我们简要地介绍关于 AdaBoost 的特征提取与上述方法的差别，AdaBoost 选择出来的特征都是有效的特征，有效特征针对性比较强，如是针对表情的或是针对身份识别的。而基于零率的选择方法并没有区分是对哪种分类，因此有一定的不足。但如果用 AdaBoost 直接在分块后的 LBP 表示特征上训练，很可能由于特征的过度稀疏而影响选择的速度，且会出现一些非显著特征。如经过上术方法处理后的特征再经 AdaBoost 选择，有可能取得更好的训练效果，同时也能极大地提高 AdaBoost 的训练速度。

3.3 本章小结

本章阐述 LBP 的基本概念和一些扩展方式，并分析了在分块情况下 LBP 表示特征可能存在过多的非显著特征的原因，通过给出了简单的数学理论解释，我们并提出了零率的概念以及基于这个概念的特征选择算法，最后通过严格的实验证明这种算法是有效可行的。

第四章 LBP 特征的静态分析与动态分析研究

本章将就静态与动态的表情分析进行研究，在实验上通过数据比较分析，并比较两种不同方式在优劣。为了更细致地比较两种方法的差别，本章还用不同的样本强调来进行实验对比，从而分析两在方法的分辨能力差别。

4.1 相对样本强度

比较早的研究基本上都是进行静态表情，所用的训练样本都只取一张面部图片。但在很多数据库里，提供的每一个样本都是一组连续变化的表情图片，这样一组图片称为一个 Sequence（如图 1.3）。以下我们从 CDK 数据库中取出两个 Sequence（如图 4.1，图 4.2）用来说明显实验时所用的相对样本强度的概念。

在 CDK 数据库中，每个 Sequence 都是由数量不等的若干张连续变化的人脸图片组成，其里面包含的表情强度是由中性（Neutral）慢慢地变化到就大强度（Apex），为了方便比较不同算法的优劣。本章给出相对样本强度的定义。

定义 2：当把一个 Sequence 的每一张图片按时间顺序进行从 0 开始的编号时，最后一张（Apex）的表情强度值定为 1.0，而把第一张（Neutral）的表情强度定为 0 时，从第一张到最后一张中的每一张的表情强度用相对值 RT 来表示，RT 由下式决定：

$$RT(k) = k/(N - 1) \quad (4.1)$$

其中 k 为第 $k+1$ 张的编号， N 为 Sequence 的图片数量，则 $RT(k)$ 给出了第 $k+1$ 张图片的相对样本强度。

一般情况下，实验都是在最大样本强度下进行的，即在静态情况下把每个 Sequence 的最后一张做为样本，而动态下把整个 Sequence 都做为样本（包含强度最大的那张）。在本章的后面部分，本文将就不同的样本强度的单独训练精度与混合精度进行实验上的分析，之所以将这个研究放在章节的后面部分是为了介绍本文已经实验数据后，综合这些研究结果，进一步提出问题，分析问题和解决问题。

LBP 特征的静态分析与动态分析研究

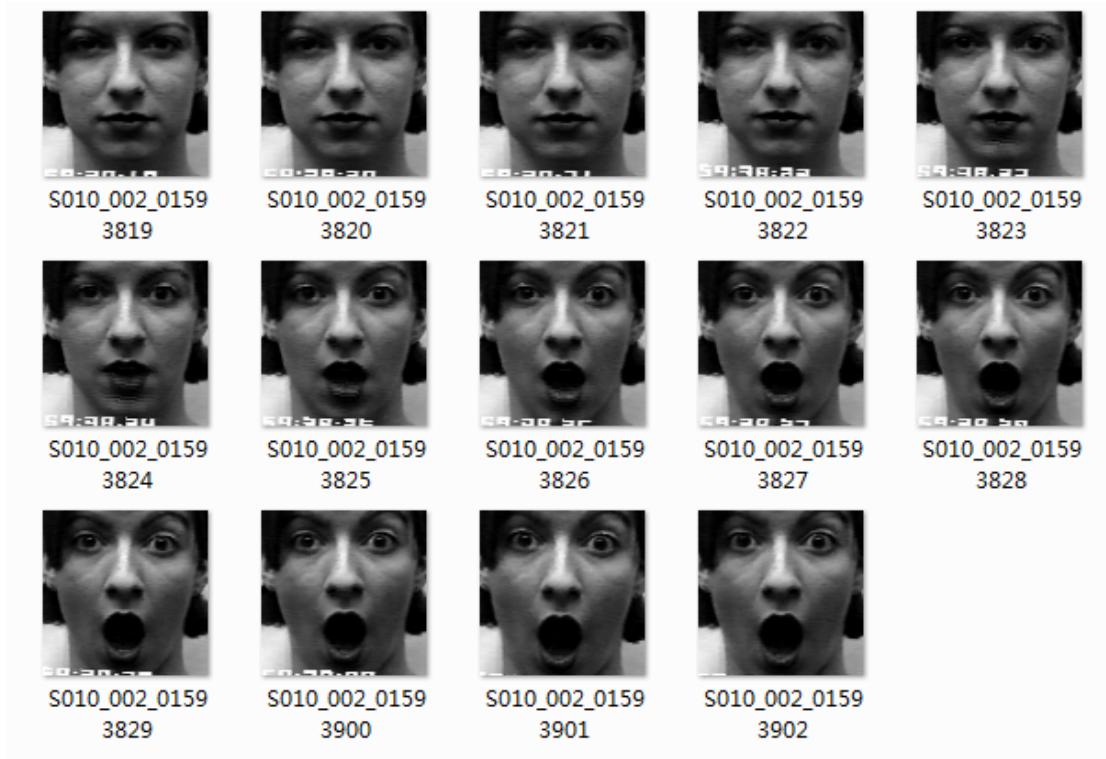


图 4.1: 从 CDK 中取出的一个样本序列 (Surprise)

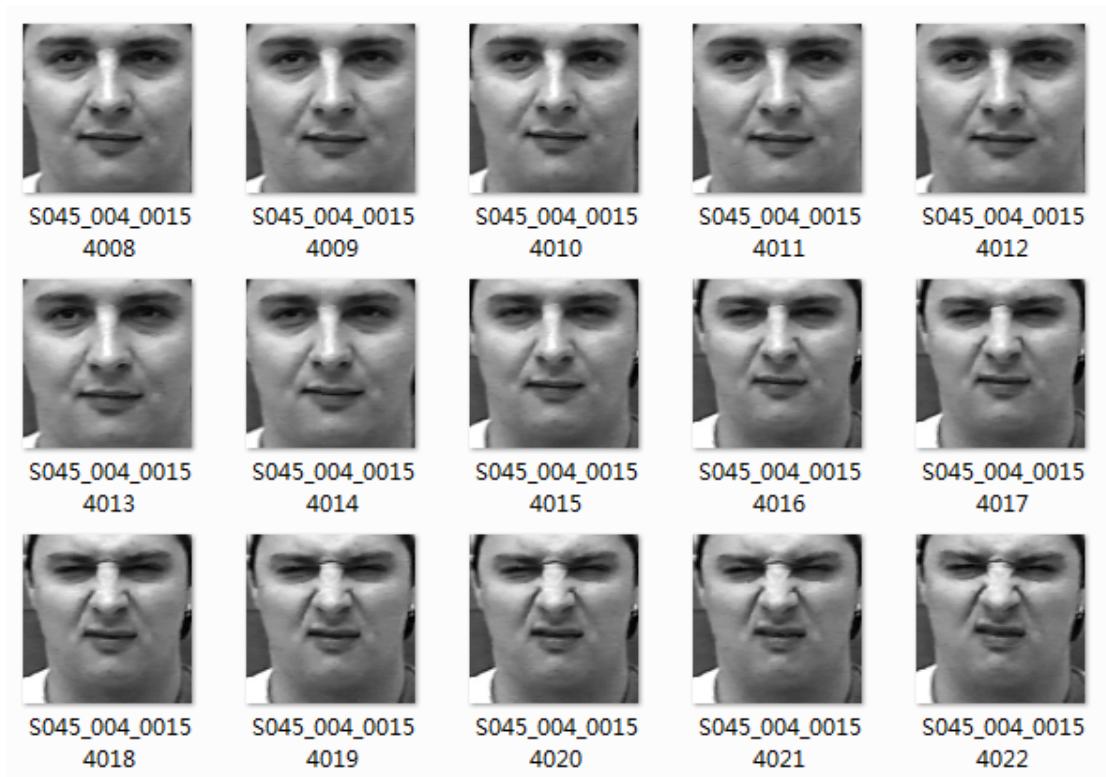


图 4.2: 从 CDK 中取出的一个样本序列 (Disgust)

4.2 前人的研究

本节将介绍我们在静态表情分析的工作和实验结果。静态表情分析只针对静态的图片进行分析，从我们的生活经验来看，很多常用的的表情我们是可以很容易的分辨出来的，而不需要有动态上的变化。但有些表情元素是蕴藏在动态的变化过程中的，而我们所研究的六种基本表情对于动态过程不应该是非常依赖于动态变化的，动态的过程起的作用在这些表情只应是有限的。2009 年，Douglas 和 Christian 在 *Journal of Vision* 发表了[33-35]，他们的研究成果也证实了我们的生活经验。图 4.3 的统计表格是他们在研究动态与静态表情在帮助常人分辨一些生活中常用的表情成分精度比较。

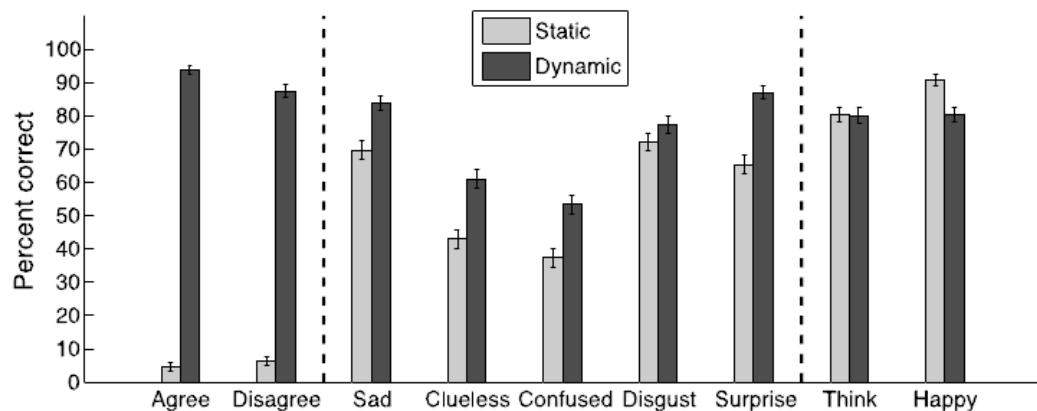


图 4.3，现实生活中人们对表情的分辨能力在动态和静态观察下的精度比对。

从图 15 可以看出，常见的 6 中表情中，Sad, Disgust, Surprise 和 Happy 在静态和动态的情况下各有差别。动态的分析有助于前三类表情的准确判断，但对于 Happy 来说并没有积极的作用，反而有碍判断。Douglas 等的实验是在实际中让非专业人在近日常的情况下进行的，因此是客观可靠的，而且它的实验结果很好地与我们的日常生活经验相符合。本章剩下的部分将就基于静态的 LBP 特征的表情分析做详细的展示，对 LBP 应用中不同的参数和分块情况进行全面的实验。

4.3 静态表情分析

在前面的章节中，本文介绍的 LBP 的实现形式，为了更方便地进行实验，我们采用了通用的圆形形式，在以中心点为圆心，半径为 R 的圆周上采集 P 个点，每个

子区域的大小的水平像素为 W , 垂直像素为 H , 子块大小一样, 无重叠。因此, 我们的 LBP 操作符就可以用这四个参数来完全决定, 并记成 $PLBP_{R-P-W-H}$ 。在没有分块的情况下, W 为图片的水平大小, H 为其高度, 即只分成一块。

4.3.1 分块 VS 无分块

表 4.1: 不同分块的训练精度

训练精度	$PLBP_{3-12-96-96}$	$PLBP_{3-12-48-48}$	$PLBP_{3-12-32-32}$	$PLBP_{3-12-32-16}$
Anger	0.8959	0.9022	1.000	1.000
Disgust	0.9148	0.9338	1.000	1.000
Fear	0.8486	0.8612	1.000	1.000
Happy	0.7287	0.9495	1.000	1.000
Sadness	0.7886	0.9085	1.000	1.000
Surprise	0.8359	0.9905	1.000	1.000

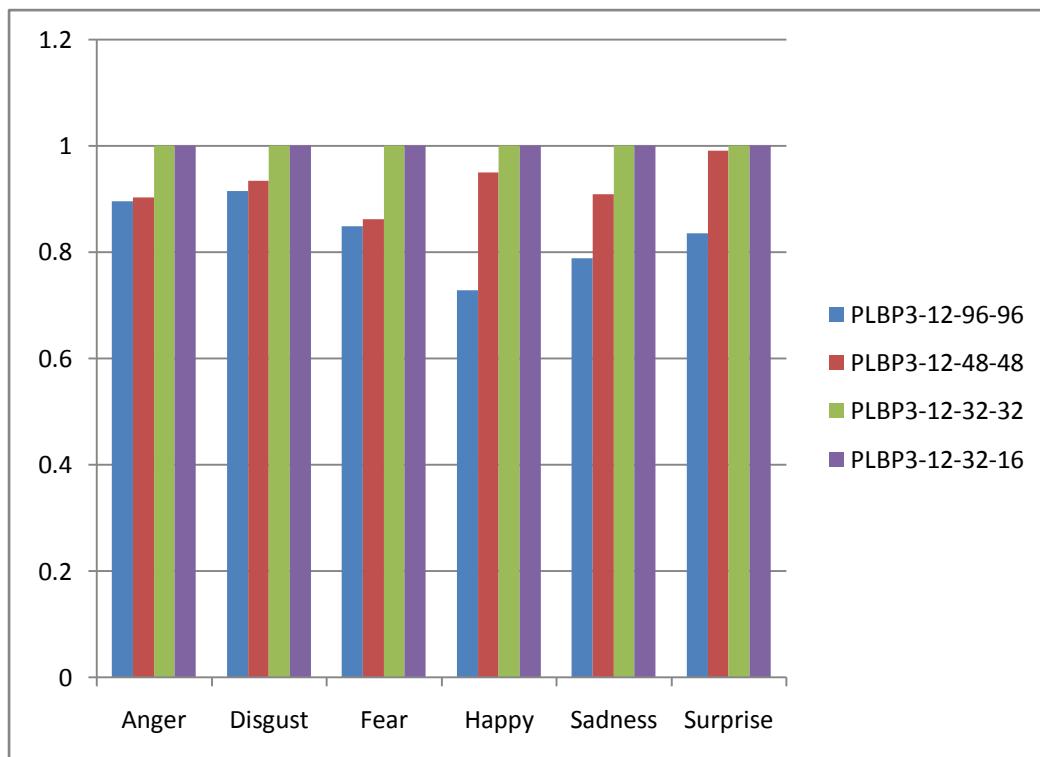
表 4.2: 40Folder 的交叉验证精度

交叉验证	$PLBP_{3-12-96-96}$	$PLBP_{3-12-48-48}$	$PLBP_{3-12-32-32}$	$PLBP_{3-12-32-16}$
Anger	0.8959	0.8959	0.9117	0.9085
Disgust	0.9148	0.9180	0.9589	0.9558
Fear	0.8486	0.8454	0.8580	0.9180
Happy	0.7287	0.8927	0.9180	0.9401
Sadness	0.7886	0.8517	0.8959	0.8959
Surprise	0.8233	0.9621	0.9874	0.9874

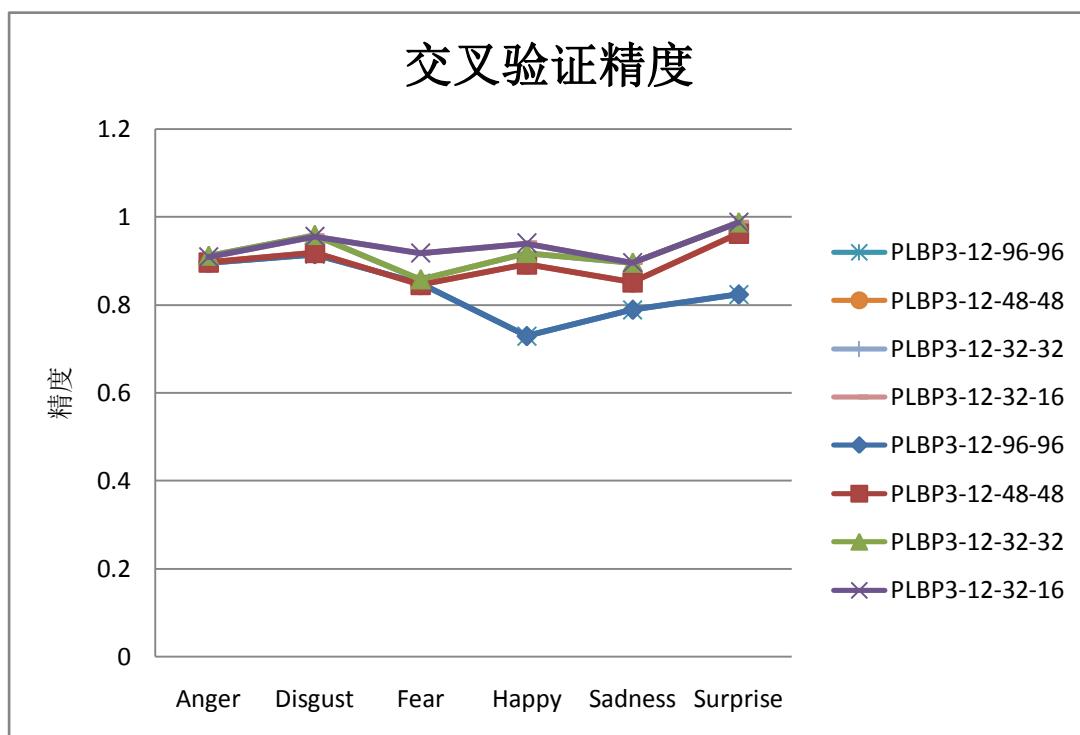
由于 LBP 特征是来描述区域里的统计特征, 而人脸表情是由脸部局部肌肉的伸缩组成的, 每个表情都与一组肌肉变化组合对应, 从这个角度来理解, 分块的 LBP 能更好地描述局部表情; 而无分块的 LBP 特征描述是脸部全局的特征, 会在一定程度模糊局部表情信息。我们通过实验来验证这一观点, 表 4.1 和图表 4.1 给出了不同子块大小的所能得到的训练精度比较; 表 4.2 和图表 4.2 给出了不同子块大小在 40 分叉的交叉验证结果; 此外, 为了检查样本有没有被正确分类, 我们开发的系统给出了样本在用 SVM 分类后的详细情况, 图 4.4-图 4.7 为四种不同 LBP 算符对应的分

LBP 特征的静态分析与动态分析研究

类分布情况。采用的 LBP 算符为 $PLBP_{3-12-W-H}$, 其中 W 和 H 为不同大小, 为了完整比较, 我们没有进行基于零率降维处理。



图表 4.1, 不同的分割方式的训练精度比对图



图表 4.2, 不同的分割方式的交叉验证精度图

LBP 特征的静态分析与动态分析研究

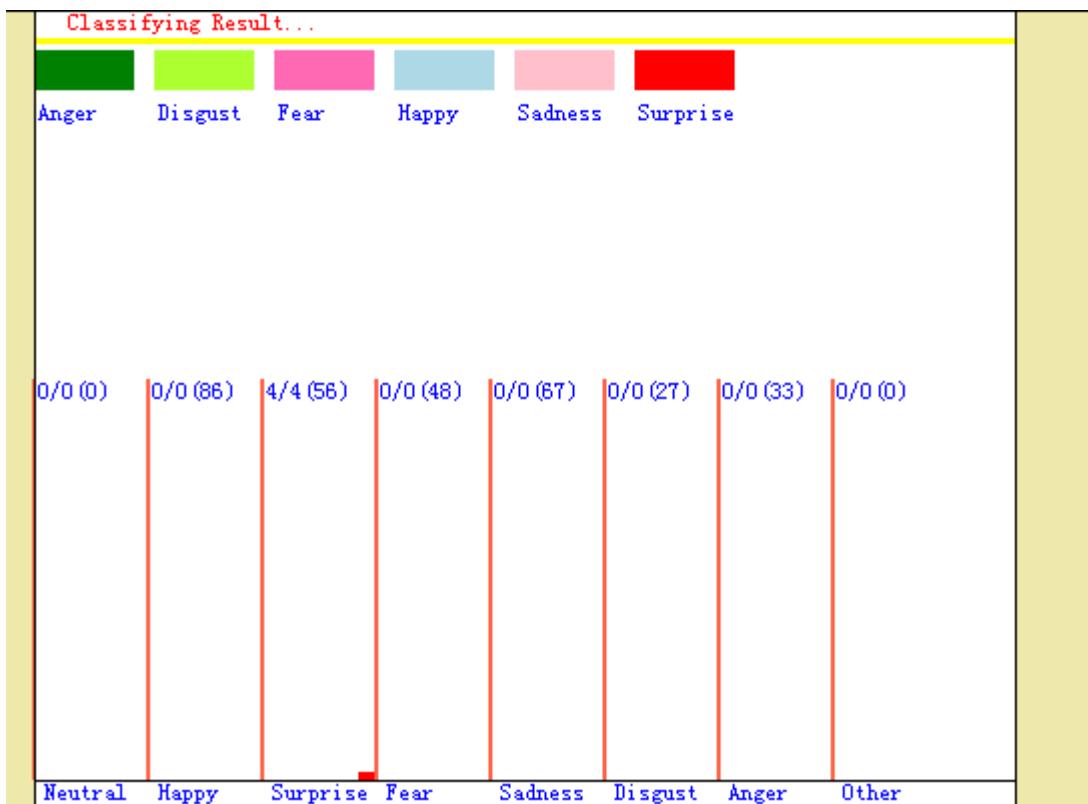


图 4.4, PLBP3-12-96-96 的样本分类后的分布结果（无分块）

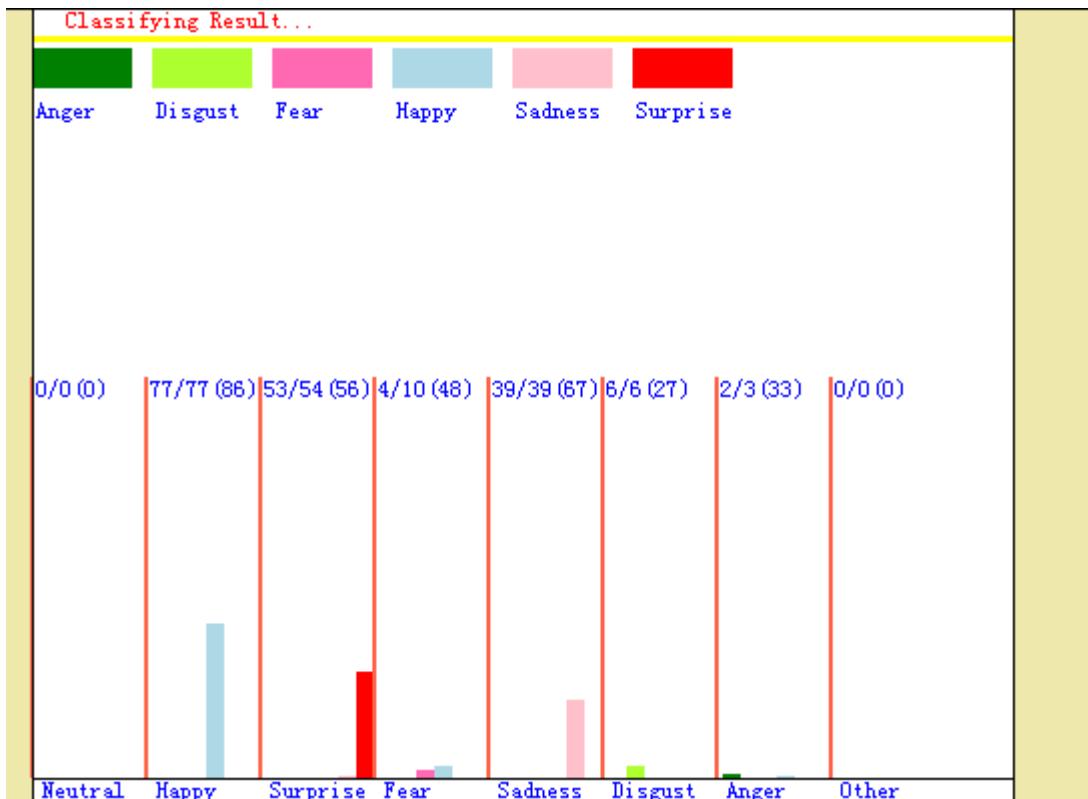


图 4.5, PLBP3-12-48-48 的样本分类后的分布结果(分成 2*2 块子块)

