

基于运动主方向特征的微表情 识别算法研究

(申请清华大学工学硕士学位论文)

培 养 单 位 ： 计 算 机 科 学 与 技 术 系
学 科 ： 计 算 机 科 学 与 技 术
研 究 生 ： 张 金 凯
指 导 教 师 ： 刘 永 进 副 教 授

二〇一五年五月

Research on the Micro-expression Recognition based on a Main Directional Motion Feature

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Master of Science

in

Computer Science and Technology

by

Zhang Jinkai

Thesis Supervisor: Professor Liu Yongjin

May, 2015

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

（保密的论文在解密后遵守此规定）

作者签名： _____

导师签名： _____

日 期： _____

日 期： _____

摘 要

微表情是一种短暂的,不易被人察觉的面部表情。作为一种日常生活中随处可见,却难以捕捉的面部表情,它在谈判、医疗、教育、刑侦甚至国家安全等方面都有着非常广泛的应用前景。但是,由于微表情的持续时间短,运动幅度小,人们难以通过肉眼捕捉并对微表情进行检测和判断。为了解决上述问题,本文提出了一种基于光流域矫正的微表情识别方法,使用这种方法能够有效地解决微表情识别率低下的问题。

大量的已有工作表明,微表情能够反映人们内心世界对外部世界的真实情感,具有丰富的应用价值。在面对面的交流中,有经验的心理学家可以通过捕捉对方的微表情而获知谈话者的内心情感,从而掌握谈话的主动权。然而,正如前文所述,微表情持续时间非常短,即使是有经验的心理学家也难以有效地捕捉和识别微表情,从而使得微表情难以运用到实际生活中去。因此找到一种简单有效的并能够准确检测和识别微表情的方法,对于微表情应用的推广有着至关重要的作用。为了达到这种目的,本文提出了一种基于光流域矫正的微表情识别方法,并在测试数据集上获得了非常好的识别率。

该算法首先使用稠密光流获取微表情序列中图像的运动特征,然后使用光流域矫正的方法对运动特征进行矫正,去除图像序列运动噪声,接着将图像区域根据面部运动单元分成若干感兴趣区域,并提取这些感兴趣区域的运动主方向特征表征微表情的运动特征,最后使用加权平均的方法对微表情特征序列的长度进行归一化。

本文提出的基于光流域矫正的微表情识别方法具有以下优点:1、相比目前世界上最好的方法,本文的方法得到了更高的准确率;2、相比于目前世界上最好的方法,本文的方法采用的特征更少,计算更为效率;3)本文的方法在微表情识别率与数据集的帧率之间没有特别的相关性,在低帧率下也有着相当的识别率。。

关键词: 微表情识别; 运动主方向; 光流域矫正; 加权平均

Abstract

Micro-expressions are brief facial movements characterized by short duration, involuntariness and low intensity. As a kind of facial movements can be seen at anytime and anywhere but recognized hardly by naked eyes, they have potential use in negotiation, medical, education, criminal investigation, national security and some other important fields. However, it is difficult for people to catch and recognize micro-expressions because its short duration and low intensity. In this case, we propose a recognition method based on correction in optical flow domain. The experiments show effective results with our method.

Plenty of researches indicate that micro-expressions would provide us concealed emotions during a talk. Experienced psychologists can get one's suppressed genuine affect by its involuntary micro-expressions, and then control the situation during an important conversation. Hence to find a simple but effective method for micro-expression recognition is of greatest importance. In order to solve this problem, we propose a method based on correction in optical flow domain, which get effective results by experiments.

We first get motion features by dense optical flow algorithm for video clips. Then an affine transformation is applied in optical flow domain for motion feature correction. After the correction, main directional feature for regions of interest are extracted by orientation histogram. Finally weighted mean method is applied for feature normalization.

Bellow are the advantages of our recognition method: 1) the method gets higher recognition rate compared with the state-of-art method; 2) the length fo feature vector is much shorter than other methods, which would be more effective; 3) the method is irrelevant with the fps of datasets.

Key words: recognition for micro-expression; main directional motion feature; feature correction based on optical flow domain; Normalization by weighted mean method

目 录

第 1 章 引言	1
1.1 研究背景与意义	1
1.1.1 基本概念与背景	1
1.1.2 研究意义	2
1.2 国内外研究现状	2
1.2.1 基于心理学实验方法的微表情检测和识别研究	3
1.2.2 基于特征提取的微表情检测和识别方法	5
1.3 本文主要工作	9
第 2 章 基于运动主方向的微表情特征的提取方法	10
2.1 图像预处理	11
2.1.1 局部约束模型与人脸检测	11
2.1.2 基于 FACS 的人脸区域编码	12
2.2 基于主运动方向的 ROI 特征提取方法	13
2.2.1 基于光流的面部运动特征的提取	15
2.2.2 基于方向直方图的 ROI 主方向特征的提取	16
2.3 基于加权平均的微表情运动特征的提取方法	19
2.4 测试评估	20
2.4.1 测试方法	21
2.4.2 测试数据	21
2.4.3 在 CASME 上的测试结果和分析	24
2.4.4 在 CASME2 上的测试结果和分析	26
2.4.5 在 SMIC 上的测试结果和分析	27
2.5 本章小结	29
第 3 章 基于光流域矫正的微表情运动特征的提取方法	31
3.1 仿射变换简介	31
3.2 基于仿射变换的微表情特征的矫正方法	32
3.2.1 基于图像域的微表情运动特征的矫正方法	33

3.2.2 基于光流域的微表情运动特征的矫正方法.....	35
3.3 测试评估	38
3.3.1 测试方法.....	39
3.3.2 测试数据.....	39
3.3.3 在 CASME 上的测试结果和分析	39
3.3.4 在 CASME2 上的测试结果和分析	41
3.3.5 在 SMIC 上的测试结果和分析.....	43
3.4 本章小结	44
第 4 章 实验与结论.....	46
4.1 LBP-TOP 特征简介	46
4.2 在原始帧率的数据集上的对比实验	48
4.2.1 实验方法.....	48
4.2.2 在 CASME 上的对比结果	49
4.2.3 在 CASME2 上的对比结果	50
4.2.4 在 SMIC 上的对比结果.....	52
4.3 在帧率为 25fps 的数据集上的对比实验.....	54
4.3.1 实验方法.....	54
4.3.2 在 CASME 上的对比结果	54
4.3.3 在 CASME2 上的对比结果	55
4.3.4 在 SMIC 上的对比结果.....	56
第 5 章 总结与展望.....	59
参考文献	60
致 谢.....	63
声 明.....	64
个人简历、在学期间发表的学术论文与研究成果	65

第1章 引言

微表情一直以来都是心理学领域研究的热点。与此同时，随着心理学和计算科学这两门交叉学科的不断深入的合作研究，微表情的检测和识别工作也逐渐受到了计算科学尤其是计算机视觉、模式识别和机器学习领域研究者的关注。近年来，世界上多个研究机构先后发布了数个通过高速数码摄像机录制的微表情库。这些种类更加丰富、影像更加清晰的数据集，更加方便了微表情特征的提取和分析工作，为最终微表情的检测和识别工作提供了良好的前置条件。

本章将介绍微表情的基本概念、特征，以及到目前为止，国内外对微表情的研究现状。

1.1 研究背景与意义

1.1.1 基本概念与背景

面部表情 (Facial expression) 是指人脸上通过位于面部皮下肌肉的一处或多处的运动变化，表达人类情绪的一种运动状态^[1]。它是一种人与人之间无声的交流方式，在日常生活中扮演着非常重要的角色。然而，心理学领域里的研究发现，面部表情不仅仅可以由人类无意识的产生。通过有意识地控制表情的产生，人类可以掩盖自己内心的真实情绪，从而达到欺诈的效果。不过，心理学的其他研究发现，人们在日常生活交流中试图掩盖自己内心真实情感的同时，往往伴随着一些运动幅度非常小，持续时间非常短的面部表情。这些伴随而生的面部表情在心理学研究中，被定义为微表情 (Micro expression) ^[2-4]。

微表情是一种持续时间非常短的面部表情。一般的面部表情 (为了书写方便，下文中使用的面部表情都不包含微表情)，也即是人们在日常生活中可以通过肉眼捕捉和识别的面部表情，往往有着数秒的持续时间。相较于这些表情，微表情的持续时间非常短，约为 $1/25 \sim 1/3$ 秒^[4]。1966年，Haggard 和 Isaacs 在一次数小时的影像观测中首次捕捉到了微表情的产生^[5]，然而这次发现并未受到学术界的重视。三年后，Ekman 和 Friesen 在对一份存在自杀倾向病人的调查影像的研究分析中，也独立捕捉到了微表情的产生，并在最终的报告中，使用术语“微表情” (Micro-expression) 来表述这种短暂发生的面部表情^[3]。

除持续时间短之外，微表情的另一个特征为它是人们自发的，也即无意识产生

的面部表情。相比于可控的面部表情而言，微表情更能准确地反映人们的真实情绪。事实上，到目前为止的许多心理学研究，尤其是 Ekman 的研究团队的研究成果，都表明了微表情表达了人们潜在的情绪，能够给观察者提供更丰富、更有价值的信息^[3,4]。

1.1.2 研究意义

由于微表情能够真实反映人们内心对外部世界的情感，使得人们，尤其是有经验的专职工作者，如审讯官、谈判专家、心理学家等，在与人交流的过程中能够通过捕捉这些微表情，获知对方的真实想法，从而根据经验判断，将审讯/谈判的走向掌握在自己手中。

除了在刑侦领域，微表情在安全、医疗、教育等方面都有应用价值。医生在检查患者的时候，可以通过观察患者的反应从而对病情有进一步的了解；心理医生可以通过一系列的语言诱导以及微表情观察，了解患者的心理状况，从而进行更好的心理治疗；教师可以通过学生的神情而选择的不同的教育方式等。

微表情丰富的应用价值，使得微表情的检测和识别一直以来都是心理学领域的研究热点。但是由于微表情的幅度小，持续时间短，人们很难通过肉眼有效地捕捉到微表情并加以判断区分。而随着计算科学的发展，尤其是近年来研究者通过图像处理、计算机视觉、模式识别等手段，有效地解决了面部表情特征的提取、检测和识别工作之后，越来越多的科研工作者尝试使用这些方法对微表情进行研究，并取得了一定的成果。

但同样由于微表情特征幅度小和持续时间短的特性，使得大多数用于面部表情识别的办法无法直接应用到微表情的研究中去。因此，找寻一种合适的表征微表情的特征，并通过合适的识别手段对微表情进行有效的检测和识别，是目前微表情研究中的难点。

1.2 国内外研究现状

微表情的研究起源于 Ekman，同时他的团队也是目前世界上从事微表情研究的主要力量之一。除了 Ekman 团队，目前世界上在微表情研究方面有影响力的团队还有美国的 Matsumoto 和 Shreve 团队，日本的 Polikovsky 团队，加拿大的 Porter 团队，芬兰的赵国英团队以及中国心理所的傅小兰团队。在下文对微表情的现状分析中，我们会提到这些团队在微表情方向所做的工作。

目前世界上对微表情的识别研究大体上可以分为两种：一种是通过心理学实验的方法对人进行训练，提高人对微表情的检测和识别率；另一种是使用计算机领

域的方法,通过提取微表情特征,并采用有效的分类和识别手段对特征进行训练,达到提高计算机对微表情的检测和识别率的效果。本小节将从这两个方面对微表情的研究进展进行阐述。

1.2.1 基于心理学实验方法的微表情检测和识别研究

在早期的心理学研究中,因为计算机学科的发展还不完善,人们对于微表情的研究工作并不依赖于计算机。在当时计算机通常只是作为存储工具或演示工具。在本小节,我们将从心理学实验方法的角度,阐述微表情检测和识别方法的进展。

在心理学研究中,通常有三个类型的基本变量以及它们相互之间的关系需要分析,分别是机体变量,刺激变量和反应变量^[6]。简单来说,机体变量表征的是被试的个体特征,刺激变量表征的是环境特征,反应变量表征的是特定的个体在特定的环境下所表现的反应。

根据对这三个变量的控制和使用方式的不同,心理学研究者通常将心理学实验方式分为两个类别:实验研究和相关研究。实验研究与物理实验中的控制变量法类似。它在保证外部环境一致的情况下,通过改变自变量的值,完成对因变量的测量和分析,从而得到自变量和因变量之间的关系。

相关研究则是通过两个变量的相关性来研究其中某个未知变量的性质。通常,相关研究中的两个变量一个是性质已知的变量,另一个是未知变量。通过两个变量之间的相关性,未知变量的性质可以通过已知变量体现出来。这种研究方法就是相关研究方法。

在已有的微表情识别的心理学研究中,大部分实验过程都采用了相关研究方法进行心理学实验。在实验结果的分析中,使用线性相关系数 r 与假设检验中的 p 值(p -value)衡量两者的相关性。线性相关系数 r 是表征两个变量线性相关程度的值,它的值的绝对值越大,则表明相关性越强。而 p 值表示一种在原假设为真的前提下出现观察样本更极端情况的概率。也即 p 值越小,越表明观测的数据之间是存在关联的。在统计学的显著性检验中,一般以 $p < 0.05$ 表示对两组数据的观测具有显著意义,反之则表示这种观测无显著意义。

利用相关性检验,Ekman 和他的团队使用相关研究的方法发现了很多微表情识别和谎言识别存在显著正相关的证据。1974和1991年,Ekman先后两次与Friesen和Sullivan合作完成了短暂表情识别测验(Brief Affect Recognition Test, BART)^[7,8]。在后一次的实验中,他们发现被试者在微表情识别上的得分与谎言测试的成绩呈显著正相关,其中相关系数 $r = 0.27, p < 0.02$ 。1997年,Frank和Ekman组织了短暂表情识别测验后续的实验,用以支持微表情识别和谎言识别的相关性^[9]。在

这次实验中，被试者的微表情识别成绩和谎言识别成绩的相关系数 $r = 0.34, p < 0.04$ ，这个结果表明了两者之间显著的正相关。而之后，随着实验方法的完善，Ekman、Frank 和 Matsumoto 等研究者先后对上述实验进行了改进，这些实验结果均表明了人们在微表情识别上有与谎言识别相一致的能力，为微表情识别工作的研究提供了重大的实验依据^[10]。

随着对微表情识别能力的实验方法的完善，研究者也开始着力于构建一套微表情识别工具用于人们对微表情识别的辅助训练。2002 年，Ekman 团队率先构建出了世界上第一个，也是目前为止影响力最大的一个微表情训练工具，并以 METT (Micro Expression Training Tool)命名^[11]。METT 包含了 Pre-Test, Training, Practice, Review 和 Post-Test 五个功能单元，如图 1.1 所示，通过预测试和使用 METT 训练之后的测试结果，实验者可以得知受试者使用 METT 的效果。Ekman 团队分别在 2002 年和 2009 年先后对 METT 做了两次评估，评估显示，受训者在 1.5 小时的 METT 工具训练之后，能提高 30% ~ 40% 的微表情识别率^[11,12]。但同时他们也得出了一个结论：由于微表情固有的持续时间短、幅度小的特性，即使是受到过 METT 训练过的人，对微表情的识别率也不会超过 47%。

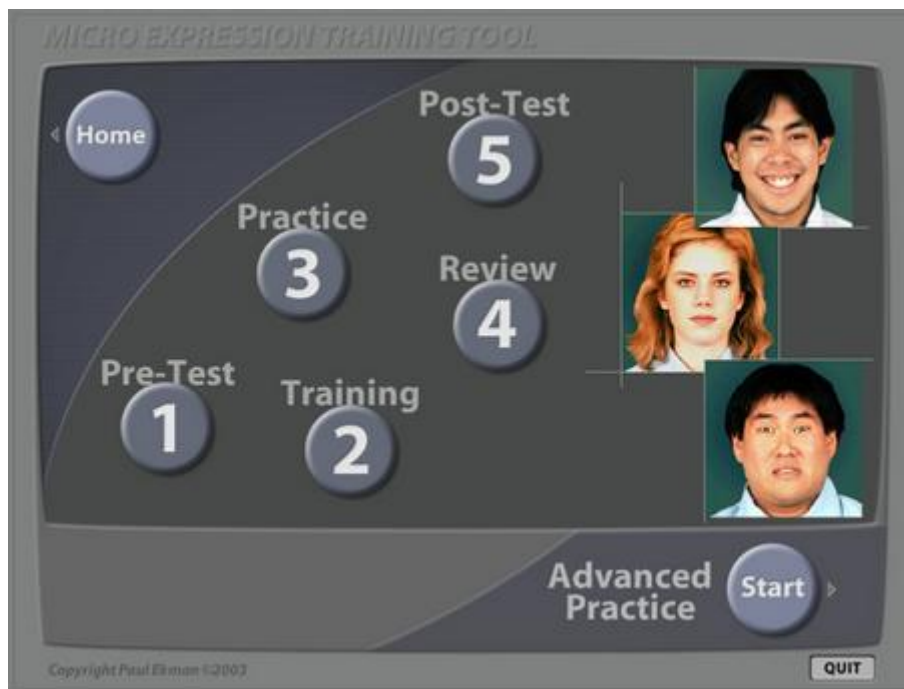


图 1.1 微表情训练工具

由于未受过微表情识别训练的人难以使用肉眼捕捉微表情的发生，因此，在 2002 年之前，许多有关微表情的应用研究都因为人眼对微表情识别率的低下，而难以展开。随着 METT 的出现，许多微表情应用领域的研究者通过使用 METT 完