LBP 特征的静态分析与动态分析研究

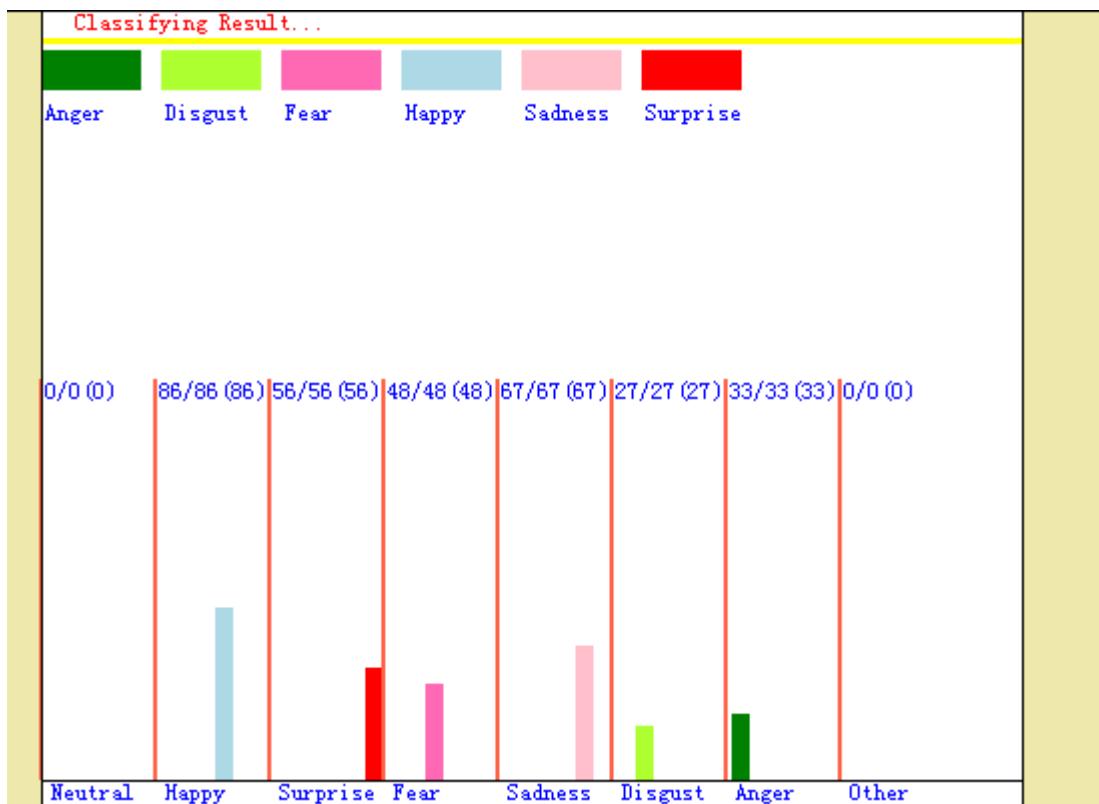


图 4.6: PLBP3-12-32-32 的样本分类后的分布结果(分成 3*3 块子块)

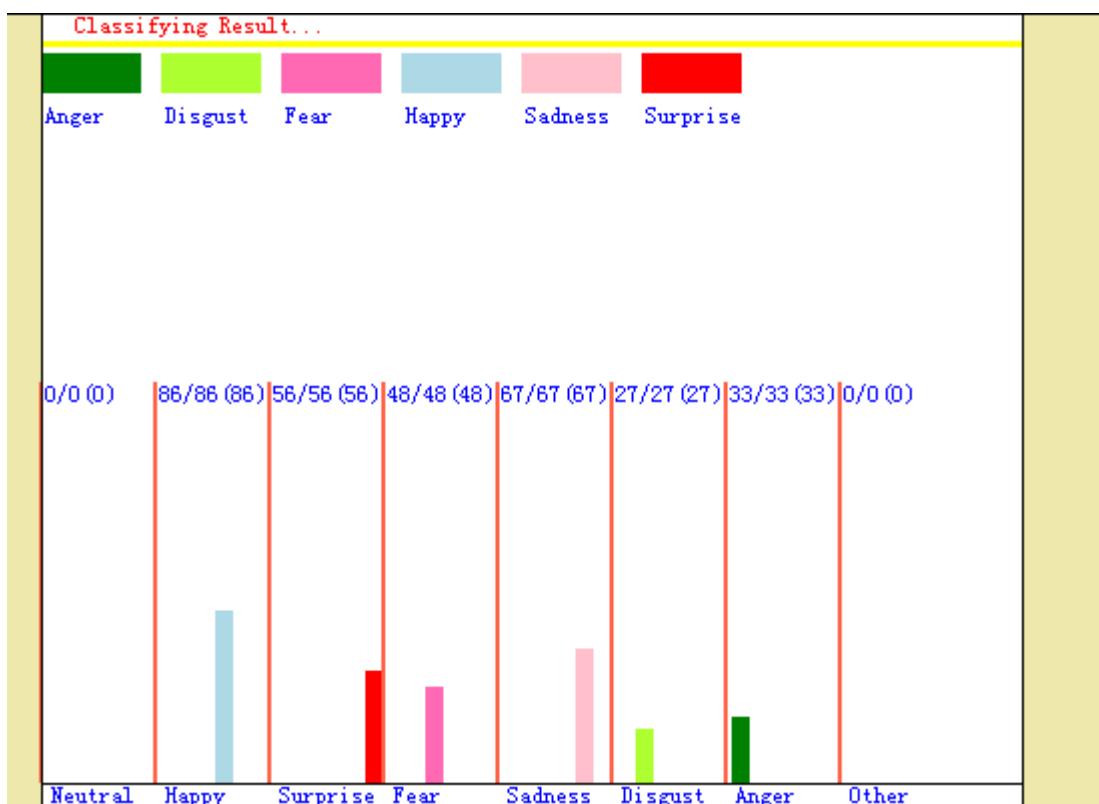


图 4.7, PLBP3-12-32-16 的样本分类后的分布结果(分成 3*6 块子块)

图 4.4-图 4.7 显示的是经过训练后，训练样本被正确分到所属的类别的统计直方图。由于系统定义了 6 种基本表情和中性表情，其它没有被定义的表情分成其它，由于采用了 MPT 实验室的标注的 CDK 子集，里面只有 6 种基本表情。中性和其它的数量为 0。每个类别图上有三个数据[数字 1/数字 2 (数字 3)]，数字 1 表示被正确归为该类的样本数，为数字 2 表示所有被归为本类的样本数(包括错分的样本数量)，数字 3 表示在训练样本中，此类样本的实际数量。理想的分类结果应该是这三个数字相等，即没有错分，漏分的样本。可以看到，图 4.4 显示的是没有分块的 LBP 特征分类的结果，虽然表 2 和表 3 的训练和交叉验证的精度都达到 70%以上，但实际的结果却让我们大为惊讶，几乎没有几个样本被正确放到所属的类别中，只有 Surprise 这个类有 4 个被正确归到该类。原因是这个精度是指所有样本的训练后的分类精度，而每个表情的样本数占总样本数并没有点绝对优势，虽说 70%以上的正确率，其实 70%并不能正确反应这个数字与我们的心理感觉是不一样的，因为我们采用的是 One-Against-Other 的训练策略。举例说，300 多个样本，6 种基本表情各占 50 个，在训练 Anger 表情的 SVM 模型时，50 为正样本，其它 250 个就是负校本，由于这样特征无法被有效地分开，结果是训练后，全被归到负样本了。那么那 50 个正样本就被错分了，而 250 其它样本被正确分类，因此分类精度就是 $250/300=75\%$ ，这解释了我们心理上对 70%这相对高的正确率与实际结果的落差。图 4.7 给出的是当人脸图片被分成 4 块大小相同的子块后的结果，可以看出，效果远远好于没有分块的。随着分块的增多，到了 9 块时，已经达到没有错分的漏分了。分成 16 块时，也有同样的效果。因此，适当地选择分块数量能有效地提高分类的效果。从表 4.1 和表 4.2 也可以看到，不论是训练精度和交叉验证的精度，只有都大于 85%以上才能取得良好的分类效果，在没有分块的情况下，这两项指标都太低了，以至无法有效分类。最后，交叉验证的精度能更加客观，而训练精度经常能达到 100%，这是由于样本数量还不是特别多的原因。所以，在比较时，一般只拿交叉验证精度来做比较。在实际应用中，为以提高准确度，往往可以通过适当地增加分块来获取，但这种方式不能是无限制，而且也需要代价的。一方面不可能分得太细，这样会使得特征变得非常稀疏，解决的方法是进行特征选择，可以 AdaBoost 或前一章的基于零率选择策略，甚至可以两者结合来选择少量显著特征。另一方面，分得太细给训练带

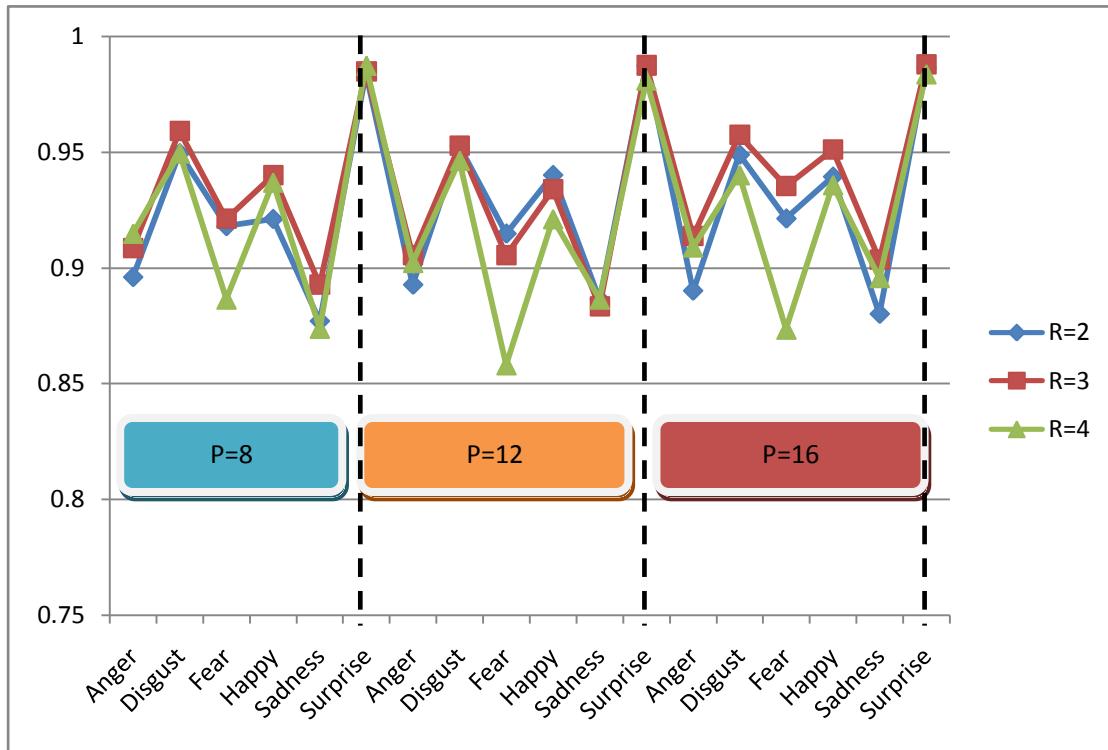
来了很大困难，会极大地增加训练时间，且在实际应用中会影响系统的实时性能。从表 4.1 可以看到，当分割的数量到达 3×6 块时，其交叉验证的精度要明显好于其它三种，为此，我们在此后的研究也采用这样的分割方式。

4.3.2 不同参数的精度对比

在上一节中，我们分析了不同的分块方式对最后的分类精度的影响，可以看出，没有分割的 LBP 特征情况很难实现有效分类，为了有效地分析表情成份，我们采用了 3×6 的分割方式。在本节中，我们将就不同的 R 值 和 P 值对识别的精度进行研究。R 越大，反应的局部特征范围变大，R 越小则会使表征的局部特征范围变小，一般情况下，R 不能太大，经常的取值为 2, 3 或者 4 个像素就可以了。P 的值决定着采点的密度，P 太小无法反应局部特征，越大能获得的有效特征就越多，但同时也会影响 LBP 特征表示过长而造成训练困难和影响实时性能。因此，P 的值一般取 8~16 之间，我们将就这样些不同参数进行实验对比，通过实验结果分析不同的值对实验结果的影响情况。表 4.3 和表格 4.3 给出了我们的实验结果。

表 4.3：不同参数对实验精度的影响

ZR=0.5	PLBP ₂ - 8...	PLBP ₃ - 8...	PLBP ₄ - 8...	PLBP ₂ - 12...	PLBP ₃ - 12...	PLBP ₄ - 12...	PLBP ₂ - 16...	PLBP ₃ - 16...	PLBP ₄ - 16...
Anger	0.896	0.908	0.915	0.893	0.905	0.902	0.890	0.914	0.909
Disgust	0.949	0.959	0.949	0.953	0.953	0.946	0.949	0.957	0.940
Fear	0.918	0.921	0.886	0.915	0.905	0.858	0.921	0.935	0.873
Happy	0.921	0.940	0.937	0.940	0.934	0.921	0.939	0.951	0.936
Sadness	0.877	0.893	0.874	0.886	0.883	0.886	0.880	0.904	0.896
Surprise	0.984	0.985	0.987	0.987	0.987	0.981	0.987	0.988	0.984



图表 4.3：不同的 R 取值对比

从以上表格的数据可以看到，当 $P=8$ 时， R 取 3 效果最好；而当 $P=12$ 时， $R=3, 4$ 都优于 $R=2$ ，但 $R=3$ 略优于 $R=4$ ，因此，我们在实验时在 P 不大于 16 时， R 匀取 3 这个值。不过这里要指出的， R 不一定要取整数，也要以 2.5, 3.5 之类的分数，本文之所以没有这样取是因此在图片中像素的个数是整数的，并没有必要最一个分数的半径，为了使计算的目标点，结果还是要通过近似归整到整数（采用插值的方法可以采用分数的半径来实验，本文没有就这些方法做验证，有待以后的工作中进行研究）。由于所有的特征都进行的零率的降维处理，实验结果指出，不管取 8, 12 还是 16，当 ZR 最 0.5 时，最终的有效维度都在 1500 左右，这也保证了比较的公平性，但也带来了疑问，公平性必然会使那些有更多信息的 P 值的效果受到影响。然而，这里要指出的是，在前一章中，我们已经证明，当 zr 取 0.5 这个值的零率降维只会微小地影响结果，这个微小的变化不足以影响表 4.3 中的实验数据，因此表 4.3 中的实验数据是可以真实反映不同参数的效果差异。从图表 4.3 也可以清楚地看到， $R=3$ 总体上要优于 $R=2$ 和 $R=4$ 。

4.3.3 本节小结

本节主要介绍平面 LBP 模型在表情分析方面的情况，并从不同的分块形式，不同的参数对比了实验效果。发现有分块的 PLBP 效果要远远好于无分块的 LBP，并在实验结果的基础上结合理论分析了这样的差异原因。不同参数的实验效果对比也为我们后面的研究提供很好的启发。

4.4 动态表情分析研究

本节将介绍我们在表情分析的另一方面的工作，即关于动态表情的分析研究，由于动态表情的分析需要涉及到连续变化的 Sequence 来做为样本，因此如何对这样的 Sequence 进行特征变换也是本节考虑的主要问题，对于特征选择和分类由于在后继的章节中我们将通过不同的方法进行实验比较。在本小节中我们依旧采用我们在前一章提出来的基于零率的降维算法，然后比较其与静态分析的结果的差别，并能实验进行讨论，然后为后续的工作做好铺垫。

4.3.1 Volume Local Binary Patter (VLBP)

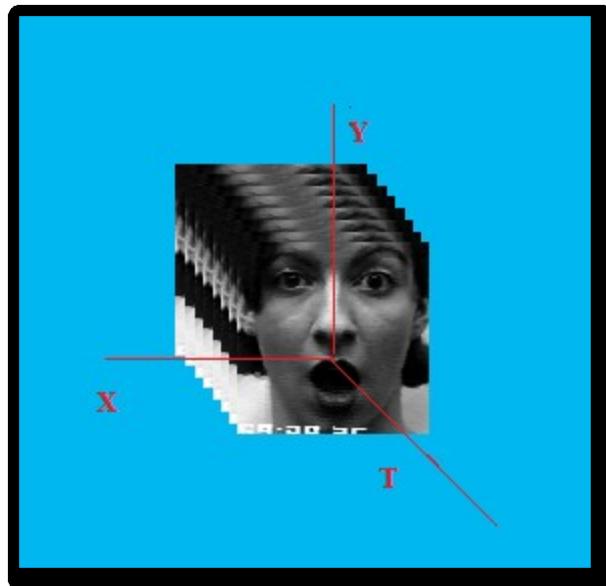


图 4.8：一个 Sequence 组成了一个具有立体效果的体积空间，在表情强度变化的方向为时间 T 轴的方面，因此，X，Y，T 组成一个 3 维的坐标系。

VLBP 即是把 LBP 原来的二维操作扩展为三维的特征操作。VLBP 是由 Zhao 提出

的用来分析时变纹理特征的，在 Zhao 的文章中，她指出 VLBP 的分析能力要明显好于我们先前使用的平面 LBP 操作符，此外，VLBP 还能有效应对纹理的旋转变化，光照变化等。然而，由于对 VLBP 的实现形式有非常多的种，在实现细节上各有各的特色。本小节仅介绍本文在实验时使用的方式，对于其它的实现方法会在本节最后做简单的介绍。为了方便后面的概念解释和其它说明，图 4.8 给出了一个 Sequence 的体空间分布图示，

4.4.1 XYT-VLBP

在这个小节中，我们将详细介绍我们实验中使用的一种用来描述动态纹理的 LBP 算符。由于要将原来的平面 LBP 扩展到时间变化这一维度，因此，我们采用了三个交叉的采样平面（如图 4.8），这三个平面分别称为 XY 平面，我们所选的 XY 平面是在一个 Sequence 样本中，表情强度最大那张图片的平面空间，这样是为了把尽可能大的有效空间特征考虑进去，这一点在处理上与 Zhao 原来的作法有点不同。Zhao 她采用的 XY 平面是一个 Sequence 中，表情强度为中间值的那个图片的平面空间作为 XY 平面。第二个平面称为 XT 平面，第三个平面是 YT 平面。图 4.8 中的 T 轴是通过整个 Sequence 中所有图片的中心的一条由最小表情强度的图片指向最大表情图片的方向的直线，是动态纹理变化的时间轴。如果把 XT 平面与整个 Sequence 的所有相片所在的平面的所有交线上的像素一一取出，并按在 T 轴方向的原顺序排成一个平面（如图 4.9 (c) 所示），那么这个平面上的数据也可以进行二维的 LBP 算符操作。同理，我们在 YT 方向上也可以进行 2 维的 LBP 变换，如果我们把这三个平面上的 LBP 特征联结起来，就构成了动态的表情特征表示，也即 VLBP 算符作用在这一 Sequence 后的特征变换。因此，从这个意义上说，VLBP 只是一般 LBP 在三维空间上的扩展应用。

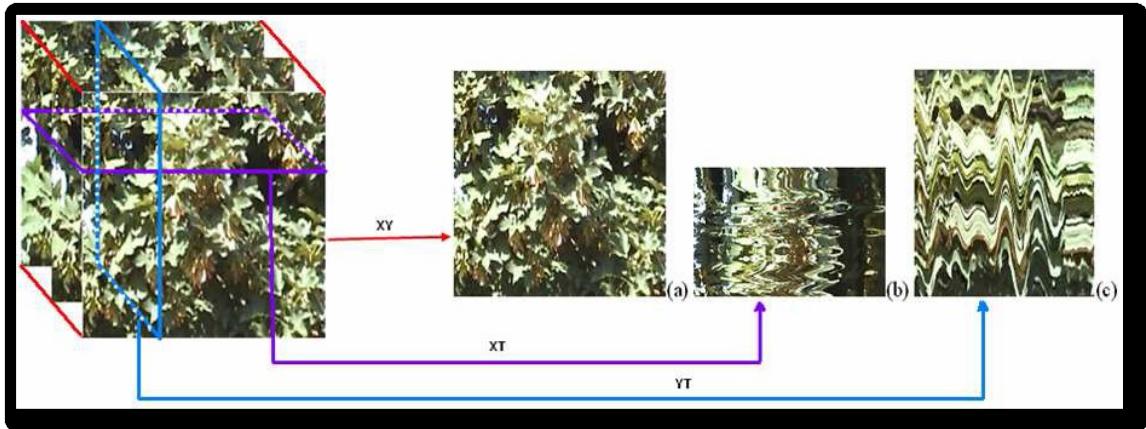


图 4.9:一个 Sequence 组成的体空间可以分出三个正交的二维平面

从图 4.8 和 4.9 可以看出, LBP 算符作用在 YT 平面上时, 得到的是垂直方向上的纹理时间变化信息, 而作用在 XT 上则得到的是纹理在水平时间变化信息。在人脸表情分析时, 三个平面上的 LBP 特征则组成了表情随时间在垂直和水平方向的变化信息, 并且也包含了空间平面上的纹理信息。在阐明了 VLBP 的基本概念后, 本文的 VLBP 算附就是在这样三个平面上进行常规的 LBP 运算。于是, 我们定义 VLBP 的一般操作符 $VLBP_{(xr,yr,tr,Pxy,Pxt,Pyt)}$, xr , yr , tr 椭球的三个方向上的半径 (见图 4.10), 如果这三个半径各不相同, 我们在三个平面上采样就是在三个椭圆周上进行; Pxy , Pxt 和 Pyt 分别代码在三个不同平面上采点的数量。