成了相关领域的应用研究。

Russell 率先于临床医学领域展开对微表情的应用研究。他分别于 2006 年和 2008 年使用 METT，展开了针对精神分裂患者的情绪识别和微表情识别能力，并获得了精神分裂患者和正常人一样有着相当的情绪识别和微表情识别能力这一成果^[13,14]。除了临床医学领域，2007 年，Fellner 将 METT 的应用拓展到了情绪智力研究领域^[15]。2009 年，METT 被应用到了国家安全和政治心理学领域^[16]。

从上述 METT 的应用研究可以了解到，通过构建出一套高效的提高微表情识别率的训练工具，可以有效的将微表情的研究推广至应用领域，使得越来越多的研究可以通过识别微表情来得到令人满意的成果。

1.2.2 基于特征提取的微表情检测和识别方法

随着计算机图像、视觉和模式识别等领域的发展，使用特征来表征图像和视频信息并加以检测和识别的方法，得到了越来越多的研究者的关注。同时由于使用目前世界上效果最好的微表情训练工具 METT，也不能将人对微表情的识别率提高到 50% 以上，使得越来越多的研究者意识到了通过人眼识别微表情的地板效应^[11,12]。因此，使用计算机来解决微表情的检测和识别问题成为了研究者研究微表情识别问题的又一途径。

不过如前文所述，尽管目前世界上已经有了成熟的检测和识别面部表情的方法，但是由于微表情持续时间短，运动幅度小的特性，使得这些方法难以运用到微表情的检测和识别方法中去。因此找到一种可以提高计算机识别能力的微表情特征是目前使用计算机自动识别微表情研究中的重点。

2009 年，Polikovsky 等人使用三维梯度向量作为微表情特征的描述符，在微表情的自动检测方面完成了突破^[17]。他们的方法如下，对于一个给定的微表情序列：

- 1) 确定特征区域，如图 1.2 所示，方法如下：a) 通过手工标定的方法，在人脸上选择 12 个特征点；b) 根据这 12 个点，计算特征区域的中心；c) 根据区域中心，确定人脸上的矩形区域，并将这些区域组合成 8 个特征区域。这些确定的特征区域即为文章中认为的能表征微表情特征的有效区域。

- 2) 提取特征区域的三维梯度向量，方法如下：a) 将人脸上的 8 个特征区域拓展至时域上，变成一个三维特征区域，如图 1.3 所示；b) 对三维特征区域计算其 3D 梯度特征；c) 使用方向直方图统计特征区域内 3D 梯度特征，生成 3D 梯度直方图向量。

- 3) 使用 K 均值聚类的方法，对每一个三维特征区域计算其聚类中心，并使用

这些聚类中心对其他微表情序列进行分类。

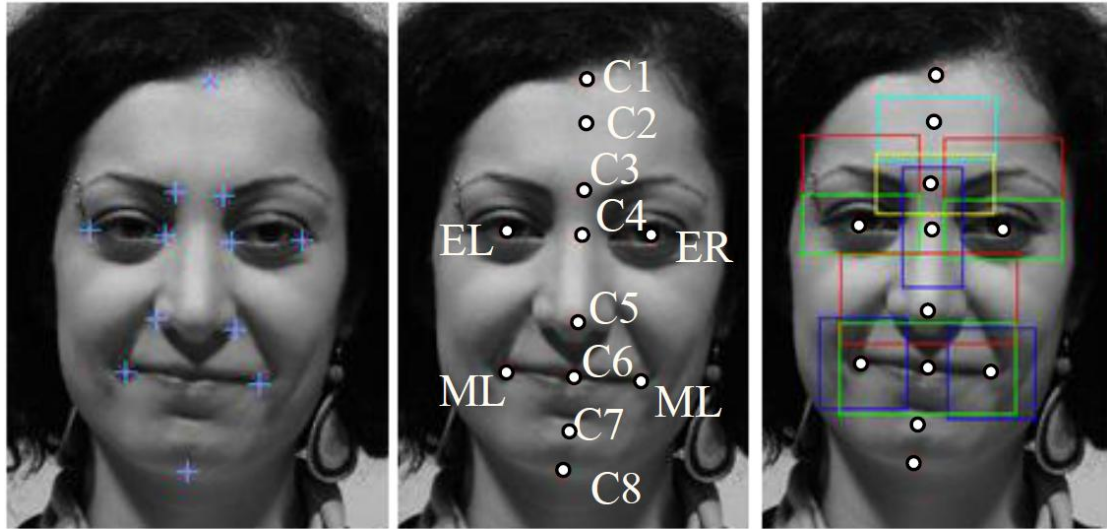


图 1.2 计算微表情特征区域，摘自[17]

Polikovsky 的方法在检测微表情产生上得出了很好的结果，但是其中同样存在着许多问题：1) 绘制的特征区域在平面上有重合；2) 由于录制微表情的过程中，被试者往往会产生轻微的头动，该文章没有考虑这一影响。3) 文章中使用的微表情数据集是非自发产生的微表情，在采集过程中，研究人员要求被试者尽快的使用尽可能小的幅度去模拟 7 个基本的面部表情，而 Ekman 的研究中已经表明模拟的微表情和自发的微表情有着本质的区别^[2]。

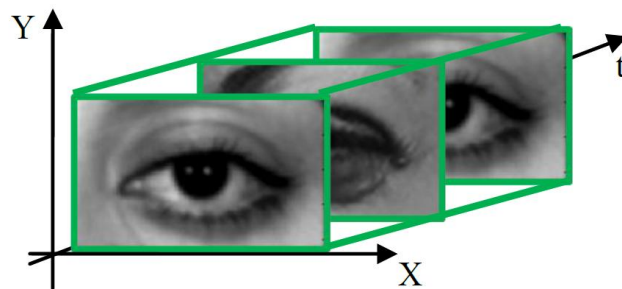


图 1.3 三维特征区域，摘自[17]

随后，Shreve 在 2010 年和 2011 年相继给出了分别在短视频和长视频中检测和区分微表情和表情的方法^[18,19]。Shreve 使用图像的张力模型来表示运动的幅度，并设定张力的阈值来区分微表情和表情，具体方法如下：

- 1) 首先使用光流法获得人脸在视频序列中的运动特征；
- 2) 通过获得运动特征，获取图像的张力大小，并使用阈值加以界定。

如图 1.4 所示，为表示微表情“轻蔑”的张力示意图。其中，a)表示原始图像，b)表示光流图，c)表示张力图。可以明显看出，在表情开始的时候，人脸（嘴角区域）的张力比较小，而在表情中，人脸（嘴角区域）的张力比较大。

Shreve 给出的使用张力模型在视频序列中检测表情和微表情的方法同样存在着以下问题：1) 只能检测和区分表情和微表情，无法做到微表情类内的识别工作；2) 在求张力模型的同时也没有考虑被试者轻微头动的问题；3) 同 Polikovsky^[17]使用的是数据集一样，Shreve 使用的 USF-HD 数据集也是通过被试者模拟基本表情，获得的非自发的微表情库。

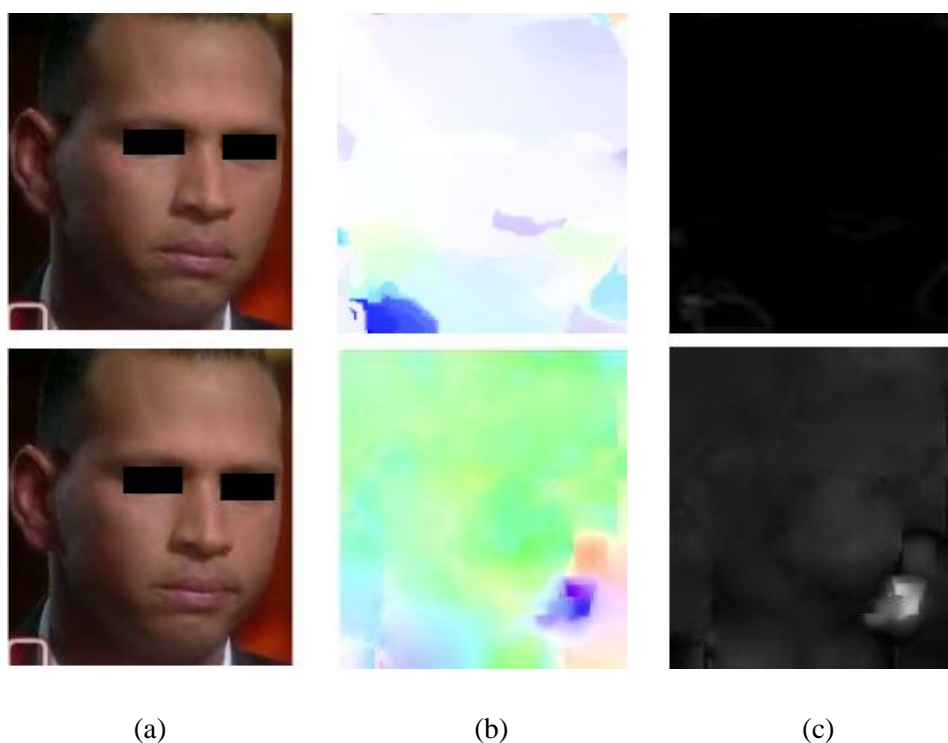


图 1.4 微表“轻蔑”的光流图和张力图；其中 (a)原始图形，(b)光流图像，(c)张力图像，摘自[18]

2011 年，Pfister 等发表了第一篇检测和识别自发产生的微表情研究论文^[20]。文中使用 LBP-TOP(Local Binary Patterns on Three Orthogonal Planes)^[21]作为表征微表情特征的描述符，并在一个自发的微表情库 SMIC 上完成了微表情的检测和识别的实验，获得了比较好的识别率，同时发现了 LBP-TOP 的识别率与视频序列的帧率不相关的特性。如图 1.6 所示，他们的方法如下：

1) 首先将一个微表情序列的每一帧根据标准脸进行标准化，获得标准化后的视频序列，如图 1.5 所示；

2) 通过时域插值模型，将视频序列的长度进行归一化；

- 3) 根据人脸图像上的特征点，计算微表情的 LBP-TOP 特征；
- 4) 使用多核分类器对微表情序列进行检测和识别。

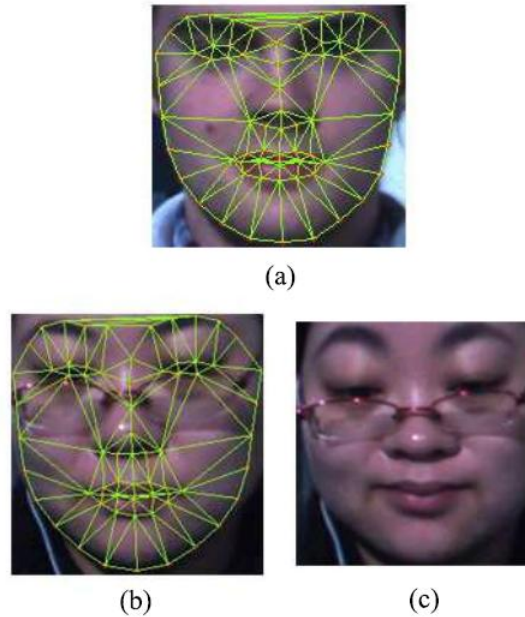


图 1.5 图像序列的标准化；a)标准脸及其特征点；b)需要标准化的人脸图像及其特征点；c)标准化后的人脸图像，摘自[20]

通过对人脸图像的标准化，Pfister 的方法在一定程度上解决了因为被试者轻微头动对特征提取带来的影响。但他们的方法依然存在两个问题：1) 在使用标准脸对人脸图像进行标准化的同时，由于图像保存时的截断误差，破坏了人脸的结构，影响了对 LBP-TOP 特征的提取；2) 对人脸特征区域的选择比较粗糙，没有考虑到人脸各个区域的运动方向的不一致性。

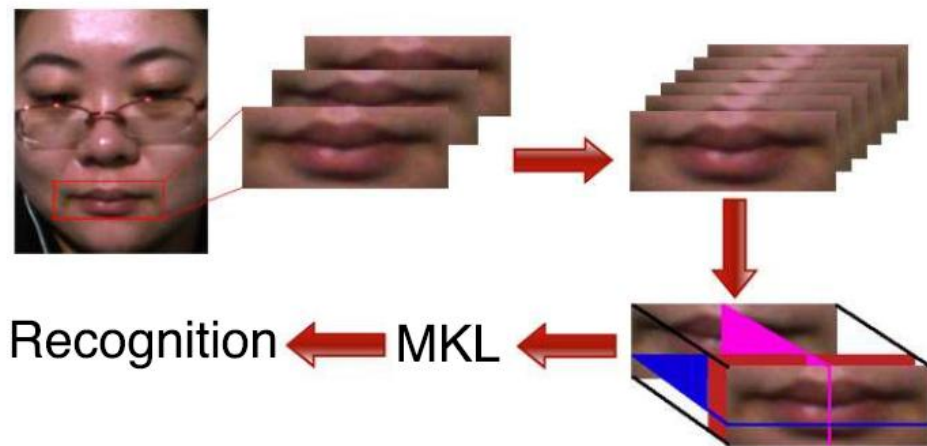


图 1.6 使用 LBP-TOP 特征的微表情识别方法，摘自[20]

2014 年, Wang^[22]将 LBP-TOP 特征应用到张量独立的颜色空间 (Tensor Independent Color Space, TICS), 并与 LBP-TOP 特征与传统的 RGB 空间和灰度空间的实验结果进行比较。并得出了基于张量独立颜色空间的 LBP 特征的识别率稍高于 RGB 空间和灰度空间的结论。

1.3 本文主要工作

由于微表情持续时间短, 运动幅度小, 人们很难在日常生活中通过肉眼捕捉和识别微表情。同时 Ekman 等人的研究发现, 即使是使用过 METT 训练的人, 他们对围标的识别率也不会超过 47%, 因此找到一种通过计算机完成自动检测和识别微表情的方法是非常迫切的事情。

为了解决上述问题, 本文提出了一种基于运动主方向特征的微表情识别方法。该方法通过 1) 将人脸分成若干个感兴趣区域, 2) 提取感兴趣区域的主方向特征, 3) 标准化特征, 并使用 SVM 完成微表情的训练和识别工作。本文提出的运动主方向特征具有下述特点: 1) 对于帧数长度不敏感; 2) 对于角色不敏感; 3) 计算方便。

本文的结构如下, 第一章为引言, 介绍研究背景和研究现状; 第二章讲述基于运动主方向的微表情特征的提取方法和实验结果; 第三章讲述通过光流域矫正的微表情运动特征的提取和识别方法; 第四章为本文方法与 LBP-TOP 特征的对比结果, 第五章为全文总结。

第2章 基于运动主方向的微表情特征的提取方法

如前文所述, 由于在医疗、教育、刑侦、国家安全等许多方面都有着潜在的用途, 微表情的研究工作在心理学领域里一直都扮演者非常重要的角色。但是, 心理学研究发现, 即使是受到过严格训练的人, 在检测和识别微表情方面都存在着严重的不足, 他们表现出来的识别率仅有 47%^[11,12]。这样的识别率, 以及严苛的训练手段, 严重制约了微表情在各个应用领域中的使用。

随着近年来计算机学科的发展, 心理学和计算科学之间的交叉领域越来越多, 交叉学科的研究使得越来越多的心理学问题可以通过计算机解决。微表情的检测和识别问题同样如此。尤其是近年来, 计算机视觉和机器学习等学科的不断发展, 有了更多和更适合的图片特征提取方法和机器学习方法, 使得越来越多的从事计算机视觉和学习的研究者投入到微表情的检测和识别工作中。