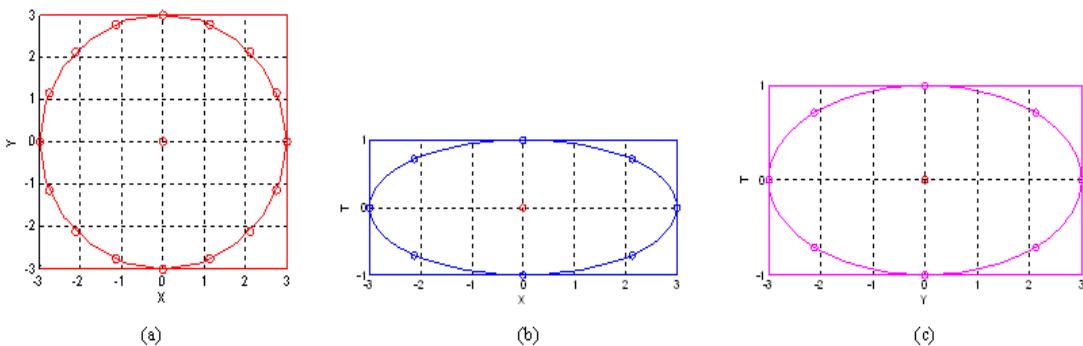


图 4.10: 在 Sequence 组成的体空间中, 截取 XY, XT 和 YT 三个平面

为了有效地减少操作后的特征长度, 本文采用直接将三个平面上得到的 LBP 特征相连 (如图 4.11 所示), 如果对样本还在空间上进行分块 (如图 4.12 所示), 则最终的特征长度将由下面的式子决定:

$$Len = (2^{Pxy} + 2^{Pxt} + 2^{Pyt}) \times m \times n, \text{ 其中 } m \times n \text{ 为分块数量}$$

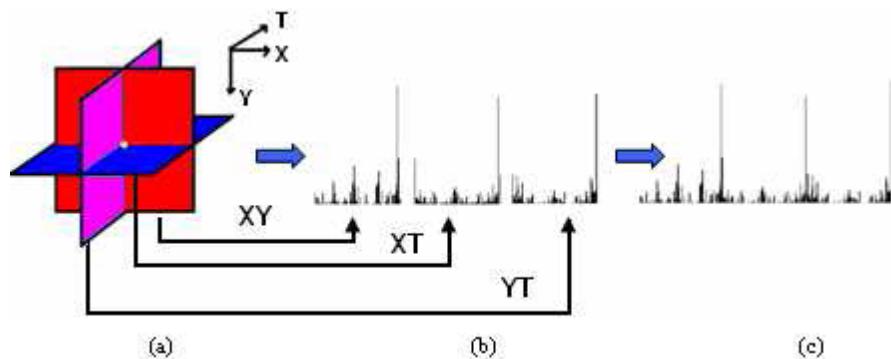


图 4.11: 三个平面上的 Histogram 直接相连生成 VLBP 的特征 Histogram

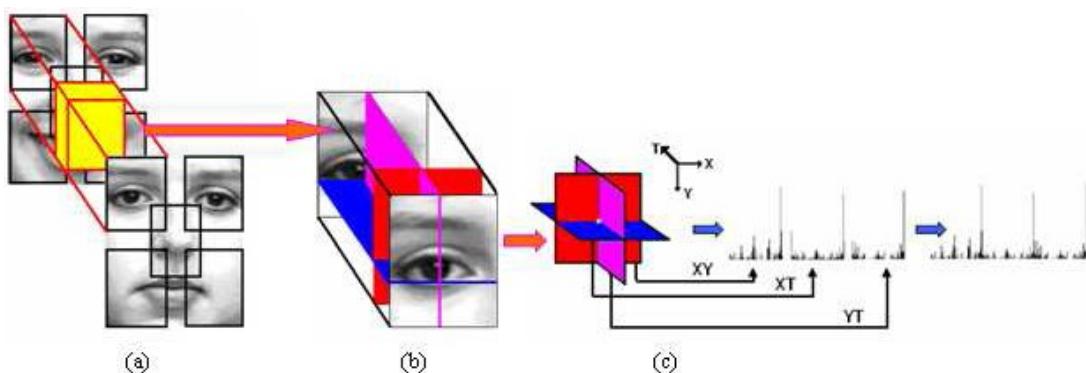


图 4.12: 分块的 VLBP 变换过程

4.4.2 旋转不变性

由数帧连续的图片组成的 **Sequence** 在采集是可能会由于视频中的人脸会有前后低头或抬头，我们在 **YT** 上的特征就会因为这样的旋转产生不确定的一些特征，不利对表情的分析。同样，当视频中的人脸左右摇头时，也会在 **XT** 平面上产生旋转变化，也会造成不利的影响；**XY** 上的也会因为人脸摆动而发生旋转变化。为了消除这些因素对特征的不利影响，需要进行旋转不变性操作，这样的操作主要有两种方法。

1, 使用 Uniform Pattern 的 LBP

在对 LBP 算符操作后的结果进行研究时, Ojala^[36,37]等发现生成的 LBP 特征有相当大的一部分（大于 70%）是 Uniform 的，也可以称为样式一致。所谓的样式一致是指当对一点为中心的局部进行阀处理后得到的由 2 进制表示的数字最多只包含两处 0 和 1 的变化。如 1000111、11110000 都是 Uniform 的；因为第一个 0 和 1 的交替变化只在左边第 1 位第 2 们，第 4 们与第 5 位存在 0 和 1 的变化，第二个也只在

第 4 与第 5 位有变化。而 10101010, 1001001001 不是 Uniform 的，因为它们不止存在 2 个 0 和 1 变化。

有了 Uniform 后，所有的具有同样数量的 0 或 1 的 LBP 特征的点都放在同一个 Bin 里面，如 11110000, 10000111, 11000011 都放在同一个 Bin 里面，而不需要三个不同的 Bin，其它非 Uniform 的放在另一个 Bin 里面，这样不仅能有效地减少旋转带来的不利影响，同时也可以有效地减少特征的长度。

2. 使用移位技术进行旋转不变性操作

学过汇编语言我们知道，一般的 CPU 都会有循环移位指令，即把一内存内的 2 进制表示进行左（右）移位，被移出的那一位数被就接在最右（左）边。（如图 4.13 所示）

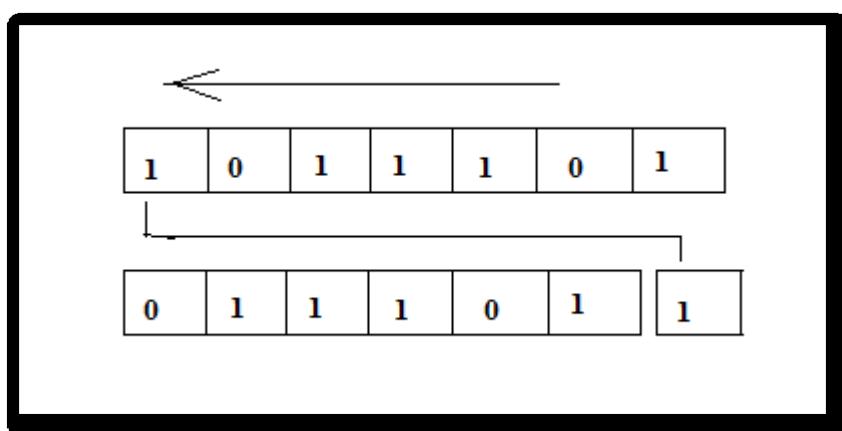


图 4.13：一个左循环移动操作示意图

从图 4.13 可以看到，循环移位指令的操作数有多长，同一个数就可能对应多少种不同的值。如拿 8 位的 1000 0000，能过循环移位操作，我们可以得到另外 7 个不同的数，即 0100 0000、0010 0000、0001 0000、0000 1000、0000 0100、0000 0010、0000 0001。如果我们把这 8 个数中的最小的数取出来做为它们的代表，即可以消除由于旋转操作所带来的这么多种不确定结果。因此，通过循环移动操作把所有的 LBP 特征都操作到所代表的数的最小值，可以有效地消除由于样本在空间和时间上的旋转所造成的影响。此外，通过这样的操作，可以让原本没有通过此操作的 LBP 特征表示中的一些 Bin 里面的点都合并到最小编号的那个 Bin 中，就可以让很多 Bin 变成空的，从而极大地减少有效 Bin 的数量，也即有效特征的数量。我们定义这样的操作作为 CIRToMin (V)，即将 V 这个数值，通过循环移动将其变成移动过程中最小的那

个数值。如 $\text{CIRToMin}(1000\ 0000) = 1$, 有经过此操作的 LBP 特征表示在每个子块的所对应的 Bin 中会放比没有经过此操作的 LBP 特征更多的点, 在数值上反应出来就是比较大的数(如图 4.14 所示)。

在图 4.14 是在其它参数完全一样的情况下, 有进行移位操作与无此操作的 VLBP 特征对比, 是在特征的同一个窗口处进行的。红线以上的部分无 CIR 操作的, 而红线以下的是有此操作的。蓝色的窗口标出了每一个子块中三个不同平面的 LBP 特征, 每个平面的特征长度为 256, 窗口中是三个平面的开头若干大小的特征块, 红线上下的窗口是为了对比的。可以看出有经过移位旋转不变化性操作的特征的开头部分明显要有更密集的点分布 (4.14 的蓝色窗口)。这样操作的有效性在我们所使用的 CDK 数据库不能得到验证, 因此 CDK 的所有样本都是很正的样本, 无明显的旋转角度。因此, 我们只能通过实验来比较有些操作与无此操作对在此数据库上的分析精度的影响。

在研究有无旋转操作对精度的影响时, 我们选用了 $\text{VLBP}_{(3,3,3,8,8,8)}$ 和 PLBP_{3-8} 来作为实验模型, 分块大小均为 32×32 , 即把 96×96 大小的图片分割成 9 个小块。由于 VLBP 是采用 Sequence 作为样本, 每个 Sequence 的长度不一样, 为了克服这个差异给特征生成造成的困难。本文采用了的方法是: 设一个 Sequence 的长度为 N, 把这些图片编上 $0, 1, \dots, N-1$ 的编号, 最小表情强度 (中性表情的) 的一张编为 0, 以此类推。VLBP 中的 T 轴上的半径为 R, 当 R 与 T 轴成 k 度角时, R 在 T 轴上的投影为 P = $R \times \cos(k)$, 我们用以下公式将 P 映射到这 N 张图片上的某一张编号为 M 的图片:

$$M = \frac{(N-1) \times P}{2 \times R} + \frac{N-1}{2} \quad (4.12)$$

其对应关系可以从图 4.15 中看到, 即中间的那张定义为 P=0, 最右那边的坐标为 P=R, 而最左边那张为 P=-R。表 4.4, 图表 4.4 与图表 4.5 给出了实验结果, 其中带有 RI (Rotation Invariant) 为有进行旋转不变操作的。

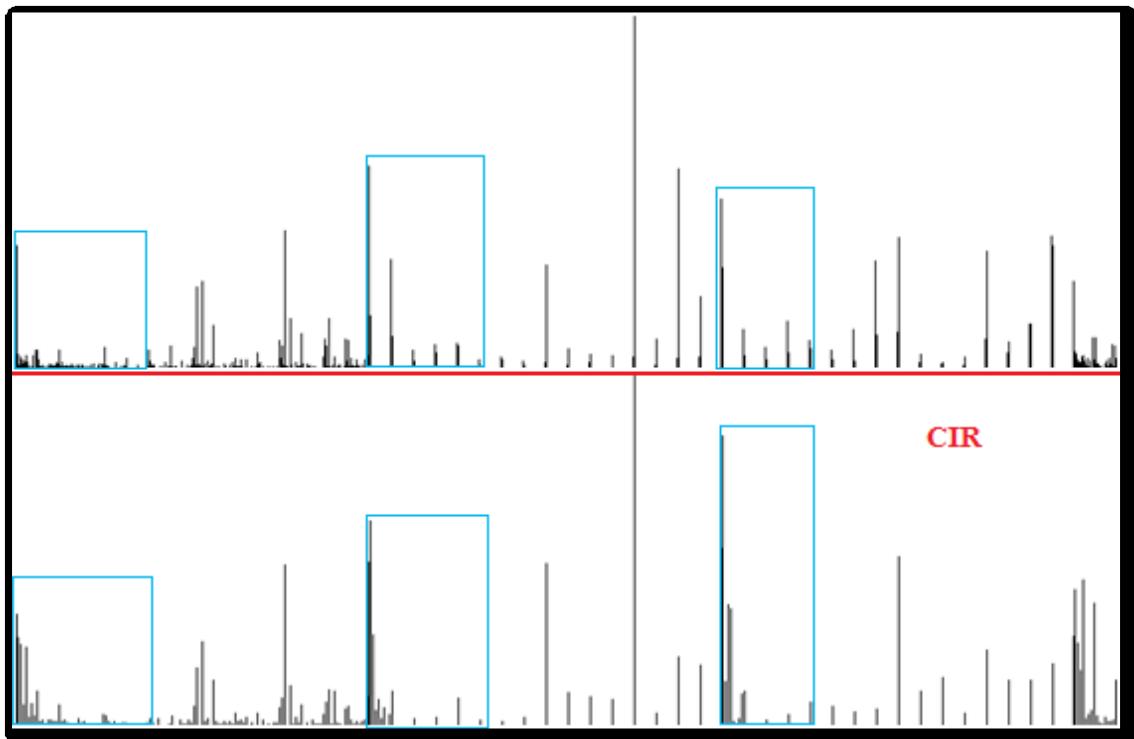


图 4.14：循环移位操作与无此操作的特征比较

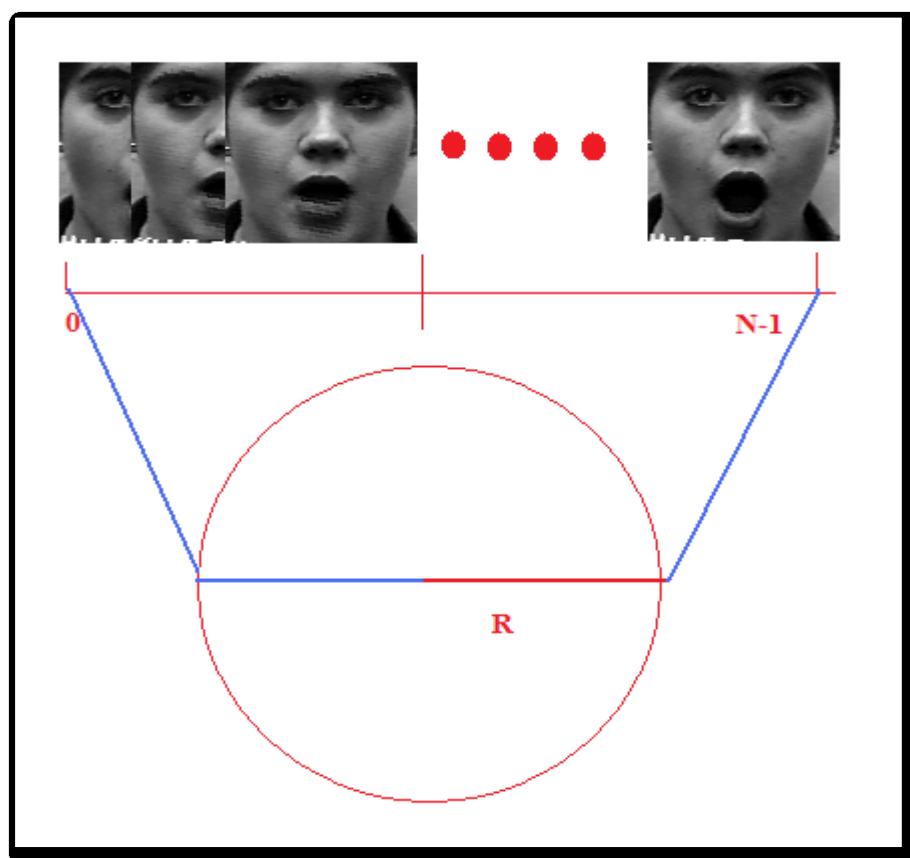


图 4.15：Sequence 中的图片与 T 轴上的坐标对应关系

表 4.4: 有旋转变换与无此操作的实验对比

RI	VLBPI _{RI} (zr=0.5)	VLBP (zr=0.5)	PLBPI _{RI} (zr=1.0)	PLBP (zr=1.0)	PLBP(AIR-RI)
Anger	0.8864	0.8864	0.8770	0.9148	
Disgust	0.9112	0.9274	0.9464	0.9590	
Happy	0.9112	0.9242	0.9180	0.9242	
Fear	0.8927	0.8864	0.8391	0.8769	
Sadness	0.8485	0.8675	0.8643	0.8738	
Surprise	0.9437	0.9526	0.9779	0.9842	

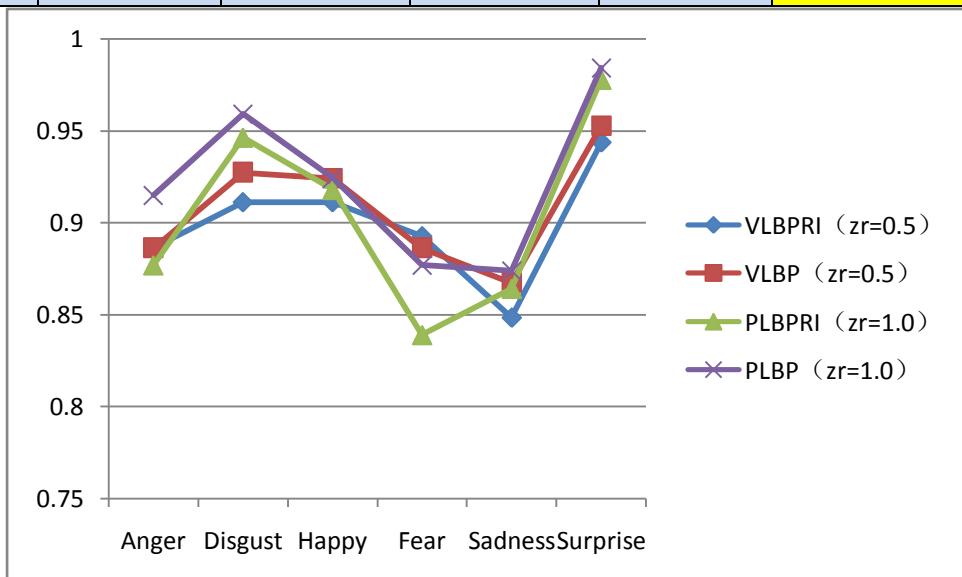


表 4.4: VLBP 与 PLBP 中有进行旋转不变性操作与无这操作的对比

LBP 特征的静态分析与动态分析研究

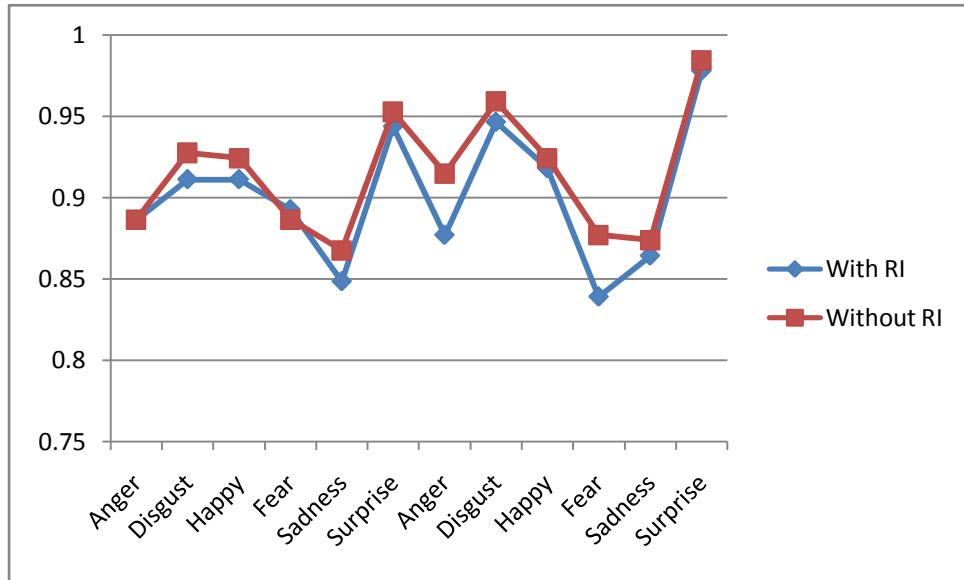


表 4.5：有无旋转操作的精度比较

有旋转不变性操作后的特征与无此操作后的特征明显的区别是前都会有更多的空箱子 (Bin)，因此在系统生成的有效特征 VS 零率的统计图会在一开始变化很慢，到后面时急速上升，而后者基本是比较平稳地增大（对照图 4.16、图 4.17、图 4.18 和图 4.19）。因此，如果采用同样的零率值 ($zr < 1.0$)，则后者的有效特征会较前者多，有效特征的增多就有可能使得后者的精度更高。

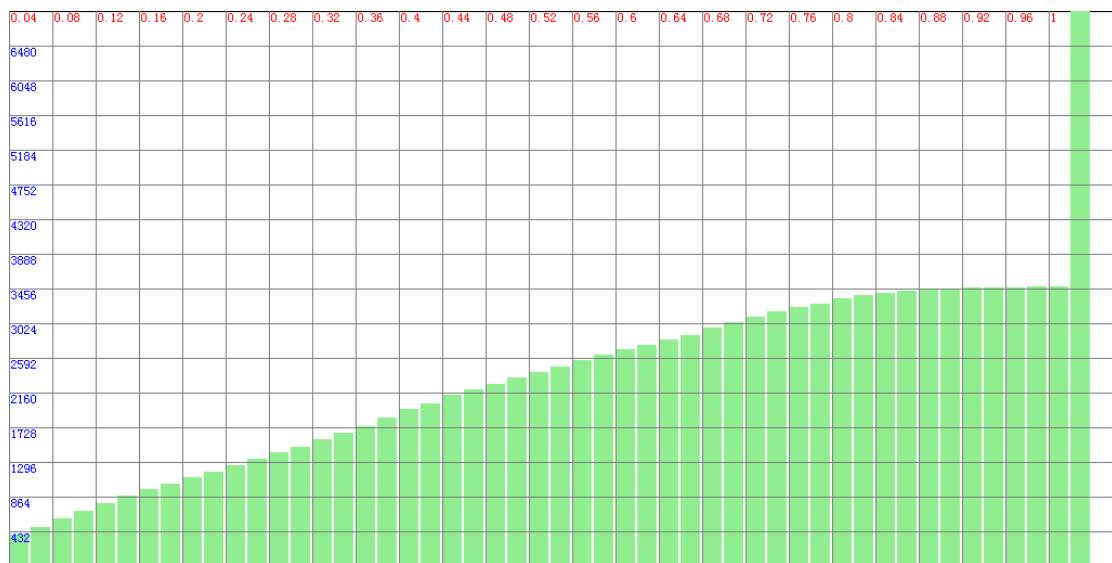


图 4.16：VLBP 有旋转的有效特征数 VS 零率

LBP 特征的静态分析与动态分析研究

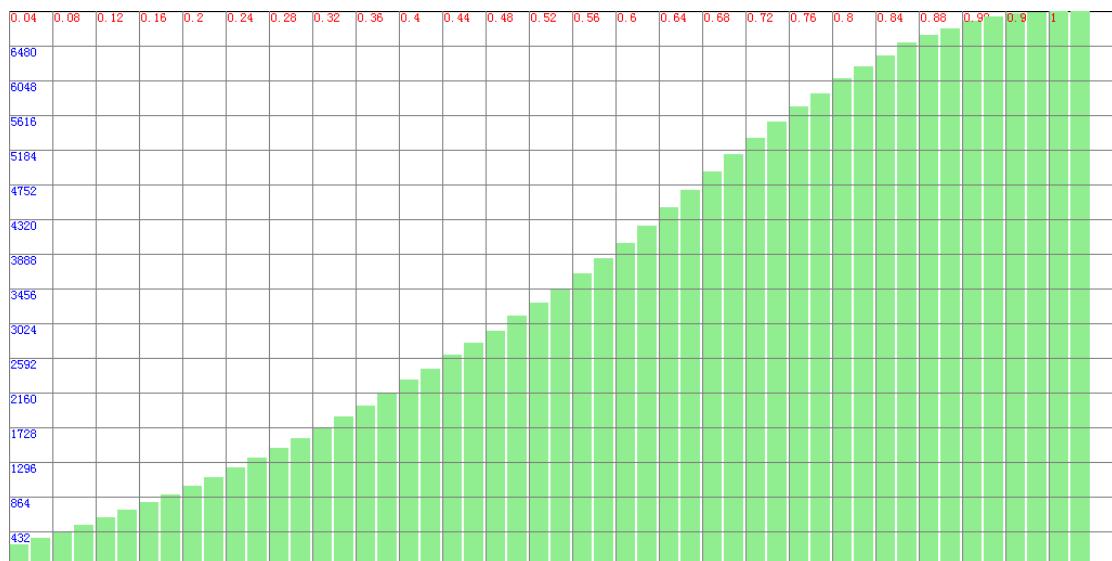


图 4.17: VLBP 无旋转的有效特征数 VS 零率

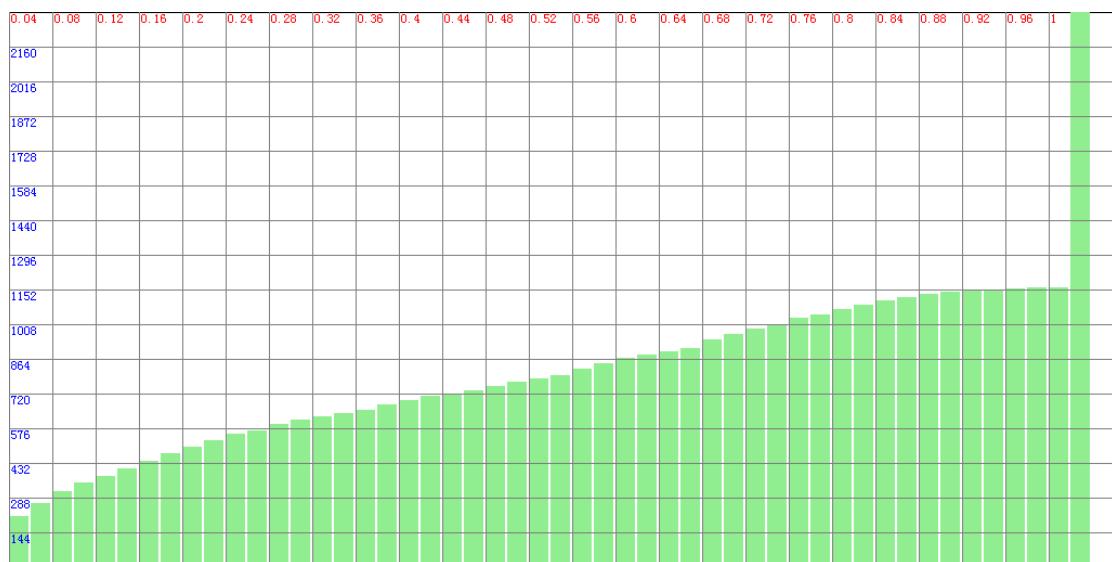


图 4.18: PLBP 有旋转的有效特征数 VS 零率

LBP 特征的静态分析与动态分析研究

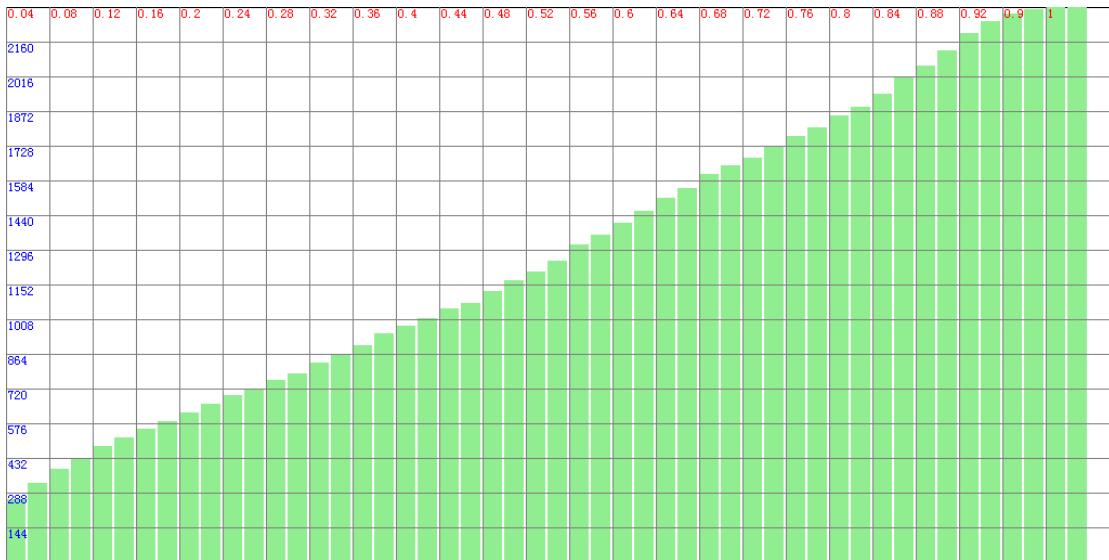


图 4.19：PLBP 无旋转的有效特征数 VS 零率

从表 5 我们也可以发现，在这个实验中，VLBP 的精度并没有比 PLBP 的要高，有些表情反而要低于后者。这样的结果并不奇怪，一个很重要的原因是在生成的 LBP 特征表示中，我们对于非常大的特征用基于零率的方法去除那些无效子特征，并没有选出对于某个表情的显著特征，而是把所有的有效特征都包括进去，不管这个有效特征对于所分析的表情是否是起积极的显著性或消极的，本文将在后面的章节中采用 AdaBoost 选取出不同表情的显著特征，然后再比较 VLBP 与 PLBP 的差别。此外，还可以发现不同的表情对于有无旋转不变操作的敏感度是不同的，PLBP 中采用的是全部特征来训练的，因此有无旋转不变操作的精度差异就全部来自这个操作本身，与有效特征数无关。表 4.4 的后两列明显地表情有旋转不变操作对于 PLBP 来说影响非常大，进行过旋转不变操作的 LBP 特征的训练精度会大大地低于不进行此操作的精度。这个问题的原因可以做如下分析：

旋转不变操作本身是利用循环移位操作来找到最小的数值，由原来的数变化到最小数所需要的移动位数决定着旋转的角度的大小。需要的位数越多，旋转的角度越大，对于一个可以用 8 位二进制表示的 LBP 特征，每移动 N 位对应的旋转角度为：

$$\theta = \frac{2\pi}{8} \times N \quad (4.13)$$

当 N=8 时，意味需要在某个平面上人脸旋转 180 度才可能需要移动 4 位。然而，现实的情况是我们几乎不可能旋转如此大的角度(即使是左右摇头，一般人也很难做到左右的摇动角度为达到 180 度)。更一般的情况是我们的脸部在很短的时间内就可

以完成一个表情的生成和消退，一般就短短几秒甚至几百毫秒；这样短的时间内我们是不会有太大的左右、前后或上下旋转的。因此，Zhao 在其文章中指出的旋转不变操作很大情况下并不会有利于表情的分析，从图表 4.5 可以清晰地得出这样的结论。当然，对于那些旋转角度不定的样本，在用它们来进行训练时，进行旋转不变操作也许有可能提高训练的精度，本文目前由于条件限制无法验证这个猜想。

4.4.3 本节小结

本节中介绍了两种用来消除由于样本旋转造成的差异，并通过实验和数据分析给出了这种操作在特征生成、训练精度及现实应用中的差异分析。在本章随后的小节中，本文将分析更多的动态表情分析技术或模型。

4.5 使用 AdaBoost 进行特征选择

为了说明前一部分提出的关于显著差别的想法，我们在没有进行过零率降维处理过的 VLBP 特征上使用 AdaBoost 特征选择器进行，选出 1000 个有效特征，然后将结果输入到 SVM 进行分类训练，并进行 leave-one-out-all^{3[38]}的交叉验证，这样的策略不存在着分组差异带来的精度不确定性，每个组只有一个样本，因此每次验证的结果是一样的。为了客观比较与没用 AdaBoost 的精度，我们利用零率降维（取 $zr=0.1$ ，使得有效特征差不多也为 1000 维）的方式获取另一种结果，并进行 leave-one-out-all 验证，由于这样的交叉验证需要非常多的时间，我们之前的实验并没有采用这样的方法，而是应用比较折中的方法，但之前的精度都是多次实验的平均值，因此也保证了客观可靠。实验的数据如表 4.6 和图表 4.7 所示。

此外，由之前的说明可知，LBP 特征特征有一个特点就是很多子特征基本都是零，尤其是进行过旋转不变性操作后的特征更是密集地分布在编号小的箱子里，使用后面的很多箱子全为空。从 AdaBoost 的弱分类器选择方式可以知道这样的特征分布很不利于 AdaBoost 的特征选取，在用 AdaBoost 选取出显著特征后，我们通过实验方法分析了最后特征的阀值情况，发现有很多弱分类器的阀值都是 0，这样的值对于每个子特征都大于或等于 0 的特征来说，相当于无效，因为它会把任何样本统统归

³即每次训练只有一个样本不参与训练，再用其它样本训练的结果来对这未参加训练的样本进行分类

到正样本或负样本。Boost 理论的弱分类器是指一个函数 $h(x)$, 它会把特征为 x 的一个样本分类到正样本或者负样本, $h(x)$ 由下式定义:

$$h(x) = \begin{cases} 1, & \text{if } p_j f_j(x) < p_j \theta_j \\ 0, & \text{otherwise} \end{cases} \quad (4.14)$$

其中 θ_j 就是弱分类器 $h(x)$ 的阀值, $f_j(x)$ 特征 x 的第 j 个子特征, p_j 由 $h(x)$ 的方向偏置, 只取 +1 或 -1, 具体的说明在此不再赘述。从式 4.14 可以很容易就得到我们上面说到的无效分类器的原因: 由于 $f_j(x) \geq 0$, 假设 $\theta_j = 0$, 那么无论对于什么样本,

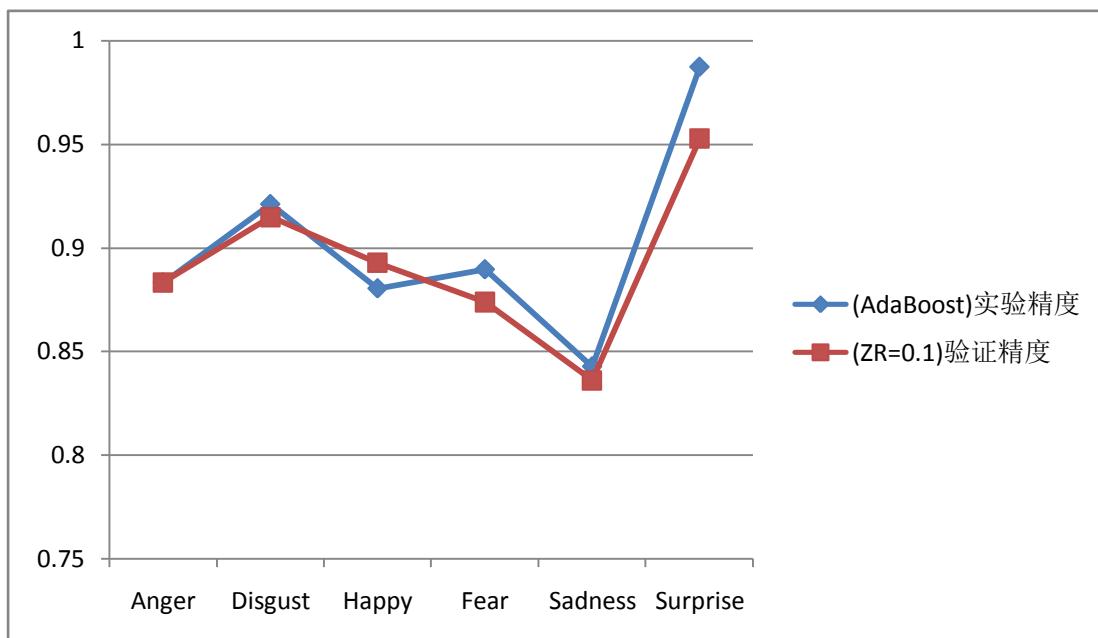
$$h(x) = \begin{cases} 1, & \text{if } p_j < 0 \\ 0, & \text{otherwise} \end{cases} \quad (4.15)$$

从表 4.6 和图表 4.7 可以看出, 在同等维度下, AdaBoost 选择出来的特征略优于基于零选出的有效特征。在对比先前的实验数据, 我们也可以看到, 当 ZR 值大一点时, 基于零率的实验精度略优于 AdaBoost 选择出来的 1000 维, 再次说明基于零率的特征选择是有效的。但在相同维度上, AdaBoost 会比较出色; 基于以上讨论, 我们有必要研究在用基于零率算法去除那些无效特征后, 使用 AdaBoost 再次进行特征选择的效果, 为了尽可能地保留有效特征, 我们使用 ZR=0.8 来消去那些零率太大的子特征(每个样本的特征维度由 27648 降到 9952)。图 4.21(原来的分布图见图 4.20)给出了上述操作后的零率 vs 有效特征图。

表 4.6: 同等维度的训练精度对比。

VLBP	实验精度	显著特征数	验证精度	有效特征数
Anger	0.8832	1000	0.8833	1004
Disgust	0.9211	1000	0.9148	1004
Happy	0.8804	1000	0.8927	1004
Fear	0.8896	1000	0.8738	1004
Sadness	0.8427	1000	0.8360	1004
Surprise	0.9874	1000	0.9527	1004

LBP 特征的静态分析与动态分析研究



图表 4.7：两种不同特征选择方式图表分析

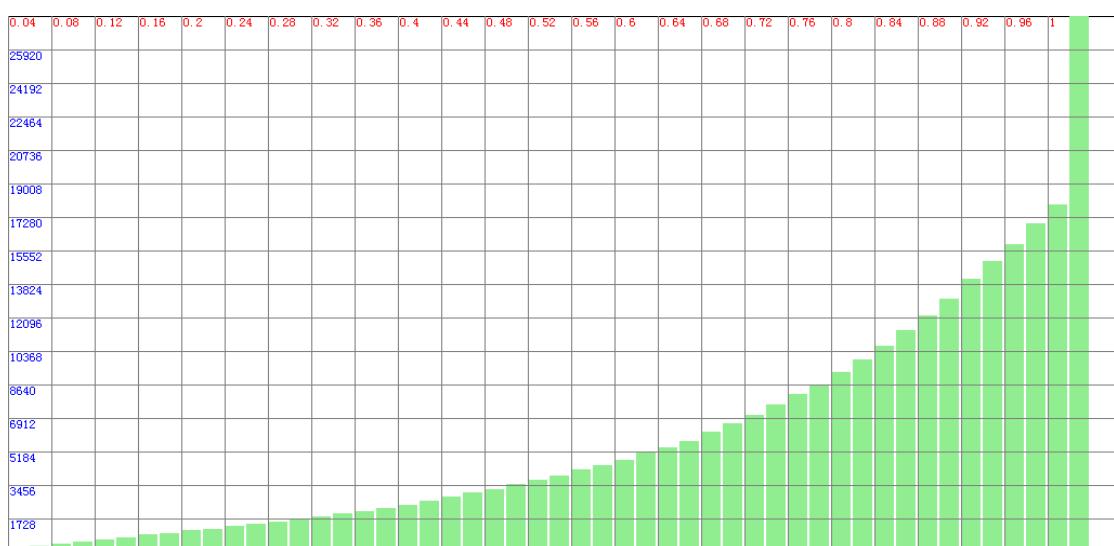


图 4.20：原始特征的零率 VS 有效特征图

LBP 特征的静态分析与动态分析研究

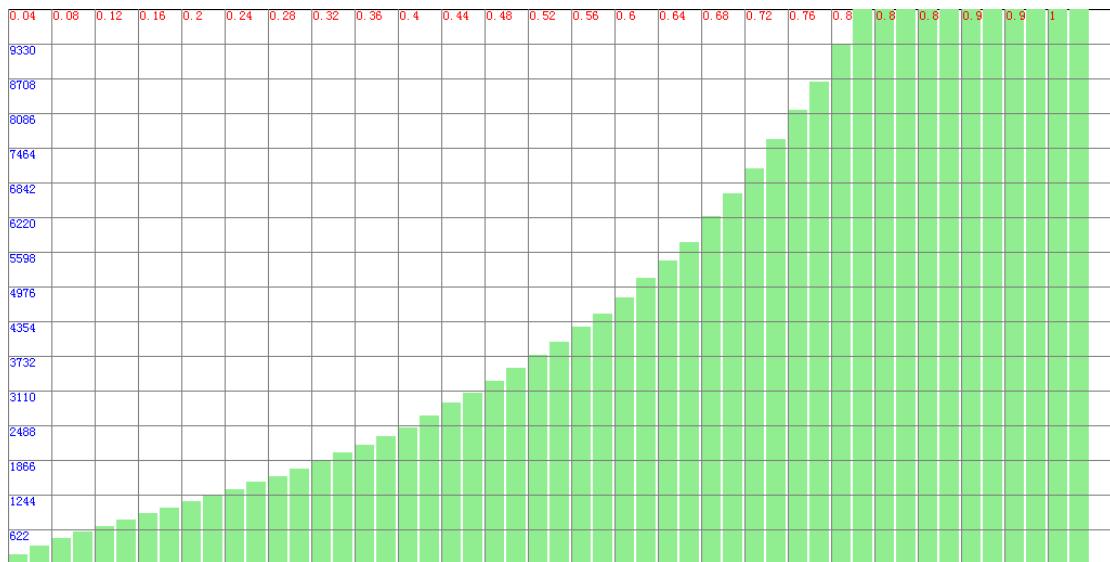


图 4.21：去除零率大于 0.8 后的零率 VS 有效特征图

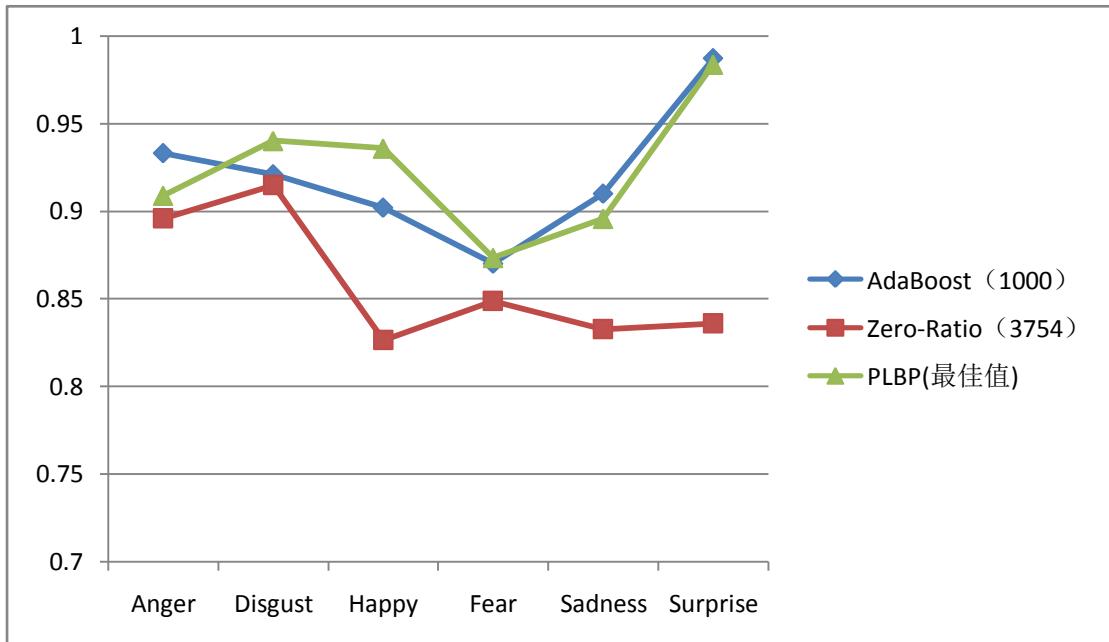
从图 4.21 可以看出，进行过零率降维后，有效特征数量不会随 RC 值变化出现急速变化，这也为 AdaBoost 的特征选取创造了良好条件，以下实验数据给出了从处理后的特征再选出 1000 维后与 ZR=0.5 的实验精度比较。

表 4.6: 不同的特征提取方法与 LBP 模型的精度比较

VLBP(36 分块)	AdaBoost (1000)	Zero-Ratio (3754)	PLBP(最佳值)
Anger	0.9332	0.8959	0.9089
Disgust	0.9211	0.9148	0.9401
Happy	0.9021	0.8265	0.9358
Fear	0.8702	0.8486	0.8734
Sadness	0.9101	0.8328	0.8957
Surprise	0.9873	0.8360	0.9837

从表 4.6 可以看出，在一般情况下，基于零率的降维方法在 VLBP 上选出的有效特征并不都是显著的，以致训练精度在很多情况下都不如 PLBP 的效果好。使用 AdaBoost 选择的特征都是显著的，能够比较好的发挥动态分析的能力，但是在实验过程中，由于 VLBP 产生的特征向量中包含有太多无效的特征，直接将这些特征输入到 AdaBoost 进行训练会得到很多无效子特征。借助基于零率的降维算法适当地消除无效特征可以很好地强化 AdaBoost 的训练效果，因此，在进行显著特征提取时，结

果两者的特点能更好地提高分类的训练精度。



图表 4.8：表 4.6 的图形化显示

有些文章指出动态分析方法比静态方法要好，这点在我们的实验中并没有得到非常好的支持，而只是得到部分支持，从图表 4.8 可以看出，对于一些表情，动态方法确实比较好，但另一些则不然，原因多少也是可以从我们的生活经验找到。举例说，对于生气这个表情，我们可以有意识地表现出生气，但如果让我们对于一个静态的表情进行判断，我们有可能无法分辨一个静态的生气表情是痛苦或是生气，我们只能通过这个表情的动态变化才能更好地分辨。而有些表情是非常明显的，如高兴，我们只要微笑，他人就能分辨出我们脸上挂的是微笑的或高兴的表情，这也解释为什么在机场进出口经常能看到一个静止微笑站立着的模型服务员，因为我们不需要其有个变化的微笑就能知道她是在微笑。因此，在动态分析对生气这个表情的分析起的是正面的效应，而对高兴这类表情并不一定是正面的，还有可能在特征选择时干扰了静态中重要的特征，因此有可能起到负面效应。

4.6 一种新的 LBP 分块模型

在经过大量的实验和对实验数据的分析后，我们发现分块对于 LBP 对表情分析的准确度影响非常的大，考虑到 LBP 特征本身是描述区域特征的，所以分块的区域

决定了特征最终描述的是那些位置的特征。回顾 FACS⁴ (Facial Action Coding System) [39] 中关于 AU (Action Unit) 的分类方法，我们不禁发现 FACS 的编码很大程度上依赖局部肌肉伸缩状态，而且不是所有肌肉对表情的作用是一样的，而是有很大的差异。如对于 Surprise 这个表情，从前面的实验分析我们可以看到这个表情的训练精度总是远远大于其它表情，一个很重要的原因是 Surprise 在 FACS 中包括了两个显著的表情单元 (AU)，一个是口型张大，另一个是眼睛睁大；而这两个表情单元所处的区域在分块时，都能比较完整地占据周围的某些子块(如图 4.22 所示)，这样，这两个表情单元的特征就能在这些子块的 LBP 特征中突显出来，在训练时在训练特征的子空间中就能很好地将 Surprise 的样本与其它的本校区分开来。对于没有分块的特征，局部的显著特征被整体的统计所模糊，以致没有分块的情况下基本没有能力有效分辨出表情的成分。