为了解决这一问题, 本章提出了一种基于运动主方向的微表情特征的提取方法, 使用该特征, 可以有效解决微表情的检测和识别问题。该方法包含: 1) 检测和分离人脸, 2) 将人脸区域分为若干个 ROI 区域; 3) 使用稠密光流法获取人脸区域的运动特征; 4) 使用方向直方图获取 ROI 区域运动主方向; 5) 通过序列特征的归一化, 获取微表情特征。对本章提取的微表情特征, 我们使用 SVM 分类器, 将其在三个微表情数据集中进行评估测验。

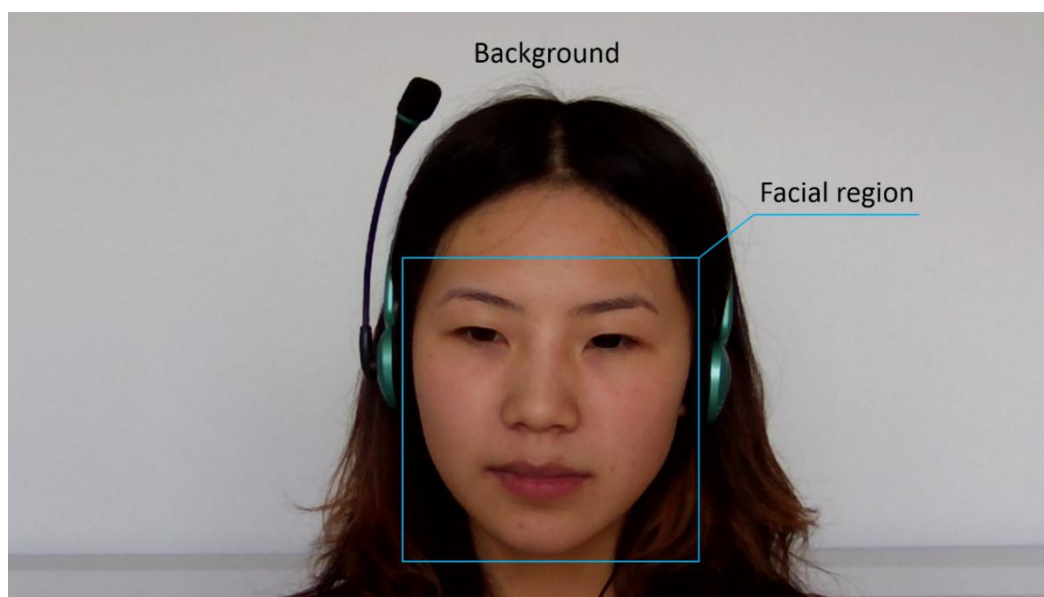


图 2.1 微表情采集图像的有效区域

2.1 图像预处理

由于微表情数据集中（以 CASME 数据集为例），其视频/图片的分辨率为 1280*720，而其中人脸的大小约为 350*350，其有效区域不超过 14%，如图 2.1 所示。考虑到计算效率，我们有必要对这些视频序列进行预处理工作。在本小节中，我们首先使用局部约束模型检测人脸区域的特征点，然后根据这些特征点将人脸进行分割以及感兴趣区域编码。

2.1.1 局部约束模型与人脸检测

局部约束模型（Constrained local model, CLM）^[23]是一种比活动外观模型（Active Appearance Models, AAM）^[24]和活动形状模型(Active Shape Model, ASM)^[25]更有效、鲁棒的模型检测工具，也是人脸检测中的经典的方法。2006 年，David C.和 Tim Coots 在 ASM 和 AAM 的基础上提出了局部约束模型：

$$x = \bar{x} + P_s b_s \quad g = \bar{g} + P_g b_g \quad (2-1)$$

CLM 是一种基于形状和纹理的混合模型，如图 2.2 所示。形状模型 x 表示人脸上特征点的坐标，纹理特征 g 表示使用如下方式得到的一个灰度向量：以训练图像上的每一个特征点为中心选择一个矩形块，并将这些矩形块灰度化之后，拉伸而成的一个灰度向量。其中 \bar{x} 表示所有训练数据中人脸上特征点的均值坐标，即平均脸坐标。 P_s 是训练数据中所有特征点的混合矩阵使用 PCA 降维后，获取的主成分特征。 b_s 表示这些特征的权值。同理， \bar{g} 表示均值纹理， P_g 表示纹理的特征矩阵， b_g 表示纹理特征的权值。

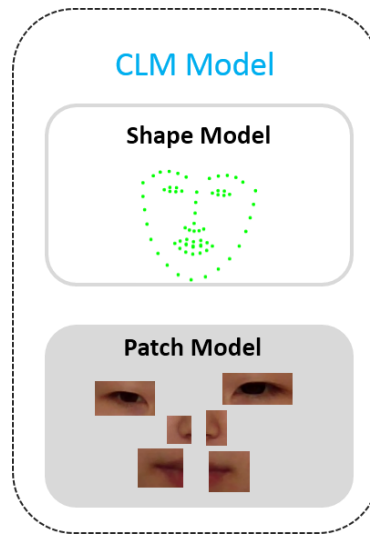


图 2.2 CLM Model

对于一张给定的人脸图片，可以使用如图 2.3 所示的 CLM 搜索算法获得图像区域表征人脸的特征点集合。该算法如下：

- 1) 在人脸图片上给定一个初始形状点集合；
- 2) 根据当前形状特征点，获取其纹理特征；
- 3) 根据形状特征和纹理特征，生成 CLM 模板；
- 4) 根据 CLM 模板在当前人脸图像上匹配图片；
- 5) 根据匹配结果，返回新的形状点集合；
- 6) 重复步骤 2)~5) 直至算法收敛。

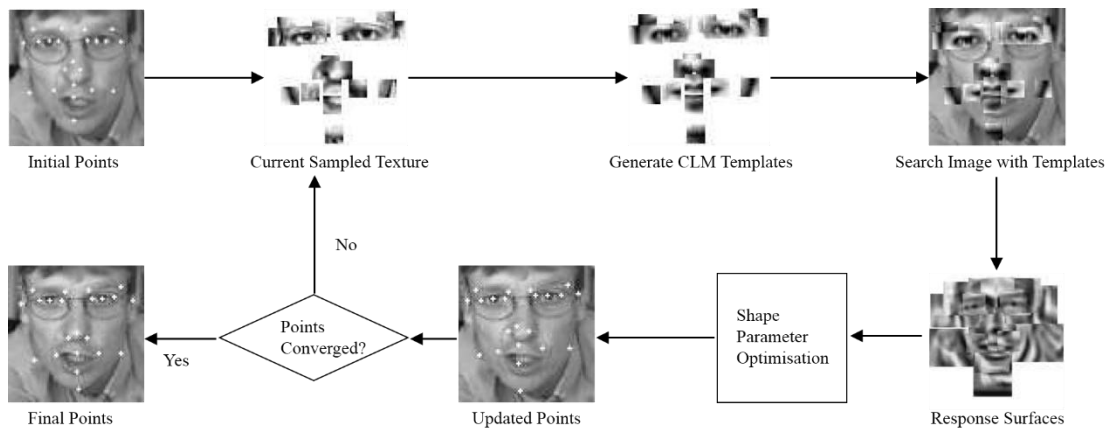


图 2.3 CLM 搜索算法，摘自[23]

在本文的实现中，我们使用 CLM 模型和搜索算法获取 66 个人脸特征点，其方法如下：首先，为了降低 CLM 在大分辨率图像上的搜索代价，我们使用 Viola-Jones^[26]人脸探测器对每个微表情序列的第一帧进行人脸的定位，并获得初始的特征点集合；然后，使用 CLM 搜索过程，根据 ASM 和 AAM 给出的形状和纹理的统计模型，不停迭代更新这些特征点的位置，直至收敛。ASM 和 AAM 给出的统计模型是针对人脸上 68 个点进行迭代计算得到的特征点集合，但其中的两个位于嘴唇上点与集合内其他的点重合，因此在对这些特征点的使用中，我们只保留了其中 66 个特征点。最后根据这 66 个点的位置，对人脸进行分割。如图 2.4 所示，其中 (a) 表示 66 个 CLM 特征点，(b) 表示切割之后的人脸区域。

2.1.2 基于 FACS 的人脸感兴趣区域编码

Facial Action Coding System(FACS)^[27] 是一个对人脸上所有的面部运动使用数字编码的形式编写而成的系统。它将相互独立的面部的运动单元 (Action Units, AUs) 作为编码的基本元素。因此，FACS 可以将绝大部分的面部表情，使用唯一