图 4.22：两种不同的图片分割方式

然而，一般的分块都是矩形的，有些区域是有重叠，有的则没有重叠，有重叠与没有重叠在实验中也被证实^[18-19]对精度有影响，是这是因为有些表情单元的显著区域是重叠的。受到这样的启发，本文提出了一种更一般的分块方法，这样的分块方法可以把人脸图片的主要区域分割成多个不规则的子区域的并集（如图 4.23）。

由于每块子区域都可以是不规则的，区域大小以蒙版 (Mask) 的方式定义，只有蒙版定义的区域内的点才参与 LBP 特征的 Histogram 生成，即只有在蒙版定义的区域内的点的 LBP 特征参与生成统计特征，因此我们称这样的模型为不规则区域 LBP (Irregular Sub-Region LBP，缩写为 ILBP)

⁴ FACS 是一种基于人脸面部肌肉对表情状态进行分割编码的一种编码规则

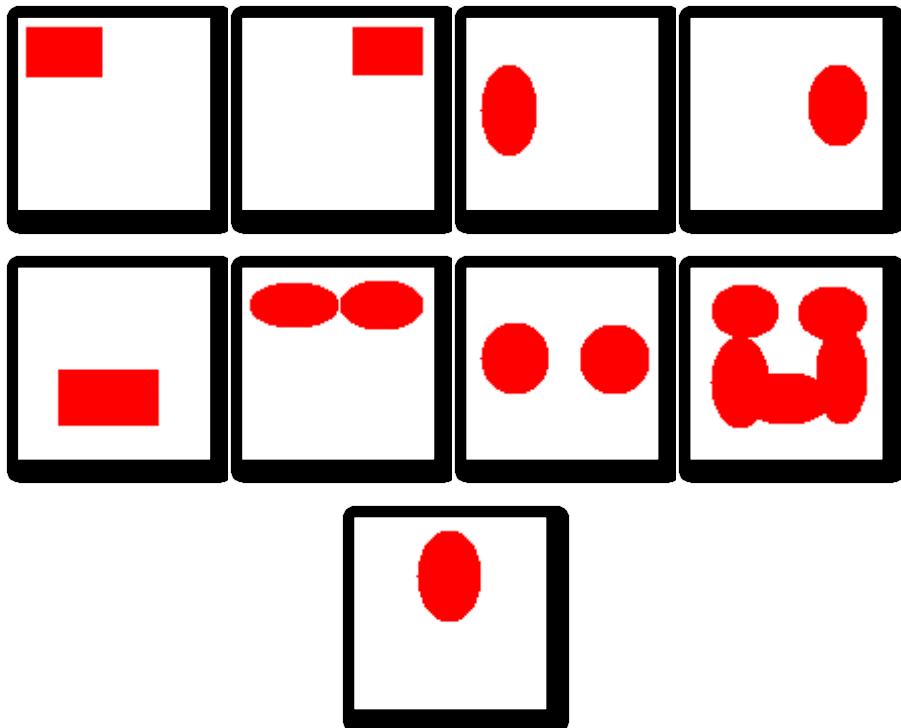


图 4.23：一组分块区域

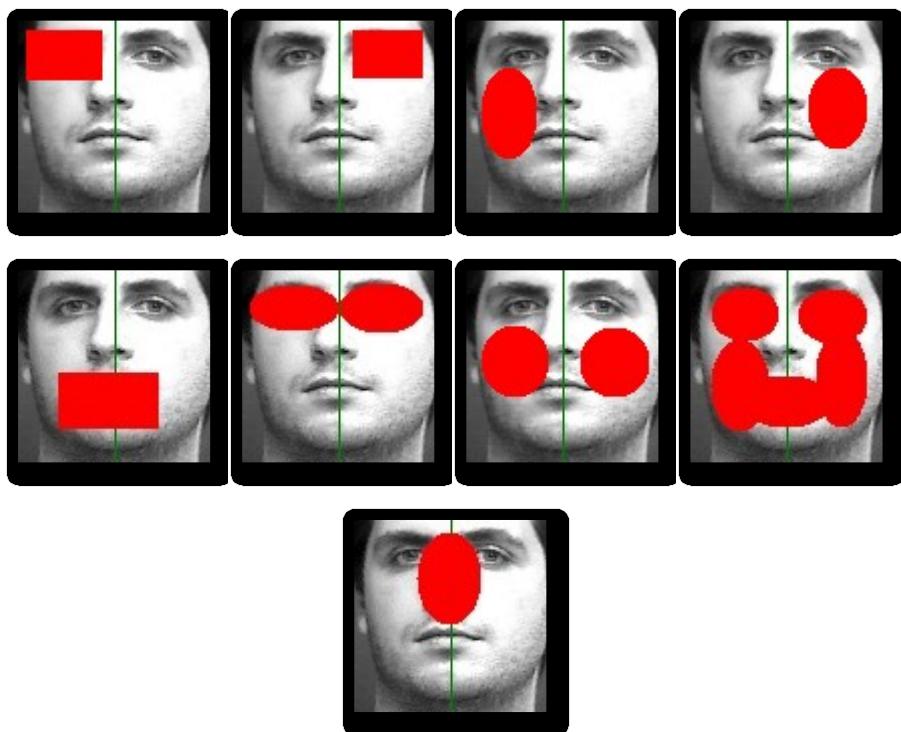


图 4.24：一组分块区域与样本的对应图示

图 4.23 中的 8 个区域定义了一个 ILBP 模型,每一个蒙版图片定义一个子区域,红色区域即为有效区域 (如图 4.24 所示), 只有这些区域内的 LBP 特征参加生成这

LBP 特征的静态分析与动态分析研究

个区域的 Histogram 特征，所有的子区域的生成特征相连形成完整的一个表示特征。可以这样理解：4.23 中的第一张蒙版图片定义的是左眼的区域，这个区域的生成特征可以很好地反应左眼的状态特征；第二张对应的是右眼的有效区域；第三张对应的是左脸颊的区域，其是椭圆区域；第四张对应右脸颊的区域；第五张对应嘴巴的区域；第六张对应双眼的区域，这样区域是要通过合并两个眼睛区域来获得更大范围内的眼睛变化特征，对于某些表情单元有比较好的作用。同理，第 7 张对应两边脸颊的主要区域；第 8 张对应更大的一个区域，它是眼睛，脸颊和嘴巴主要区域的集合，这个区域生成的特征能反应这些局部的变化情况；最后一张对应的则是鼻子的区域。由于得到的 Histogram 特征反应的是这个区域的纹理变化，也可以说是皮肤因为肌肉伸缩的变化引起的变化。

为以验证 ILBP 的有效性，本文以图 4.23 中的 9 个区域作为 ILBP 的操作区域，用平面 LBP 的方式在 9 个红色块对应的人脸区域上进行静态的 LBP 特征提取。由于是不规则的区域，9 个 Histogram 在合并后输入 SVM 之前要将整数的 Histogram 特征转换成浮点数据，这需要进行特征的校准。本文采用的方案是将得到的每个区域得到的 Histogram 用对应区域内的有效点总数除（见算法 4.1，为 C++. net 源代码），以使该区域得到的 LBP 特征是一单位向量。每个区域得到的 LBP 特征都是这样单独的单位向量，所有的这些区域的特征共同组成了一个完整的 LBP 特征表示。表 4.7 最后一列给出了基于上述 9 个区域的实验结果，图表 4.9 给出了直观的曲线对比。

LBP 特征的静态分析与动态分析研究

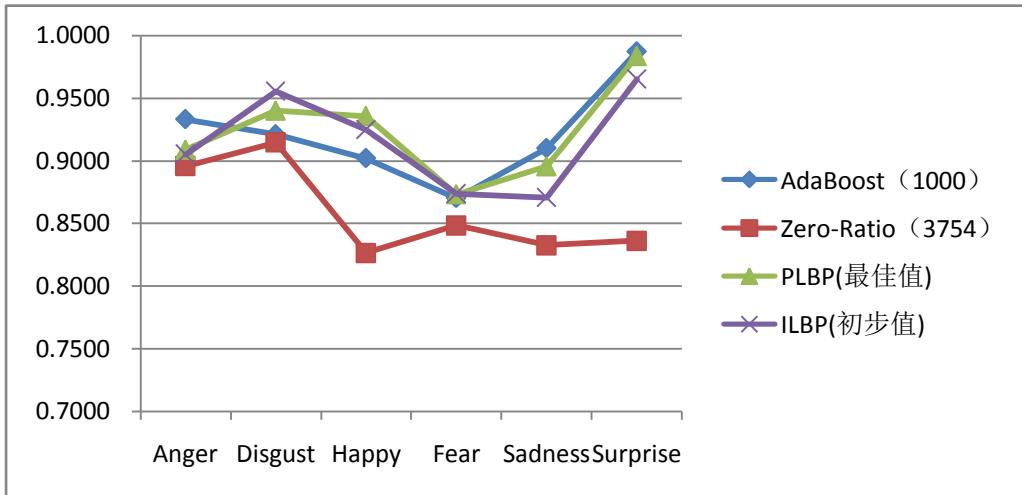
```
void Convert.ToDouble(array<int>^ input, array<double>^ output)
{
    int index = 0;
    double scaleSize = 0;
    ILBPMODE::IRegion^ reg ;
    for(int i = 0; i < subRegions->Length;i++)
    {
        reg = subRegions[i];
        scaleSize = (double)(reg->RegionHitCount ); //scaleSize为区域内的有效点数
        //RegionHistSize为区域的Histogram大小，即Bin的数量
        for(int j = 0; j < subRegions[i]->RegionHistSize;j++)
        {
            output[index] =(double)input[index] / scaleSize;
            index++;
        }
    }
}
```

算法 4.1: ILBP 中 Histogram 转换标准特征的过程

表 4.7: ILBP 初步实验结果与其它实验的对比

VLBP(36 分块)	AdaBoost (1000)	Zero-Ratio (3754)	PLBP(最佳值)	ILBP (静态)
Anger	0.9332	0.8959	0.9089	0.9052
Disgust	0.9211	0.9148	0.9401	0.9558
Happy	0.9021	0.8265	0.9358	0.9274
Fear	0.8702	0.8486	0.8734	0.8738
Sadness	0.9101	0.8328	0.8957	0.8707
Surprise	0.9873	0.8360	0.9837	0.9653

对比表 4.7 和图表 4.9 可以看到, ILBP 总体表现良好, 在 Disgust 这个表情的分析中取得了最佳的效果, 并且其它表情的分析也有很好的效果。以上实验也初步验证了 ILBP 有效性, 而接下去的工作就是要寻找能取得更好分辨能力的 ILBP 区域方案, 并将其扩展到动态分析的算法中去。在本章随后的内容中, 会在适当的时候用 ILBP 结合到新的模型中或分析流程中去。



图表 4.9: ILBP 的实验结果与其它实验的比对曲线图

ILBP 的提出是根据 LBP 本身的局部性质 (Local) 而来的, 在 LBP 提出时, 并没有说分块只能是简单的矩形。ILBP 正是充分发挥 LBP 在局部区域的强化显著特征, 在全局淡化局部特征的特点, 在有意识的进行分块后, 可以充分发挥 LBP 的特征表示能力, 以取得更好的模式识别效果。此外, 由于 LBP 这个概论是一般性, ILBP 完全可以应用到模式识别的其它领域, 如使用了 LBP 的 ASM(Active Shape Model)^[40-41], 也可以采用 ILBP 模型来进行人脸特征的定位;此外, 还可以在人脸识别中应用 ILBP 来强化人与人之间的身份差别等。

4.7 不同样本强度下的表情分析研究

本章为本文的主要章节, 在本章我们介绍了本文采用的相对样本强度的概念,之所以引入这个概念是为了研究动态分析与静态分析在相对弱的表情强度下的有效性比较。由前面的实验数据可知, 在使用如 AdaBoost 进行显著特征提取后, 就能够有效地提动态分析的精度, 发挥动态分析的优势。然而, 由于 AdaBoost 的训练时需要大量的时间, 本节在研究不同样本强度对识别精度的影响时, 对于动态模型仅就 Fear 这个表情进行分析, 除了特征选择要很多时间外, 另一个原因是这个表情的精度无论是在 PLBP 还是在 VLPB, 或静态 ILBP 上都不是很令人满意, 我们希望在动态分析中能在这个表情分析上发挥它的优势。

4.7.1 静态分析对样本强度的敏感性研究

由于每个样本都是由数张连续变化的图片组成的 Sequence, 本小节将就在不同样本强度下, 对 PLBP 的分析能力进行评估, 采用实验的强度有 0.5, 0.6, 0.7, 0.8, 0.9 和 1.0。在 1.0 下, 就是最大样本强度, 与之前的实验一样。在实验中, 我们采用了 3×3 分块方案, R=3 和 P=8, 即采用 $PLBP_{3-12-32-32}$ 这个模型来分析不同样本强度下的实验精度, 不进行任何降维处理。表 4.8 和图表 4.10 给出了实验结果。

从表 4.9 可以看出, 当样本强度小于 0.9 时, 其训练后的样本统计分布都存在

LBP 特征的静态分析与动态分析研究

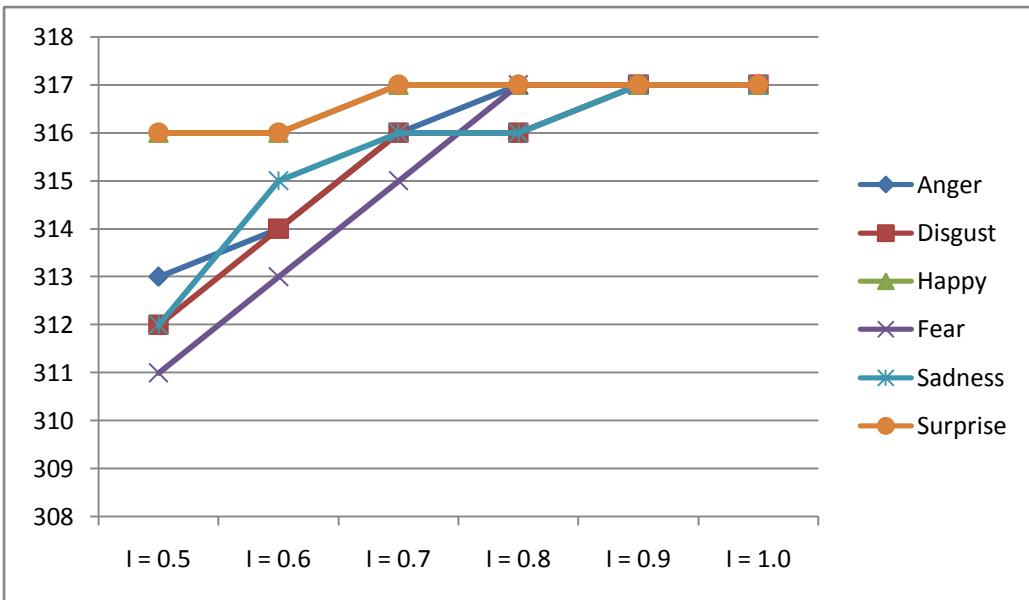
漏分或错分，从 0.5 到 0.9 这个变化过程，漏分与错分的可能性也随着减少，当样本强度为 1 是，和本文之前的实验数据一样，就不存在漏分与错分的情况。表 4.8 的交叉验证精度也是随着样本强度的增加与不断提高的。

表 4.8：不同样本强度下的交叉验证精度

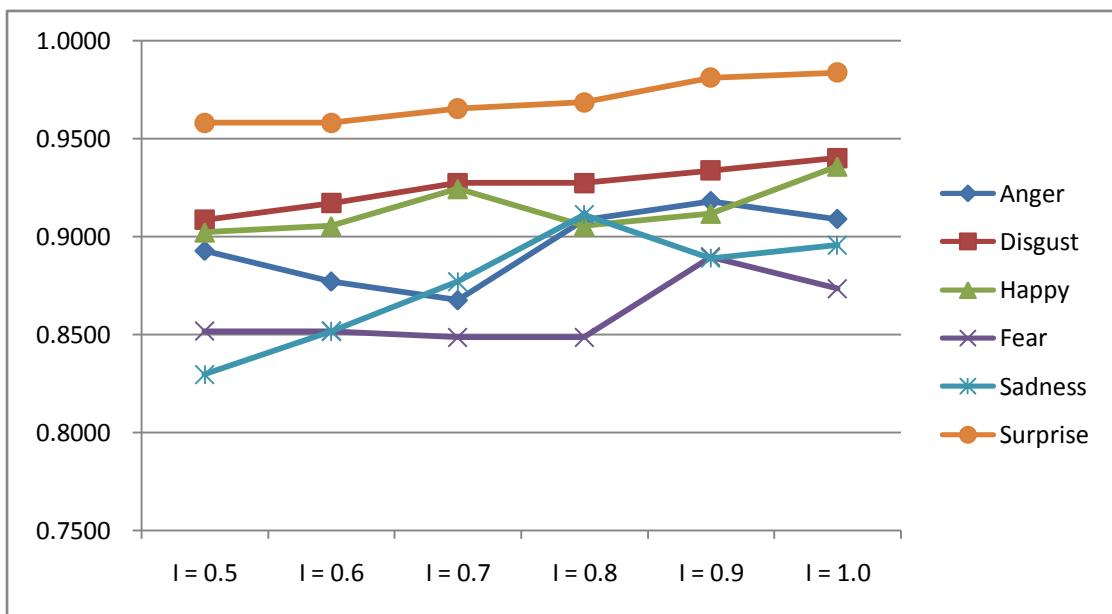
PLBP(9 分块)	$I = 0.5$	$I = 0.6$	$I = 0.7$	$I = 0.8$	$I = 0.9$	$I = 1.0$
Anger	0.8927	0.8770	0.8675	0.9085	0.9180	0.9089
Disgust	0.9085	0.9170	0.9274	0.9274	0.9337	0.9401
Happy	0.9021	0.9053	0.9243	0.9053	0.9117	0.9358
Fear	0.8517	0.8517	0.8486	0.8486	0.8896	0.8734
Sadness	0.8297	0.8517	0.8770	0.9112	0.8890	0.8957
Surprise	0.9580	0.9580	0.9653	0.9684	0.9810	0.9837

4.9：不同样本强度下的训练精度（总样本数为 317）

PLBP(9 分块)	$I = 0.5$	$I = 0.6$	$I = 0.7$	$I = 0.8$	$I = 0.9$	$I = 1.0$
Anger	313	314	316	317	317	317
Disgust	312	314	316	316	317	317
Happy	316	316	317	317	317	317
Fear	311	313	315	317	317	317
Sadness	312	315	316	316	317	317
Surprise	316	316	317	317	317	317



图表 4.10：训练精度随着样本强度地变化情况



图表 4.11：训练精度随着样本强度地变化情况

为了验证分类精度，图 4.25 到 4.30 给出了在实验平台下的分析统计结果。从这 6 个训练结果样本分布图可以更清晰地看出不同样本强度对训练的影响。只有在所有的样本的分布都是完整时（图片中每一个类所在的列上方的三个数字都相等的情况），比较交叉验证精度才是有意义的。从图 4.19 还可以看出，当样本强度是 0.9 时，即使训练精度为 100%，也存在着错分与漏的现象，因此，要评估训练结果好坏，需要结合训练精度，分类统计分布情况及交叉验证精度。本文之前的实验都是在训练精度为 100%，不存在错分与漏分的情况下，仅给出交叉验证的精度，因此只要比较交叉验证精度即可（见表 4.10 和图表 4.11）。

LBP 特征的静态分析与动态分析研究

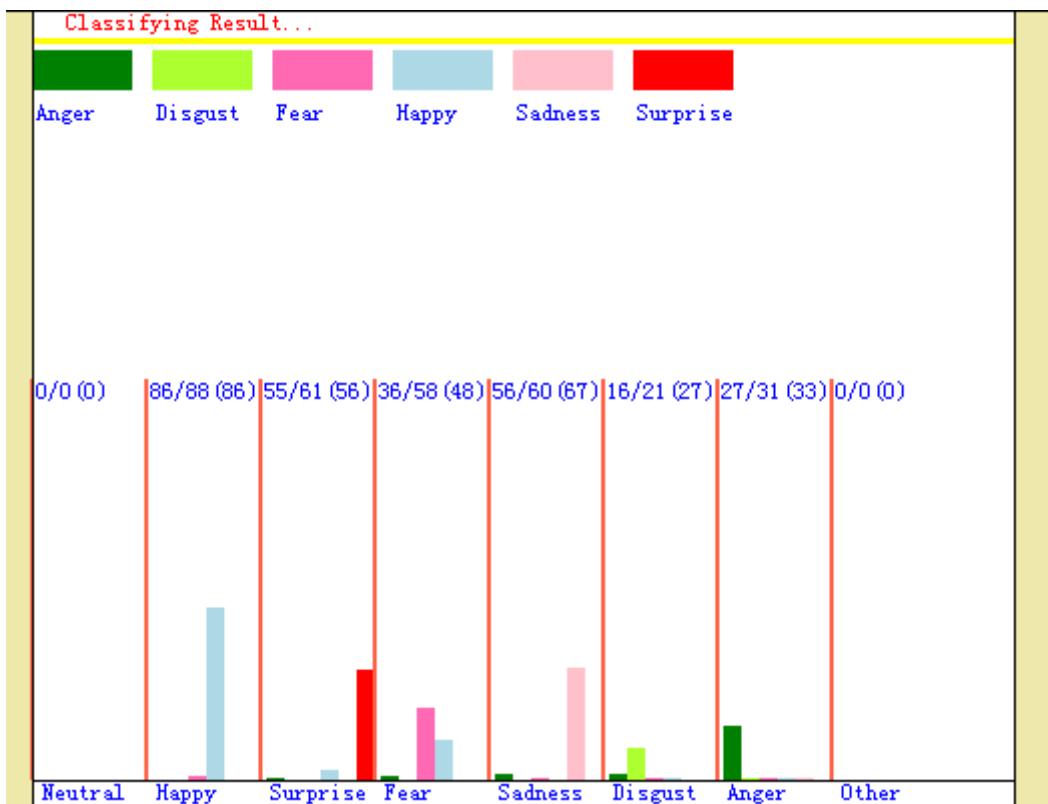


图 4.25：样本强度为 0.5 时的分类结果统计

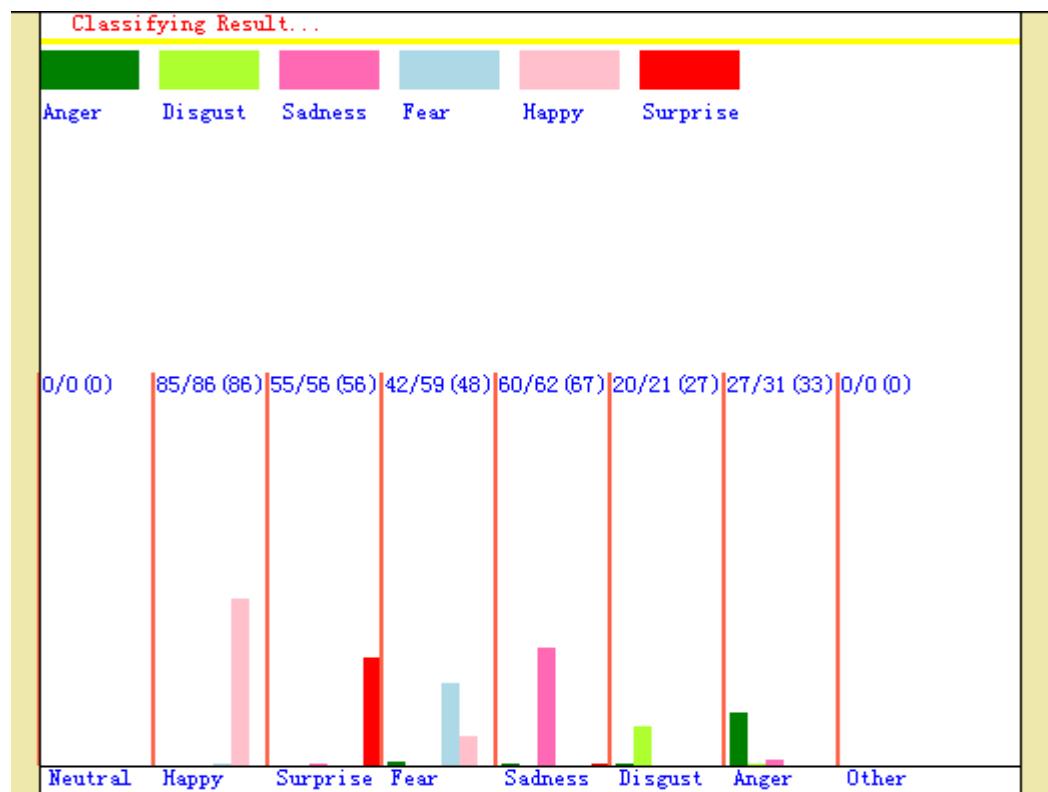


图 4.26：样本强度为 0.6 时的分类结果统计

LBP 特征的静态分析与动态分析研究

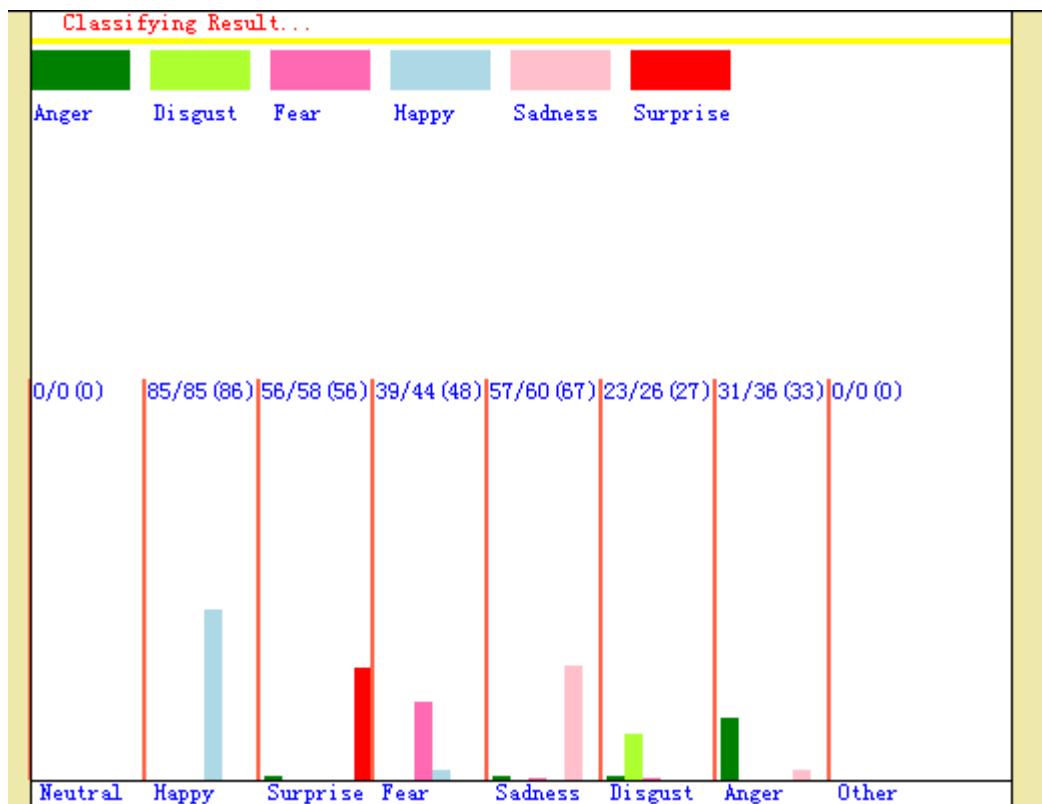


图 4.27：样本强度为 0.7 时的分类结果统计

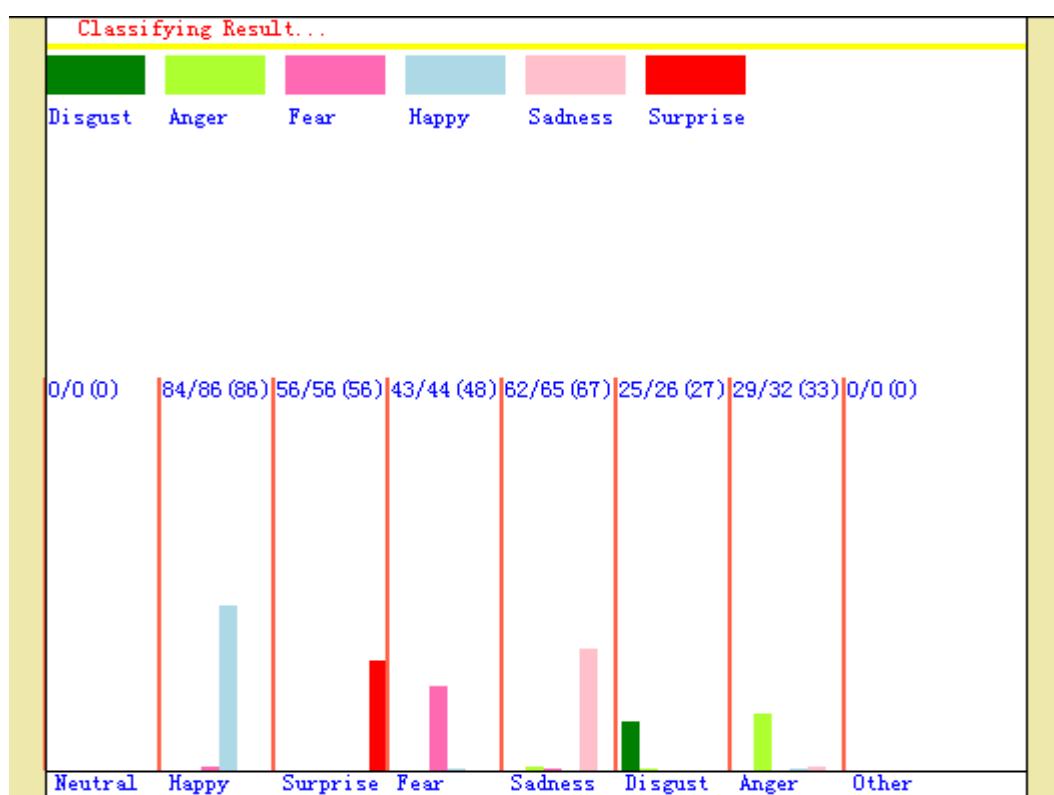


图 4.28：样本强度为 0.8 时的分类结果统计

LBP 特征的静态分析与动态分析研究

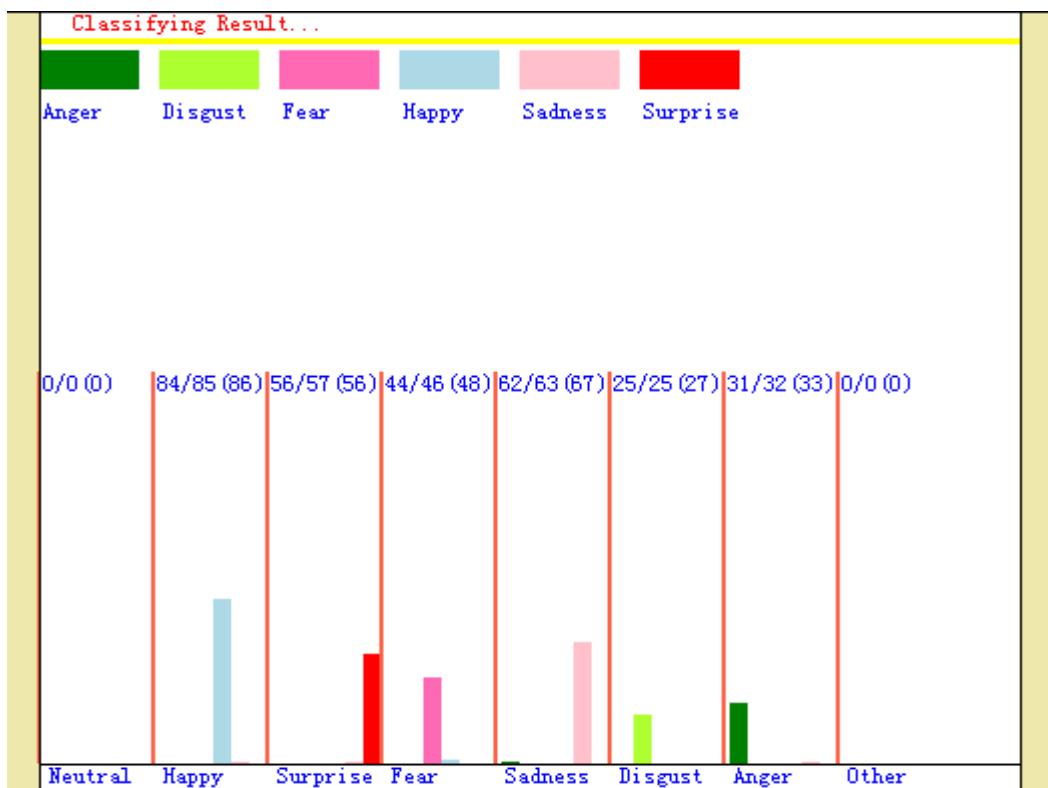


图 4.29：样本强度为 0.9 时的分类结果统计

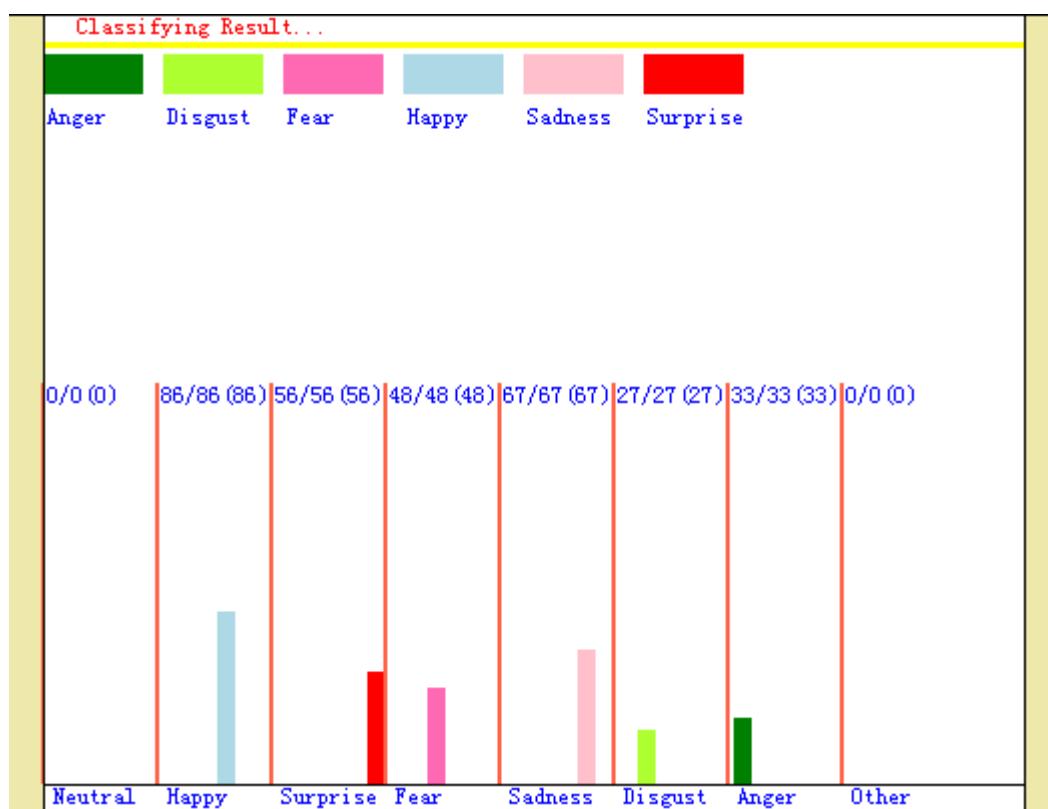


图 4.30：样本强度为 1.0 时的分类结果统计

4.7.2 动态分析对样本强度的敏感性研究

为了比较 VLBP 在不同强度下的分辨能力，本文同样做了实验验证，使用的是 $VLBP_{(3,3,3,8,8,8)}$ 模型，测试的样本强度同样有 0.5, 0.6, 0.7, 0.8, 0.9 和 1.0。利用进行了 AdaBoost 选择出 1000 维显著特征，但仅就 Fear 这个表情进行实验。实验结果如表 4.10 和表 4.11 所示。

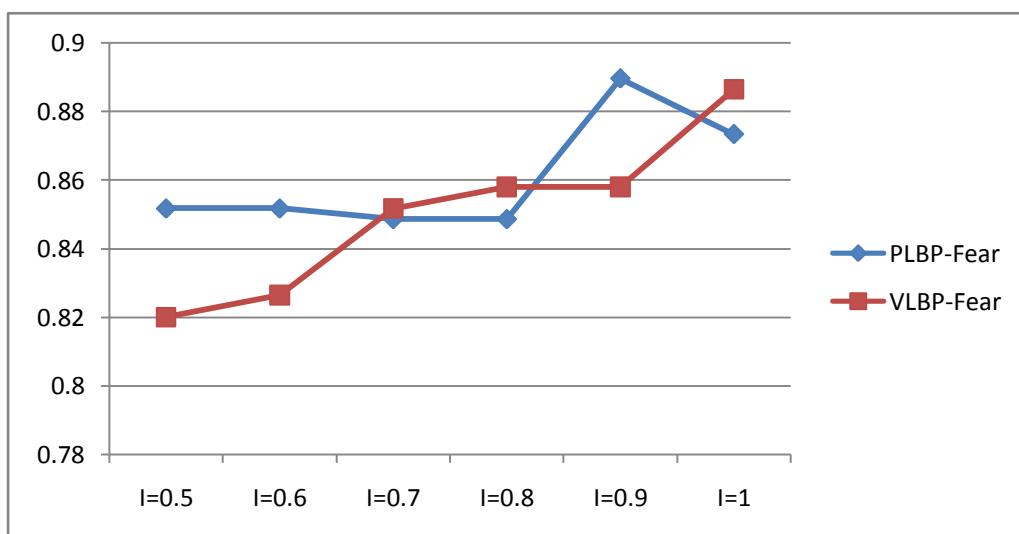
表 4.10: VLBP 不同样本强度下的交叉验证精度

VLBP(9 分块)	0.5	0.6	0.7	0.8	0.9	1.0
Fear	0.8201	0.8265	0.8517	0.8580	0.8580	0.8864

表 4.11: VLBP 不同样本强度下的训练精度

VLBP(9 分块)	0.5	0.6	0.7	0.8	0.9	1.0
Fear	317	317	317	317	317	317

从表 4.10 可以看到，VLBP 的交叉精度也是随样本强度地提高而提高，然而结果并没有总在所有强度下比 PLBP 的高（图表 4.12）。从图表 4.12 还可以看到，用 AdaBoost 选择特征的 VLBP 的交叉验证精度是随样本强度呈比较平滑的上升曲线，而 PLBP 的则比较波动，但两者总体都随样本强度提高。在样本强度为 0.6 和 0.7 时，动态分析的模型 VLBP 要优于静态的 PLBP。在这个区间中，静态的表情强度人们分辨会比较困难，但如果有动态过程就会容易一些。从表 4.11 也可以看到，在 6 总不同强度下，动态的 VLBP 的训练精度都是 100%，即能完全把样本分开，而在静态分析中，即使使用了所有特征，也无法完全分开（见表 4.9）。尤其是在较低的样本强度下，动态分析的效果优于静态分析就更明显。



图表 4.12: 交叉验证精度随着样本强度地变化情况

4.7.3 本节小结

本节对样本强度对训练效果的影响进行了初步的研究，可以发现动态分析在较低的样本强度下会优于静态分析方法，但尚不能从实验上给出很肯定的结论，需要进一步地实验分析，这些将作为以后的研究工作。由于动态分析模型还有很多种建模的方式，以及在特征选择后不同的特征长度都训练效果的影响都可能是值得关注的，因此，深入研究这些差异可能对分析效果的影响是必要的。

4.8 本章小结

本章着重研究了在表情分析过程中涉及到的相关模型以及一些模式分类过程中所用到的实验方法，同时还根据 LBP 特征的特性和针对这些特性提出的相关算法。在本章中，实验分析是最为主要的部分，有大量的实验数据比较和分析，并定量或定性地对一些研究过程中研究体验进行分析。本章提出了基于零率的降维方法，这个方法主要是针对 LBP 特征的，当然，它的数学思想允许它扩展其它可能的应用范围。本章还简单地介绍全新的区域分块模型 ILBP，分析了这个模型提出的理论基础，并给从实验上验证也它的潜在优势。本章最后研究了不同样本强度下对动态分析与静态分析的分辨能力差别和它们随样本强度的变化情况进行了实验，实验初步验证动态分析在较低样本强度下动态分析的优势。

第五章 系统的工程实现

在进行表情的分析研究过程中，为了能够对分析的数据进行处理，验证及可视化，我们开发了一系列软件系统来完成这些任务。此外，由于研究过程中要不断地变更模型和相关算法，不同的模型之前又会不同的数据要求等，这要求系统在设计与开发过程充分考虑到工程上的问题。为了测定实验模型在现实测试的效果，需要解决系统的运行效率、接口、UI 和稳定性的技术问题。其中接口设计不仅涉及到工程问题，同时也涉及到对不同的模型的理论理解和实践经验。本章将就研究过程中遇到的问题和需要工具，系统进行讨论。

5.1 开发工具和演示系统

研究表情分析模型最终的目标就是要能实现自动的，实时的表情分析。因此系统实现就不能使用如 Matlab, Java 等执行效率较低的编程进行编写。我们在研究过程中在 Visual Studio 2008 开发平台进行系统开发和测试，在编程语言上，我们选用了 C# 做为系统 UI 的开发设计，在核心算法上（如 LBP 算法的实现，眼睛定位和人脸的标准话）采用 C/C++ 进行开发，个别算法（如旋转不变性）则采用了汇编语言进行编程。由于 C# 是托管语言，而 C/C++ 及汇编语言都是 Native(非安全的)语言，为了连接不同结构的语言，我们利用了 C++.net 的优势，将非安全代码与托管代码很好地结合在一起。这样做好处是可以充分发挥非安全代码的高效性能，又可以快速地开发出良好工程结构的 UI 组件，在保证系统高效运行的情况下，实现了良好的工程开发。

系统开发最上层的是演示和分析系统（如图 5.1 所示），这个系统定义了工作模块接口，依赖于低层的功能模型，由功能模块来完成任务的执行，而演示系统完成数据的可视化（见图 5.2）。可视化是用来帮助实验过程中判断实验结果是否正确，数据处理是合正确和监视系统运行状态。前面的章节中使用实验图像都演示通过相关算法自动生成的。因此，数据的可视化对于实验研究是非常重要和必要的，如果没有这样的可视化工具，我们很难判断实验结果的好坏。如之前实验所看到的，仅仅靠训练精度是不能够说时分块方式的优劣的，还要从分类的统计结果来看最后是

系统的工程实现

否存在错分，漏分的情况，并且还要结果交叉验证精度来评估。

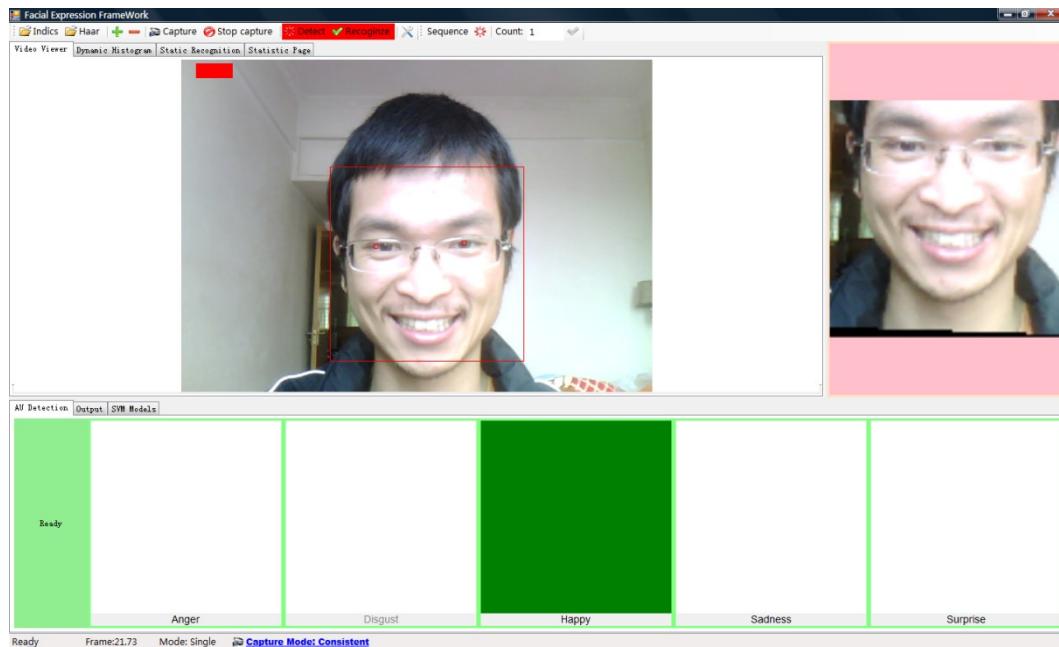


图 5.1：演示分析系统的主界面

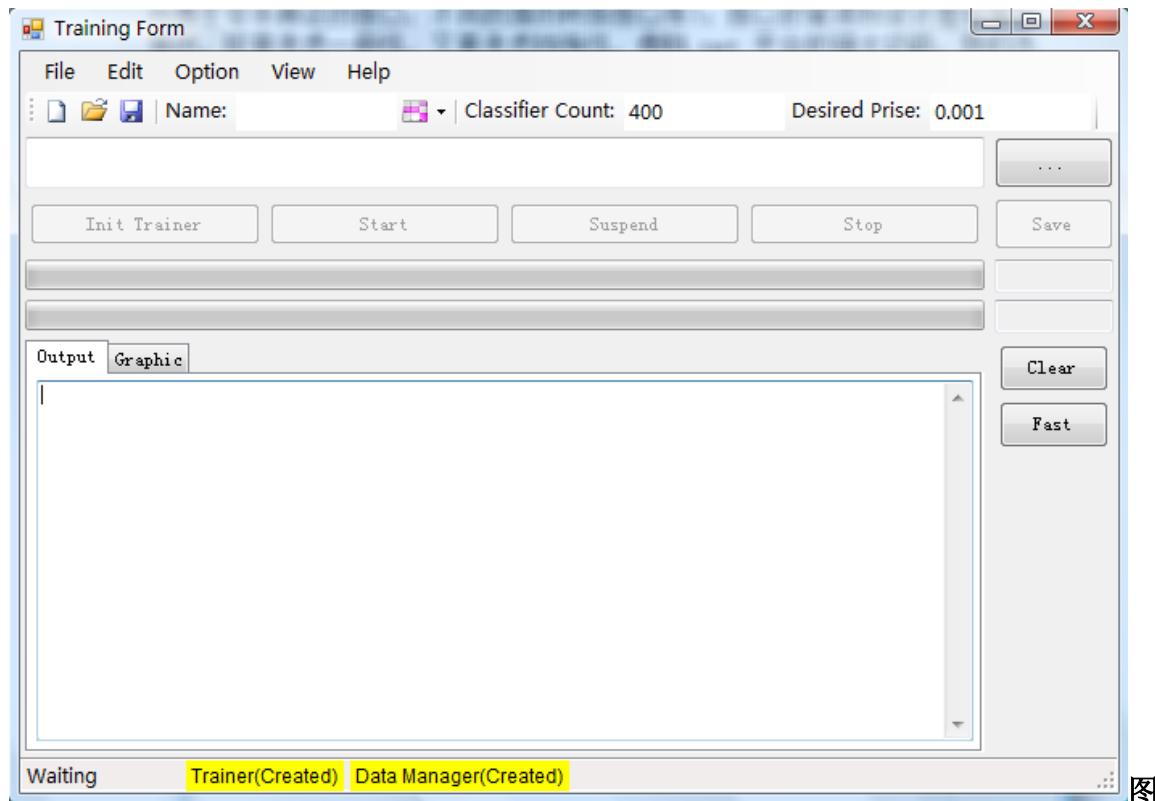


图 5.2：演示分析系统的的依赖关系

从图 5.2 的依赖关系图可以看出，演示系统需要实现接口管理，这是由于为了评估不同模型 (PLBP, VLPB, ILBP)，不同特征选择算法 (AdaBoost 和基于零率的降维方法，或两者结合等)，在不同的模型，不同算法之间实现准确连接还需要相关的适配器 (如用来连接 OpenCV 类库与 .Net 的接口，连接 AdaBoost 与基于零率算法的接口，不同数据的转接接口等)。接口的管理和设计是非常困难的，即要考虑一