的 AUs 组合进行表征。

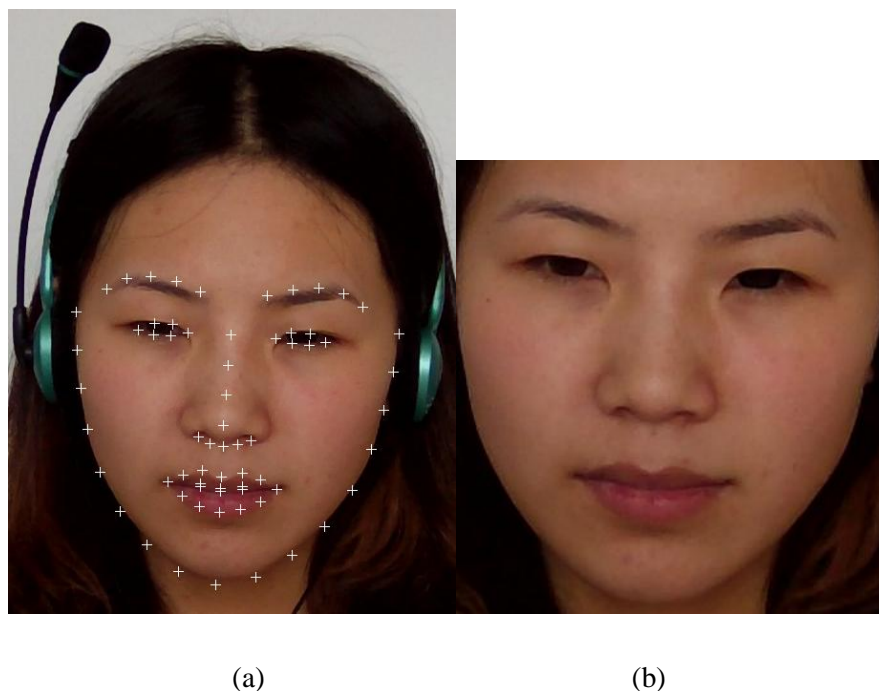


图 2.4 使用 CLM 搜索算法获取的 66 个特征点和切割结果

为了把人脸分割成数个拥有运动方向一致性的感兴趣区域（Region-of-Interest, ROI），我们根据 AUs 将人脸分割成为 36 个 ROI 区域。分割方法如下，1）通过 2.1 节的人脸检测方法，获取人脸上 66 个特征点；2）根据 66 个特征点集合，获取 22 个辅助点，如图 2.5（a）所示，其中白色的点为 CLM 特征点，绿色的点为辅助点；3）根据这 88 个点，将人脸分为 36 个 ROI 区域，分别用编号 R_1 - R_{36} 表示，如图 2.5(b)所示。这些感兴趣区域几乎覆盖了 FACS 中所有的面部运动单元，且不影响这些区域的运动一致性，它们和人脸区域的对应关系如表 2.1 中所述。

2.2 基于运动主方向的 ROI 区域运动特征的提取方法

如 1.2 节所述，之前的许多关于微表情的研究工作都是通过检测微表情运动的强度展开的，但是这些研究中经常忽略了方向特征在微表情检测和识别的作用。比如，被检测的视频序列和已知的一个微表情有着相同的运动强度，以及不同的运动方向，那么在之前的研究工作中，这个视频序列就会被检测成一个微表情，而事实上它或许只是镜头的轻微摆动所造成的。同时由于缺少表征运动方向的特征，许多类别的微表情在各个面部区域的运动幅度有着极强的相似性，因此仅仅提取运动强度作为表征微表情的特征，无法完成微表情的类内识别研究。

为了解决上述问题，本文提出了一种基于运动主方向的 ROI 区域运动特征的提取方法。通过该方法提取的运动特征，除了能够采集 ROI 区域中的方向信息之外，还可以通过统计的方法，获取了这个 ROI 中面部运动的主方向。其过程如下：

1) 使用稠密光流算法获取人脸图像的运动特征；2) 对给定的 ROI，获取其运动方向的直方图；3) 根据直方图，得到表征 ROI 区域的运动主方向特征。

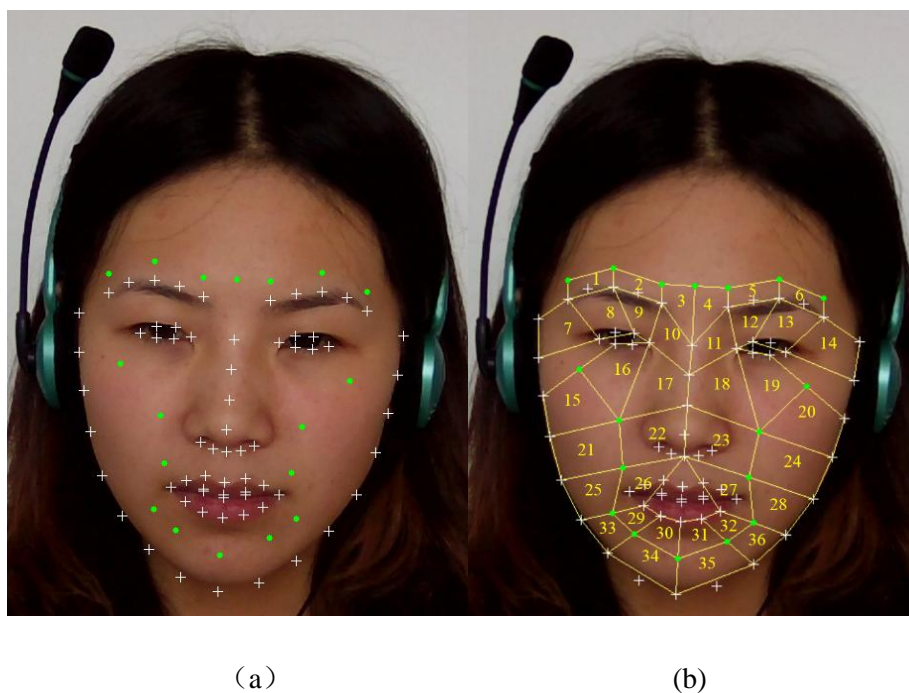


图 2.5 (a)66 个 CLM 特征点和 22 个辅助点; (b)36 个 ROI 区域编码

表 2.1 面部区域与 ROI 之间的对应关系

Region Name	ROIs
Outer eyebrows	R1, R6, R7, R8, R13, R14
Inner eyebrows	R2, R5, R9, R12
Forehead	R3, R4
Nose root	R10, R11
Nose bridge	R17, R18
Cheeks	R15, R16, R19, R20, R21, R24
Nasal horn	R22, R23
Mouth corner	R25, R26, R27, R28, R29, R32, R33, R36
Jaw	R30, R31, R34, R35

2.2.1 基于光流算法的人脸图像运动特征的提取方法

光流法^[28]是计算机视觉中一种广为人知的运动检测方法，它经常被用于运动物体的检测和跟踪的研究中。光流特征是通过光流法计算像素点在单位时间上的强度变化得到的，它表征了这个特定位置上像素的运动速率和运动方向。对一个视频片段而言，视频中的一个像素点的位置都可以使用三维坐标 (x, y, t) 表示，其中 (x, y) 表示像素点的在空间上的位置， t 表示像素点在时间上的位置。同时，我们使用符号 $I(x, y, t)$ 表示像素点的强度。根据光强不变原则，处于时刻 t 的像素点运动到下一帧位置 $(x + \Delta x, y + \Delta y, t + \Delta t)$ 的时候，它的光强不变，即：

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (2-2)$$

其中 $(\Delta x, \Delta y)$ 为这个像素点在二维平面内的位移， Δt 为相邻两帧之间的时间间隔。在位移很小的前提下，式（2-2）可以使用泰勒展开，即：

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + \tau \quad (2-3)$$

其中， τ 为泰勒展开的高阶无穷小量。根据式（2-2）和（2-3）可以得到：

$$\frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0 \quad (2-4)$$

即：

$$\frac{\partial I}{\partial x} \frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y} \frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t} = 0 \quad (2-5)$$

令 $u = \frac{\Delta x}{\Delta t}$, $v = \frac{\Delta y}{\Delta t}$ 分别表示像素点 (x, y, t) 在 x 和 y 方向的速率。则式（2-5）可以写成如下形式：

$$\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} = 0 \quad (2-6)$$

因此，可以使用向量 $V = [u, v]^T$ 则表示了像素点 (x, y, t) 在相邻两帧之间的光流值，其中，两帧之间的时间间隔为 Δt 。

本文使用 Brox^[29]算法获取视频序列的运动特征。根据 Brox 算法计算的结果，每一张图片的稠密光流特征都可以用一个与图片分辨率一致的二维矩阵表示，如式（2-7）所示：

$$M = \begin{bmatrix} V_{0,0} & \cdots & V_{w-1,0} \\ \vdots & \ddots & \vdots \\ V_{0,h-1} & \cdots & V_{w-1,h-1} \end{bmatrix} \quad (2-7)$$