系统的工程实现

般性，又要考虑特殊性，借助.net 平台的强大功能，我们不仅在演示系统上实现了优良的接口设计，也在其它工具上完成了类似的设计。如我们开发的训练器系统，就是一个可以动态加载不同训练器的系统（见图 5.3），它有非常简洁的界面，这个工具可以根据定义的接口自动判断所加载的训练器是否实现速算法，是否可暂停，何时可以保存和是否支持并行运算等，然后根据判断结果自动变更用启界面上的元素，自动更改训练类别，自动选择文件类型和更新绘图方式等。所有的这些设计都大大地加快了研究速度，同时也为以后的工具做好了充分准备。



图

5.3 训练器运行窗口

5.2 主要研究工具的开发设计

研究的进行首先是要对现有的数据进行处理，将不同的数据归整到所需要的状态。在前面的章节已经介绍了一般的分析过程（见图 1.1）。在研究过程中我们做了很多工作，其中是一个很重要的研究系统软件就是集成了数据生成模块、分析统计工具、基于零率的降维工具和 SVM 分类器 LBPStudio 程序（如图 5.4 所示）。

从图 5.4 可以看到这个软件主窗口下还包含了三个子窗口（图中显示的是 SVM 训练器），它们都是可以单独成为一个工具，只是在开发过程中用一个母窗口把它们包含进来，方便使用。第一个窗口为特征生成器（见图 5.5），它可以生成用来训练和分析的特征文件，特征文件的样本强度都可以在它上面设置，还有可以手工选择一些样本来做特征观察（如图 5.6 所示）。

系统的工程实现

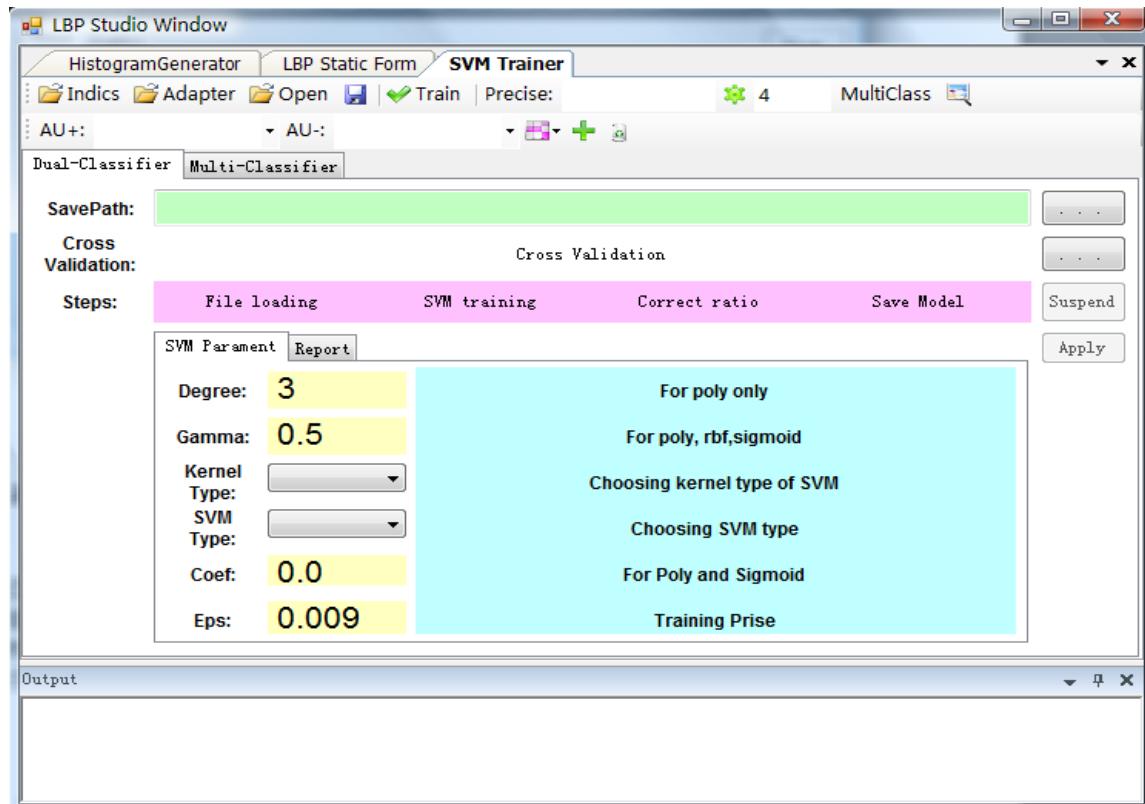


图 5.4: LBPStudio 工具软件系统

LBPStudio 内部的数据生成, 分析和训练都与 FACS 这个表情编体系码紧密相连, 虽然我们在前面的章节并没有对 FACS 做介绍, 但它在我们开发过程中都占据了重要的部分, 从图 5.4 可以看到有 AU+ 和 AU- 的下拉列表, 这些都是我们在研究基本表情时需要指定的。如我们要对 Smile 这个表情进行分析, 就需要知道它是 AU 组合是什么, 即包含哪些 AU, 不能包含哪些 AU。6 个基本表情的 AU 组合在系统内总属于内置的, 可以直接通过 AUPairManager 的一个类来获取相关信息。这样的类设置可以很方便地在整个工程中通过定义的公用接口引用。

图 5.4 首先展示的是 SVM 训练器, 它的界面简洁美观, 而且参数的设定也非常方便。只要通过鼠标操作就可以轻松完成所有的功能, 它底层调用了 C++ 算法, 上层用 C# 完成 UI 设置, 中间层用 C++.Net 连接。在训练过程中, StepIndicator 控件组一个接一个地闪烁以提示当前系统正在进行的操作, 有些操作是可以挂起 (Suspend) 和继续 (Resume) 的, Suspend 按钮会自动判断是否允许用户挂起一个操作, 训练精度可以在 Correct Ratio 的 StepIndicator 上显示。模型地交叉验证精度也是在 SVM 训练器里面完成的, 可以指定需要多少 Folder 的交叉验证, 在验证时, cross-validation 的 StepIndicator 会显示验证结果和提示验证过程状态。一般情况下 SVM 只是一个两类分类器, 但也可以通过多个 SVM 模型来实现多类分类, 系统也集成了这样的功能选择。

系统的工程实现

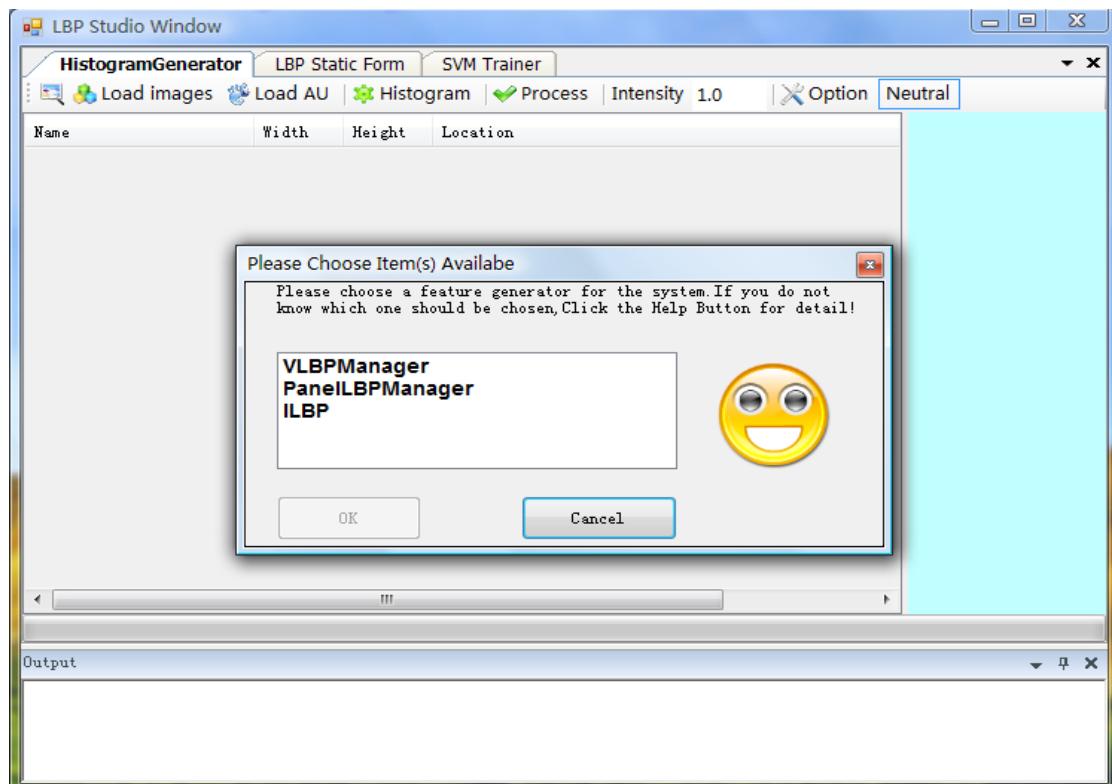


图 5.5：弹出模型选择窗口的特征生成器

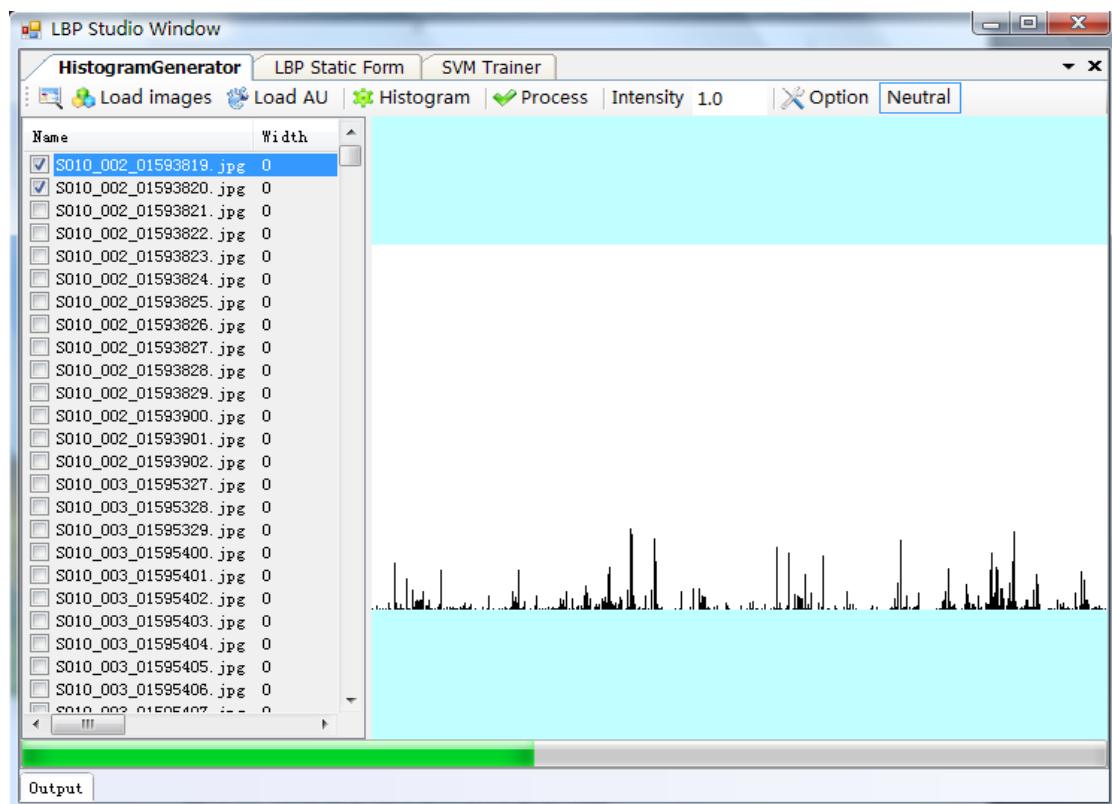
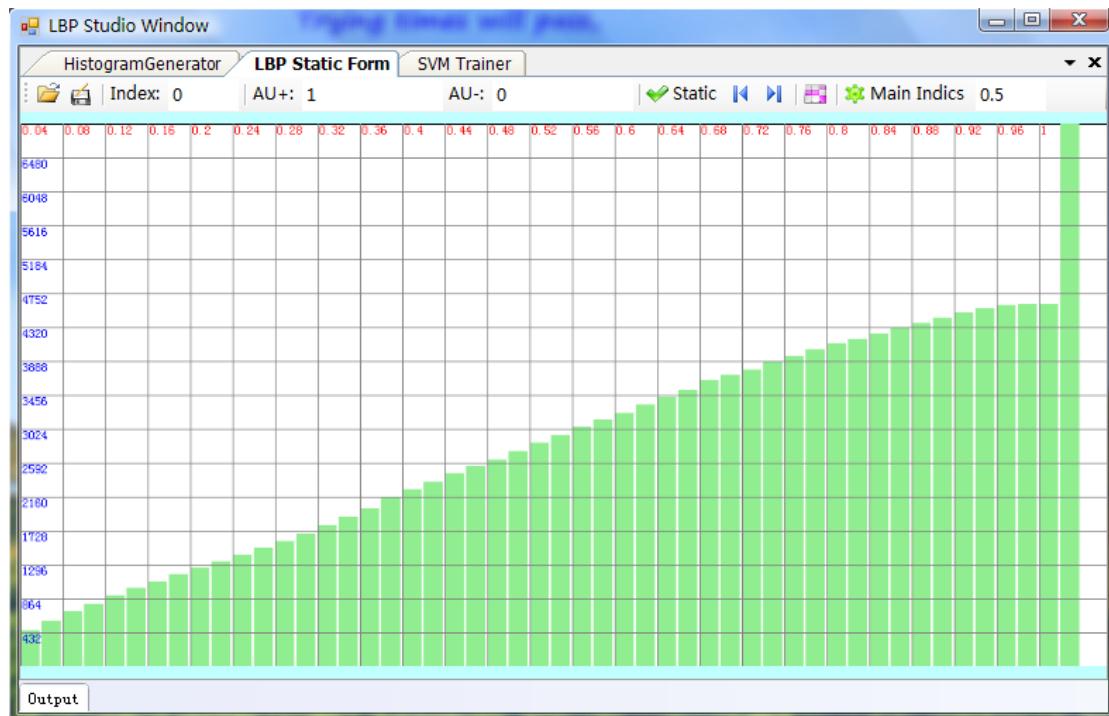


图 5.6：弹出模型选择窗口的特征生成器

系统的工程实现

LBPStudio 的第二个窗口是样本分析器（见图 5.7 和图 5.8），它是用来分析由样本生成器输入的特征文件，可以查看有效特征 VS 零率的统计图（如图 5.7 所示），也可以查看指定的 AU 组合在生成样本中的分布情况（如图 5.8）。



在样本分析器中还可以完成特征文件基于零率的降维，以方便 AdaBoost 在训练减少训练难度和提高训练精度。在实验研究过程中，本文主要是在 LBPStudio 完成基于 LBP 特征的表情分析研究，然后在演示系统中研究实时效果，这两个系统构成了研究的实验平台。

由于本文提出了 ILBP 模型，这个模型虽然有很好的实验潜力，但也由于其不规则的区域分块在实验时不好处理。为此，我们开发了 ILBP Tool 这个软件用来特征 ILBP 的区域蒙版文件，借助 ILBP 内核就可以将这个模型很好去应用到静态分析或动态分析中去。ILBP Tool 的窗口如图 5.9 所示：

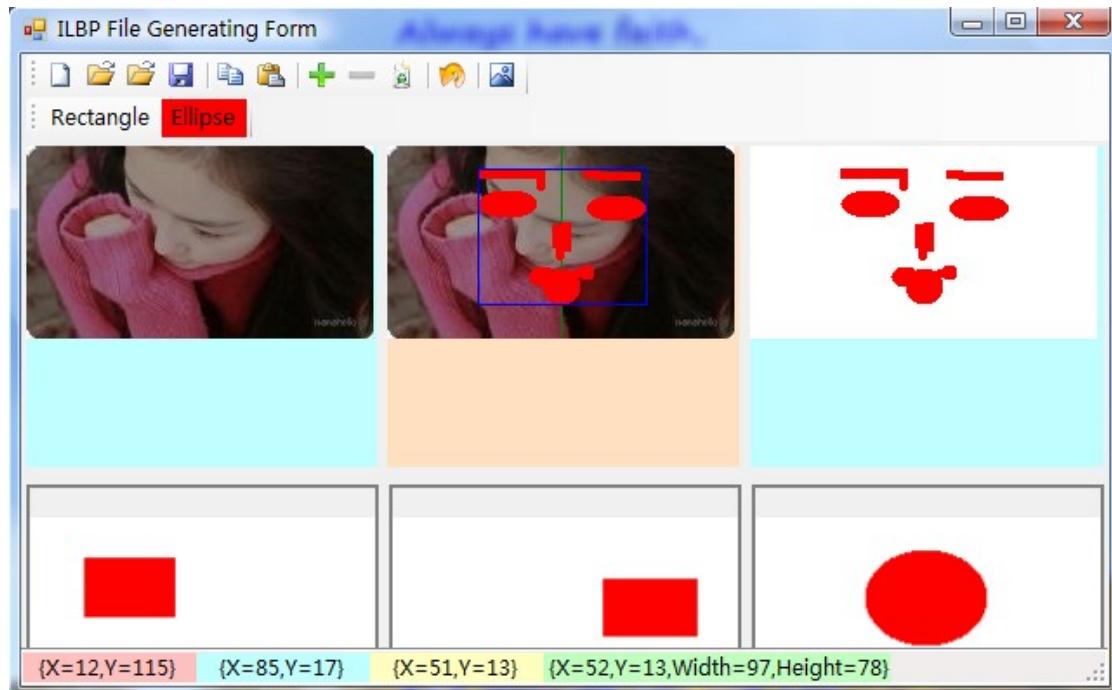


图 5.8：弹出模型选择窗口的特征生成器

ILBP Tool 的工作界面也非常简洁，在工作区域内有三个大小一样的子区域，分别来显示参考的样本图片，操作过程的效果设计图片和最终生成的蒙版图片。底下显示已经完成的不规则分块，用户可以使用绘图工具来绘制任意形状的区域分块。ILBP Tool 是典型的利用组件开发完成的软件，由于之前的工具的模块接口定义比较好，在组件开发时都充分考虑了扩展性，使用在开发 ILBP Tool 时只要上百行的代码就可以完成编码工作。此外，ILBP Tool 还支持从图片来生成目标文件，用户可以 Photoshop 等比较专业的图像编辑软件中绘制他们想要的蒙版图片，然后利用工具的导入功能自动生成蒙版，这也充分利用了图像处理软件的强大功能，所有这些功能为以后深入研究 ILBP 模型提供了必要的基础。

5.3 实时识别效果和问题

在演示系统开发完毕后，我们立即进行了实时效果实验，通过摄像头动态捕获图像，然后根据加载的模型生成图像序列（一张图片或多张），把图像序列输入到

LBP 模型进行特征生成（见图 5.9），经过特征选择输入到 SVM 进行预测，决定最终表情类别。

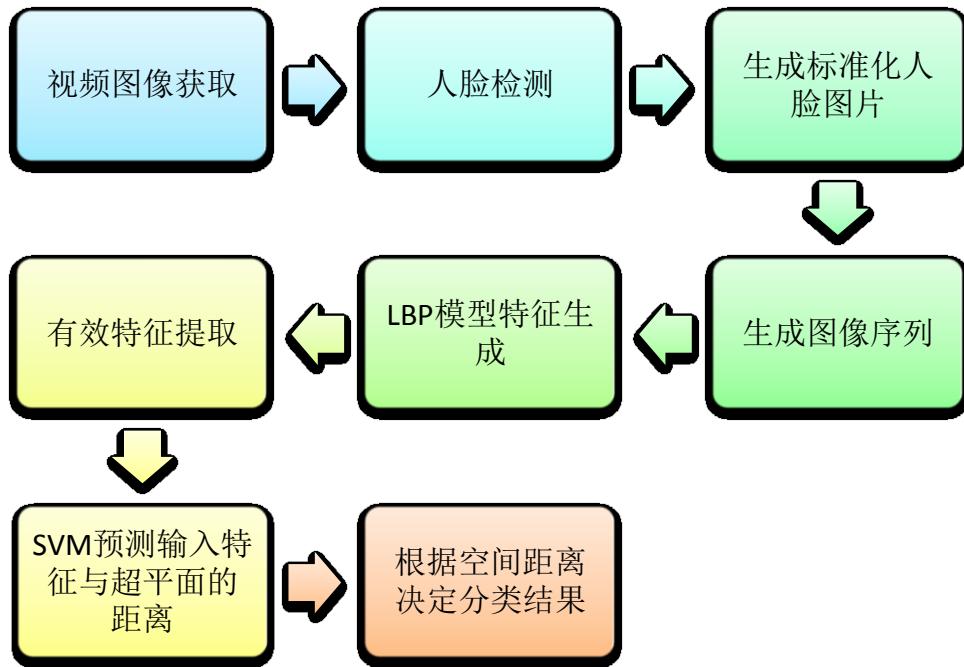


图 5.9：实时识别流程图

我们在对 PLBP 和 VLBP 进行实时效果实验发现，由于 VLBP 需要一组图片序列才能开始识别，由于摄像头一般一秒有 30 帧，但经过图 5.9 的过程后一般会降到 15 帧左右，随着 SVM 模型的向量的增加而会不断减少。当帧率小于 10 后，系统在双核的系统上运算很缓慢，以致演示表情很难把握，严重影响效果，再加序列较长的情况下，采集的标准图片容易在采集过程中受到光线和摄像头清晰度差的影响无法很准确地定位到眼睛，以致得到的标准图片不标准。