其中 M 表示图像的运动特征， $V_{i,j}$ 表示在图像上位于坐标 (x, y) 上的运动矢量

(如图 2.6 所示)，参数 w 和 h 分别表示图像的宽度和高度。

2.2.2 基于方向直方图的 ROI 主方向特征的提取

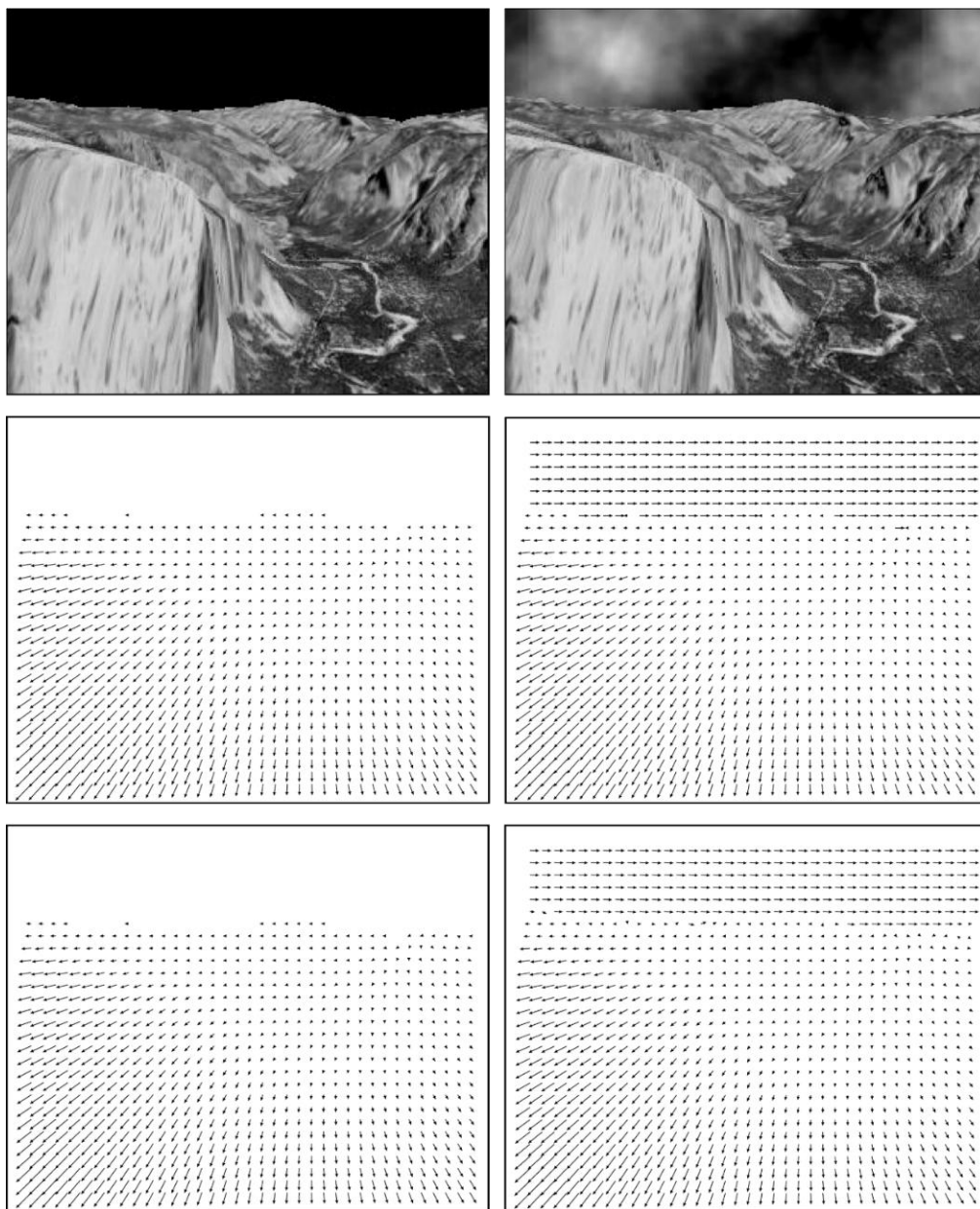


图 2.6 通过光流法获取的视频序列的运动矢量图 (a) 左上角: 原始帧。(b) 右上角: 与 a 对应的添加云雾之后的帧。(c) 中间靠左: 图 a 的光流特征真值。(d) 中间靠右: 图 b 的光流特征真值。(e) 左下角: 使用 Brox 算法得到的图 a 的光流特征。(f) 右下角: 使用 Brox 算法获得的图 b 的光流特征。摘自[29]

Polikovsky^[17]的方法, 采用方向直方图对 ROI 区域的 3D 梯度特征进行统计, 并将获得的方向直方图特征作为微表情识别的重要依据。然而, 根据 Polikovsky 的方法, 它的每一个 ROI 区域最终都会使用一个 32 维的特征向量进行表示, 对于

ROI 区域多的分割方法而言, 这样的特征表示方式容易产生维度上的灾难。为了在增加 ROI 区域的同时, 减少特征向量的维数, 本文提出了 ROI 区域的主运动方向特征的提取。

为了获取人脸上 ROI 区域的主运动方向特征, 我们首先使用方向直方图 (Histogram of Oriented Optical Flow, HOOF) [30], 如图 2.7 所示, 对 ROI 区域内所有像素点的运动矢量进行统计。

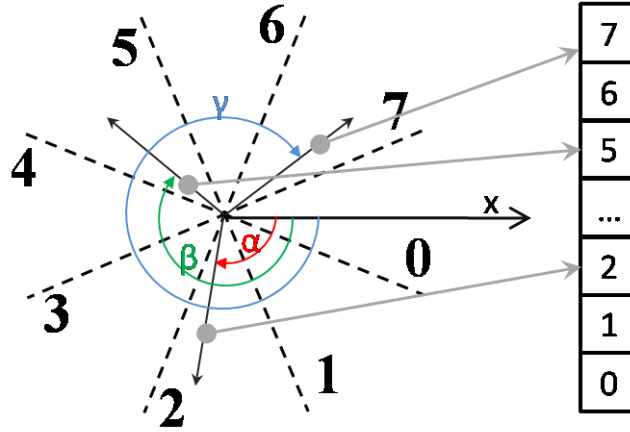


图 2.7 方向直方图

对于一个给定的 ROI 区域, 假设 p 为任意一个落在这个 ROI 区域内的一个像素点。根据式 (2-7), 可知像素点 p 的运动矢量可以用向量 $V_p = [u_p, v_p]^T$ 表示。为了便于书写, 我们在这里为符号 V_p 引入一个对偶符号 $U_p = (\rho_p, \theta_p)$, 其中符号 ρ_p 和 θ_p 分别称为为向量 V_p 的长度和方向。如图 2.7 所示, 每一个运动矢量 V_p 都根据其方向 θ_p 落入到 8 个桶中, 在这里我们使用数字 $\mathcal{N} = \{0, 1, \dots, 7\}$ 分别表示这 8 个不同的类别。其中数字标签和角度之间的关系可以使用式 (2-8) 表示:

$$\mathcal{F}(V_p) = n, \text{ if } \theta_p \in \left(\frac{2n-1}{8}\pi, \frac{2n+1}{8}\pi \right], \text{ where } n \in \mathcal{N} \quad (2-8)$$

其中函数 $\mathcal{F}(V_p)$ 是将运动矢量 V_p 进行分类的函数, n 表示集合 $\mathcal{N} = \{0, 1, \dots, 7\}$ 中的一个数字。

根据运动矢量的方向直方图, 我们可以得知该给定的 ROI 区域内像素点的总体运动方向。不过, 尽管在对人脸进行 ROI 分区的时候已经尽量考虑保证面部运动的一致性, 但是在 ROIs 的边缘衔接处, 还是会出现运动方向与整体运动方向不