VLBP 的实时效果与预期的差别比较大，影响的因素很多，如样本都是国外人的相片，一些表情中国人很难表现出来，如 Sadness，样本中的这类表情都是嘴形向下弯的，实验时本人做了半天也没有一个做出能被确定是 Sadness 的。此外，由于硬件条件有限，实验使用的摄像头清晰度差，无法分辨鼻子中间的纹理变化。目前所进行的实时实验取得最好的效果的是 PLBP，一个重要的原因是它只要求一张图片输入，在采用样本时受到环境的干扰大大减小，即使有时眼睛定位没定准，也不影响下一帧的识别，图 10，11 和图 12 给出了用 PLBP 模型的实时效果。

此外，实时识别也与之前的实验数据很好的一致，不论在 VLBP 还 PLBP 模型下，Surprise 这个表情的识别精度都是最好的，只要睁大眼睛和张大嘴巴就会立即被认为是 Surprise；此外 Happy 这表情也很高，只是嘴形不要太夸张，就能很好识别；接下去是 Disgust，但 Disgust 还会时常被同时识别成 Anger，原因是这两个表情我们肉眼也不好分辨；而 Fear 有时会被认为是 Surprise，可能的原因惊讶时我们也确定有奇怪表情的特征吧；Sadness 到目前为止没有被识别过。

系统的工程实现

从实时实验中我们发现表情分析还有很长的路要走，相对人脸识别，表情识别更加复杂，需要把握的因素也更多。但在实验的过程中，我们也慢慢地认清这些问题的所在，并在努力地寻找解决方法。

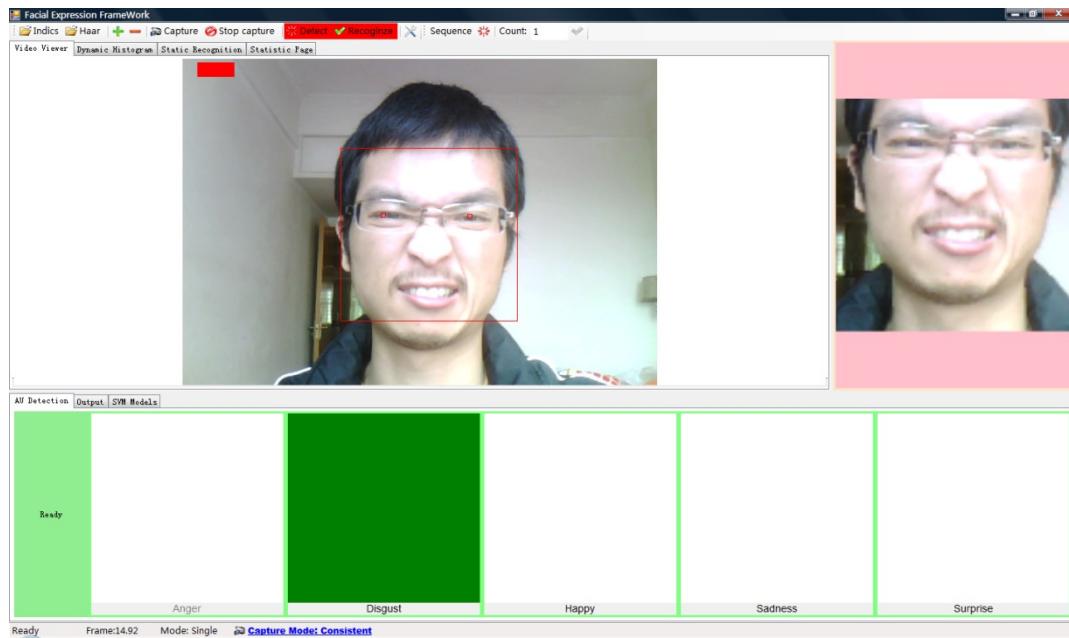


图 5.9：识别结果为 Disgust

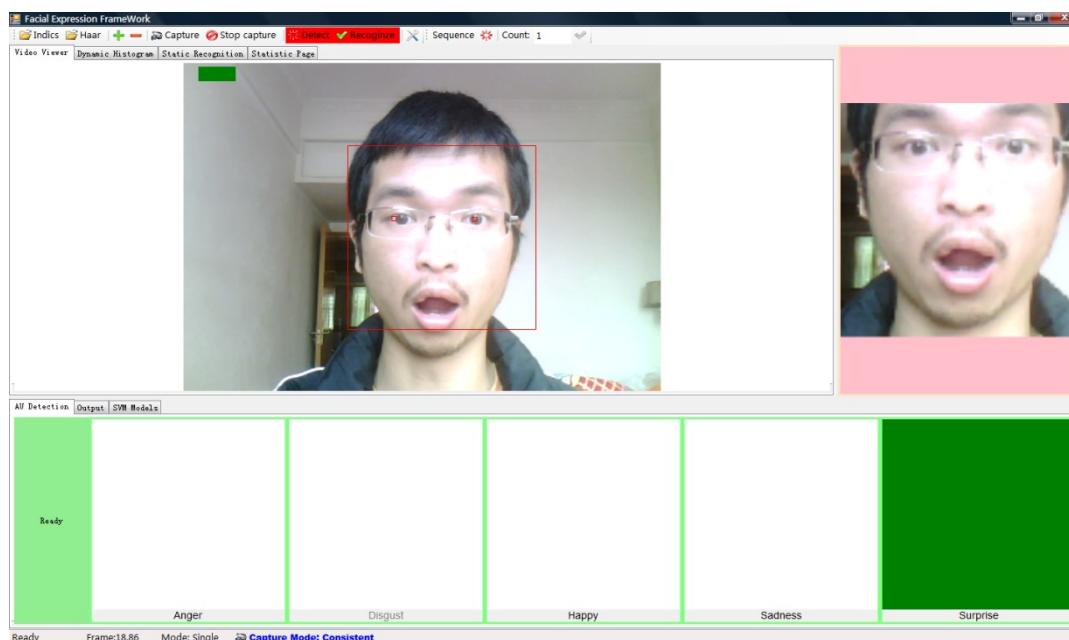


图 5.9：识别结果为 Surprise

系统的工程实现

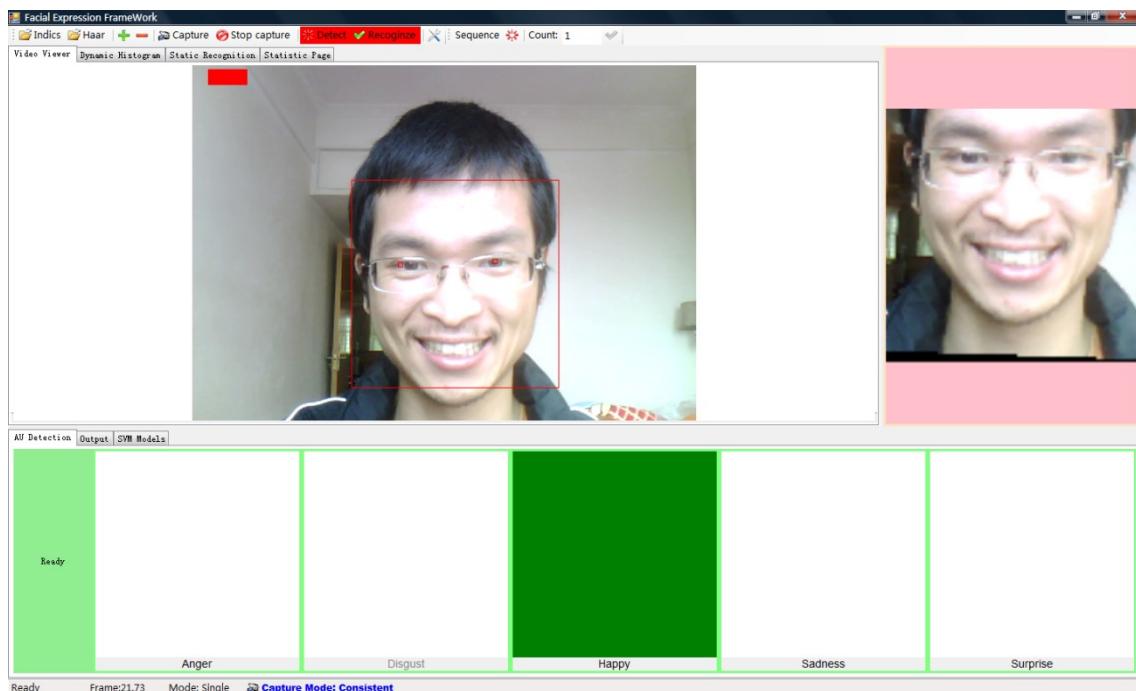


图 5.9：识别结果为 Happy

5.4 本章小结

本章介绍在研究过程中我们开发的演示系统和相关实验工具，简单阐述了系统的工程设计和实现，并通过给出实时条件下的表情分析结果，介绍识别过程的问题和可能原因。

第六章 总结与展望

本文着重研究了在用 LBP 模型进行表情分析时涉及到的各种问题，经过大量的实验分析和系统开发的不断完善，实验平台日趋成熟，对表情分析的问题认识也更加清晰。

6.1 工作总结

本文主要总结基于 LBP 特征的表情分析研究，从静态分析与动态分析来着重阐述工作成果。研究了分块策略对分析效果的影响，提出了新了更加通用的分块模型 ILBP，分析了模型的理论依据，从实验上分析其潜在的优点;根据 LBP 表示的特征特点，提出了基于零率的降维方法，给出了数学理论基础，并通过实验对其有效性加以验证。本文还就不同的样本强度对分析效果的影响进行了初步的研究，比较了动态分析模型与静态模型在低样本强度下的优劣。最后给出了实时分析的效果，分析了实际应用中遇到的问题，和可能的解决方式。

6.2 工作展望

在本文的前面章节中，我们发系统地研究 LBP 特征表示在表情分析中的各种问题，在未来的工作中我们将就下面几个方面深入研究：

- (1) 更加深入研究动态分析技术的特征，研究更多的动态分析模型，将 ILBP 扩展到动态分析中进行研究。
- (2) 深入研究基于不同样本强度的动态分析效果，通过 AdaBoost 选出显著特征，充分发挥动态分析的优劣。
- (3) 研究特征表示的分析效果（如 Gabor^[42-43]小波特征[]，Harr 特征^[44-46]等），寻找更适合表情分析的特征表示。
- (4) 开发更加实用可靠的演示系统，在工程方面不优化算法执行效率。

参考文献

- [1] Beat Fasel and Juergen Luettin. Automatic Facial Expression Analysis: A Survey[C]. IDIAP& Ascom Systec AG, Switzerland. Pattern Recognition,36(1),pp.259-275,2003.
- [2] Marian S.Bartlett, Bjorn Braathen, Gwen Littlewort-Ford, John Hershey, Ian Fasel2, Tim Marks, Evan Smith, Therrence J. Sejnowski, & Javier R. Movellan. Automatic Analysis of Spontaneous Facial Behavior: A Final Project Report[R]. UCSD MPlab. 2009.
- [3]Bartlett, M., Donato, G., Movellan, J., Hager, J., Ekman, P., & Sejnowski, T.(2000). Image Representations For Facial Expression Coding[M]. In S. Solla, T.Leen, & K.-R. Muller(Eds), Advances in Neural Information Processing System, Vol. 12. MIT Press.
- [4]Bartlett, M., Movellan, J.,& Sejnowski,T.(2001). Image Repressions for Facial Expression Recognition[C]. IEEE Transactions on Neural Network.
- [5]Caifeng Shan, Shaogang Gong,Peter W.McOwan. Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study [G]. Philips Research, High Tech Campus & Department of Computer Science, University of London, UK. Image and Vision Computing, pp.803-816, 2008.
- [6]Douglas W.Cunningham, Christian Wallraven. Dynamic Information for The Recognition of Conversational Expressions [J]. Max Plack Institute for Biological Cybernetics, Tubingen, Germany & Brandenburg Technical University, Cottbus, Germany & Department of Brain and Cognitive Engineering Korea University Seoul Korea. Journal of Vision (2009)9(13):7, pp.1-17, 2009.
- [7]Marian Stewart Bartlett. Face Image Analysis by Unsupervised Learning and Redundancy Reduction[D]. University of California, San Diego. Paper for the degree Doctor of Philosophy.
- [8]Marian Stewart Bartlett, Gwen C.Littlewort, Mark G.Frank, Claudia Lainscsek, Lan R.Fasel, Javier R.Movellan. Automatic Recognition of Facial Actions in Spontaneous Expressions [J]. Institute for neural Computation, University of California, San Diego. Journal of Multimedia, Vol.1, No.6, September 2006.
- [9] T. Ojala, M. Pietikainen, D. Harwood, A Comparative Study of Texture Measures with

参考文献

- Classification Based on Featured Distribution [G]. Pattern Recognition 29 (1) (1996) ,pp.51–59.
- [10]Pearson,k. On Lines and Planes of Closest Fit to Systems of Points in Space[M]. Philosophical Magazine 2(6), pp.559-572, 1901.
- [11] Xiao yang Tan and Bill Triggs. Fusing Gabor and LBP Feature Sets for kernel-Based Face Recognition[C]. INRIA and Laboratories Jean Kuntzmann, France.
- [12]Quan you Zhao, Bao chang Pan, JianJia Pan, Yuan Yan Tang. Facial Expression Recognition Based On Fusion of Gabor and LBP Features[C]. Department of Computer and Information Science, Hunan University of Science and Engineering. Wavelet Analysis and Pattern Recognition,2008.ICWAPR'08.International Conference, Vol 1,pp.362-367 2008.
- [13] Fablet,R., Boutheemy,P. Motion Recognition Using Spatio-Temporal Random Walks in Sequence of 2D Motion-Related Measurement. IEEE Int. Conf. On Image Procesing,(ICIP 2001). Thessalonique, Creece(2001) pp.652-655.
- [14]Rahman, A., Murshed, M. Real-Time Temporal Texture Characterization Using Black-Based Motion Co-Occurrence Statistics. IEEE International Conference On Image Processing(2004), pp. 1593-1596.
- [15] D.Chetverikov, R.Peteri. A Brief Survey of Dynamic Texture Description and Recognition. Proc. Int'l Conf. Computer Recognition System, pp.17-26, 255.
- [16]C. Shan, S. Gong and P.W. McOwan. Robust Facial Expression Recognition Using Local Binary Patterns. Proc. IEEE Int'l Conf. Image Processsing,pp.370-373,2005.
- [17] S.P. Aleksic. K.A. Katsaggelos. Automatic Facial Expression Recognition Using Facial Animation Parameters and Mutil-Stream HMMS. IEEE Trans. Signal Processing, Supplement on Secure Media 2005.
- [18]Guoying Zhao, Matti Pietikaine.Dynamic Texture Recognition Using Volume Local Binary Pattern [EB]. Machine Vision Group, InfoTech Oulu and Department of Electrical and Information Engineering, University of Oulu, Finland. IEEE Transactions on Pattern Analysis And machine Intelligence, June 2007(vol.29 no.6), pp.915-928.
- [19] Guoying Zhao, Matti Pietikaine. Boosted Multi-resolution Spatiotemporal Descriptors for Facial Expression Recognition [J]. Machine Vision Group, InfoTech Oulu and Department of Electrical and Information Engineering, University of Oulu, Finland. Pattern Recognition Letters

参考文献

- 30(2009), pp.1117-1127.
- [20]Yoav Freund, Robert E.Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting [C]. AT&T Labs Florham Park, NJ USA. Lecture Notes in Computer Science; Vol.904, Proceedings of the Second European Conference on Computational Learning Theory, pp.23-37, 1995.
- [21]H. Abdi. Discriminant Correnspondence Analysis[M]. In: “N.J. Salkind(Ed.): Encyclopedia of Measurement and Statistics”. Thousand Oaks(CA): Sage,pp. 270-275.
- [22] Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems[R]. Annals of Eugenics 7: pp. 179-188.
- [23]Mika, S. et al. Fisher Discriminat Analysis with Kernels[C]. IEEE Conference on Neural Networks for Signal Processing IX: pp.441-48, 1999.
- [24]Lawrence R. Rabiner. A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition[C]. Proceedings of the IEEE 77(2): pp. 257-286, 1989.
- [25]Li J, Najmi A, Gray RM. Image Classification by a Two Dimensional Hidden markov Model[C]. IEEE Transactions on Signal Processing 48(2): pp.517-533, 2000.
- [26] Davis RIA, Lovell BC. Comparing and Evaluating HMM Ensemble Training Algorithms Using Train and Test and Condition Number Criteria [J]. Journal of Pattern Analysis and Application0(0),pp.1-7,2000.
- [27]M.Aizerman, E.Braverman, and L.Rozonoer. Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. Automation and Remote Control 25, pp. 821-837.
- [28] B.E. Boser, I.M. Guyon, and V.N.Vapnik. A Training Algorithm for Optimal Margin Classifiers[C]. D.Haussler, Editor, 5th Annual ACM Workshop on COLT, pp. 144-152, Pittsburgh, PA, 1992. ACM Press.
- [29]C.W.Hsu, C.C. Chang, and C.J.Lin. A Practical Guide to Support Vector Classification[M]. Tech Rep. Department of Computer Science, National Taiwan University, 2005.
- [30] Paul Viola, Michael Jones.ERapid Object Detection using a Boosted Cascade of Simple Features[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01) - Volume 1, pp.511-2001,2001

参考文献

- [31]Chunde Huang,KunhuiLin,Fei Long. A Fast Eye Localization Algorithm Using Integral Image[C]. Software School of Xiamen University. ISCID 2009,pp.231-234,2009.
- [32] Peng Yang, Bo Du, Shiguang Shan, Wen Gao.A Novel Pupil Localization Method Based On Gaboreye Model And Radial Symmetry Operator[C]. Proc. of ICIP ,pp.67-70,2004.
- [33] John Daugman. Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression. IEEE Transaction on Acoustics, Speech, and Signal Processing. Vol.36.7. July 1988,pp 1169-1179.
- [34] Ambadar, Z., Schooler, J. W., & Cohn, J. Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. Psychological Science, 16, pp.403–410.
- [35] Anaki, D., Boyd, J., & Moscovitch, M. Temporal integration in face perception: Evidence of configural processing of temporally separated face parts. Journal of Experimental Psychology: Human Perception and Performance, 22, pp.1–19.
- [36]Ojala, T., PietiKainen,M.,Maenpaa,T. Multiresolution Gray-Scale And Rotation Invariant Texture Classification With Local Binary Patterns. IEEE Transactions On Pattern Analysis And Machine intelligence 24(2002) ,pp.917-987.
- [37]Ojala,T.,PietiKainen,M., Harwood,D. A Comparative Study of Texture Measures With Classification Based On Feature Distributions. Pattern Recognition 29(1996) ,pp.51-59.
- [38]Kohavi, Ron. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection[C]. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2(12), pp. 1137-1143, 1995.
- [39]P.Ekman and W.Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement[M]. Consulting Psychologists Press, Palo Alto, 1978.
- [40]T.F.Cootes, G.J.Edwards, C.J.Taylor. Active Appearance Models[C]. European Conf. Computer Vision, pp. 484-498, 1998
- [41] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active shape models – their training and applications[C]. Computer Vision and Image Understanding, 61(1), pp.38–59, 1995.
- [42]Peng Yang, Shiguang Shan, Wen Gao, Stan Li, Dong Zhang. Face Recognition Using Ada-Boost Gabor Features. Proceeding of The 6th IEEE International Conference on Automatic Face and Gesture Recognition, Korea pp. 356-361, May 2004.

参考文献

- [43] Peng Yang, Shiguang Shan, Wen Gao, Stan Li, Dong Zhang. Face Recognition Using Ada-Boosted Gabor Features[C]. Proceeding of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Korea pp.356-361, May 2004.
- [44] Mita Takeshi, Kaneko Toshimitsu, Hori Osamu. Joint Haar-like Feature Based on Feature Co-Occurrence for Face Detection[J]. IEICE Transactions on Information and System, D Vol. J89-D. NO. 8, pp. 1791-1802,2006.
- [45] Papageorgious, Oren and Poggio. A General Framework for Object Detection. International Conference on Computer Vision, 1998.
- [46] Messom, C.H. and Barczak, A.L.C. Fast and Efficient Rotated Harr-Like Features Using Rotated Integral Image. Australian Conference on Robotics and Automation (ACRA2006). pp. 1-6, 2006.

研究生期间发表的论文

- [1] Chunde Huang,Kunhui Lin,Fei Long. A Fast Eye Localization Algorithm Using Integral Image[C]. Software School of Xiamen University. ISCID 2009,231-234,2009.

致谢

硕士三年的学习生活很快就要结束了，在攻读学位期间，我成长了很多，也在努力地追求着自己的梦想。在这艰难的三年里，我得到许多老师，朋友和同学的大力帮助，在此向所有帮助过我的人们表示我最诚挚的谢意。

首先我要衷心感谢我的导师龙飞老师，他是一们年轻而具有严谨学术精神的学者，他总是能耐心地指导我在科研总遇到的问题。但是我更要感谢他的对思想不成熟做事不够稳重的我的宽容和支持，他在学习和科研上给了我极大的自由，教会我一些为人处事的能力。因为他的宽容和理解，让我参加了长汀支教，可以集中精力去追求自己的梦想。

我还要感谢厦门大学物理机电学院的吴晨旭院长在我追求理想的路上给了我重要的指点；感谢物理系系主任赵鸿教授给我讲授群理论，在理论物理学上给我的指导，以及对我的其它帮助；同时还要感谢我本科的导师段鸿老师对我的支持和帮助；感谢王亚南经济学院的牛霖琳老师的对我慷慨的帮助和关心。此外，我还要感谢中科院院士郭光灿老师给予我的帮助和鼓励。

生活学习上我要感谢好朋友们和同学对我的支持和帮助。首先感谢我的老朋友晓珊在经济和精神上对我的大力支持；感谢王亚南经济学院的同学们给我的友谊，尤其要感谢杨大侠在我无助的时候对我无私的帮助，还要感谢的亚南院的朋友还有王建军，感谢他给我真诚的友谊和支持，非常感谢曹楠，王楠，安然和侯庆峰等对我的关心和帮助以及他们的友情；老友聪哥一直以来的理解，支持以及大学7年的一路相伴相知；感谢时敏学弟在厦大期间的关心与支持；感谢宪斌，绍东，奕堆，保华，明妮和兔子等一直来关心；还有美丽可爱的好朋友林希在支教期间的相伴和照顾；感谢软件学院的黄济泳，李舒平，巍洁和学弟郭宇腾，物理系学弟王叙平等对我的帮助和关心；感谢物理系学长济舸和我分享学习物理学经验和帮助；感谢清华大学的朋友刘冠兰的友谊和帮助；感谢我在长汀的那些可爱的学生给我带来的美好回忆。此外，我还要感谢美丽善良的美子的纯真，谢谢她在她在艰难的时候的关心和鼓励。

我还要感谢我心爱的母校厦门大学，是她保护了我的梦想，让我健康地成长，让我懂得人生大爱和对这个社会的责任，让我明白大学之道，为人之本。感谢她在

致谢

这 7 年来一直的呵护和引导，让我找到了坚定人生理想，让我成为自己认可的人。

最后我要感谢我的家人和亲人们，在我自私地追求自己理想的时候给予我世间真伟大的爱和理解。感谢他们让我成长和成熟，让我变得坚强和勇敢，让我懂得责任。