一致的情况。不过，令人高兴的是，这种产生不一致运动方向的像素点个数在整个面部表情的生命周期里，总是占少数。如图 2.8 所示，为微表情“happiness”位于嘴角区域的 R_{26} 的方向直方图，其中每一列的左边是 R_{26} 的将各个运动方向用箭头表示出来的示意图，右边是对应的方向直方图。从图 2.8 可以看出，在微表情的生命周期里，随着微表情表现的越强烈，它所在区域的运动方向越趋于一致。

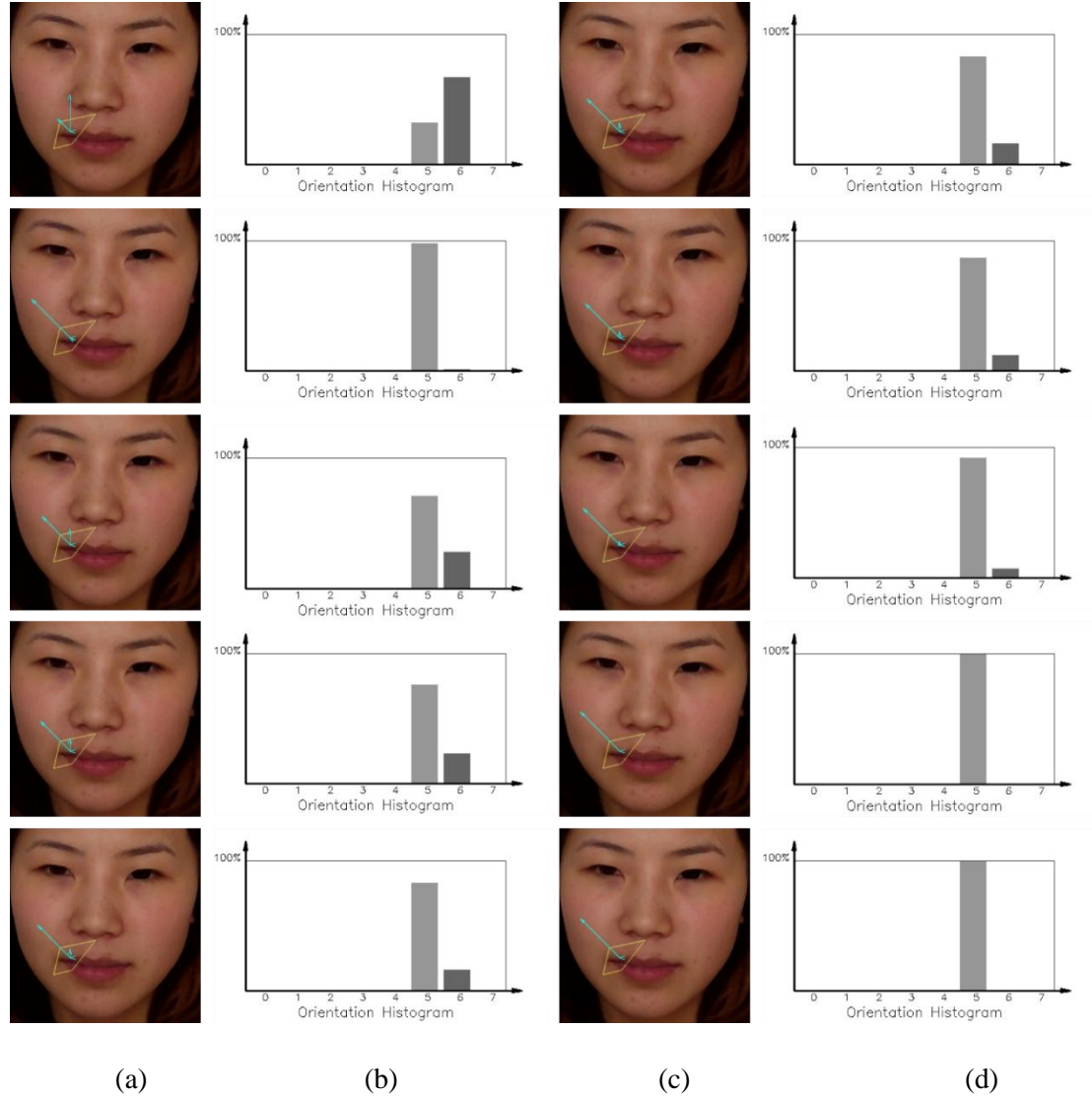


图 2.8 happiness 微表情在 R_{26} 区域的方向直方图。(a) (c) 列为微表情序列，(b) (d) 列为对应序列的方向直方图

为了减少一些 ROI 区域中与总体运动方向不一致的边缘点的影响，我们采用直方图中含像素点最多的方向为该 ROI 区域的主方向，使用与主方向一致的运动矢量的均值特征作为该 ROI 区域的主方向特征，如式 (2-9) 所示：

$$\bar{V} = \frac{1}{|B_{max}|} \sum_{V_p \in B_{max}} V_p \quad (2-9)$$

其中, B_{max} 表示方向直方图含元素个数最多的集合, 操作符 $|\cdot|$ 表示取集合的大小, $\bar{V} = [\bar{u}, \bar{v}]^T$ 表示该 ROI 区域的运动方向特征。

2.3 基于加权平均的微表情运动特征的提取方法

通过 2.2 节, 我们可以通过方向直方图的方法获取图像上任意一个 ROI 区域的运动主方向特征。将此方法运用到视频序列的每一帧, 即可获得视频序列的运动特征, 其方法如下所述。

对于一个给定的视频序列 $\Gamma = \{F_1, F_2, \dots, F_n\}$, 其中 F_f 为视频序列中的第 f 帧, 下标 n 表示视频序列的长度。我们用符号 ψ_f 来表示帧 F_f 的运动特征, 如式 (2-10) 所示:

$$\psi_f = (\bar{V}_f^1, \bar{V}_f^2, \dots, \bar{V}_f^{36}) \quad (2.10)$$

其中 \bar{V}_f^k 表示视频序列 Γ 中第 f 帧上第 k 个 ROI 区域的运动特征。

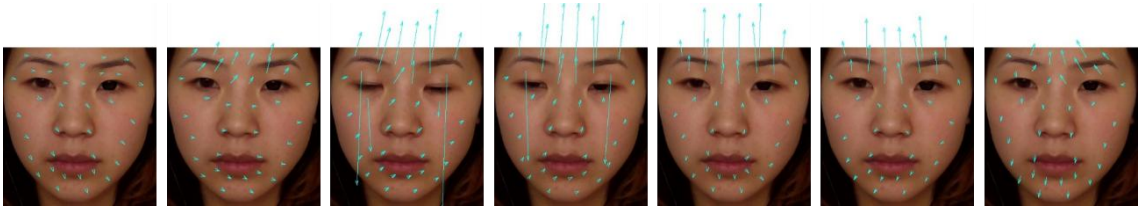


图 2.9 表示“惊讶”的微表情的运动特征序列

因此, 微表情序列 Γ 可以用一组特征序列 $\Psi = (\psi_1, \psi_2, \dots, \psi_n)$ 表示, 如图 2.9 所示, 为表征“惊讶”的表情的运动特征序列。

考虑到数据集中每个视频序列的长度都不尽相同, 同时也为了降低特征向量的维度, 我们使用加权平均的方法, 对微表情序列的特征进行标准化, 标准化公式如下:

$$\bar{\Psi} = \sum_{f=1}^n \alpha_f \psi_f, \text{ 其中 } \alpha_f = \frac{\sum_{k=1}^{36} \bar{\rho}_f^k}{\sum_n \sum_{k=1}^{36} \bar{\rho}_n^k} \text{ 且 } \sum_{f=1}^F \alpha_f = 1 \quad (2-11)$$

其中 $\bar{\rho}_f^k$ 为 $\bar{V}_f^k = [u_f^k, v_f^k]^T$ 的对偶形式中 $\bar{U}_f^k = (\bar{\rho}_f^k, \bar{\theta}_f^k)$ 的表征矢量长度的值。 α_f 是第 f 帧的权值, 表征这一帧的运动强度在整个视频序列中所占的比重, 总和为 1。

从式 (2-11) 中可以看出, 视频序列中, 运动强度越高的帧, 在最终获得的微

表情运动特征中所占的比例就越高。从而使得微表情的加权特征总体呈现视频序列中表现强度最强的那几帧，于此同时，使用人脸上的所有 ROI 区域强度的总和作为权值的加权方法，可以在一定意义上去除人眼眨动的影响。如图 2.10，为图 2.9 的微表情序列加权平均得到的微表情运动特征的示意图。

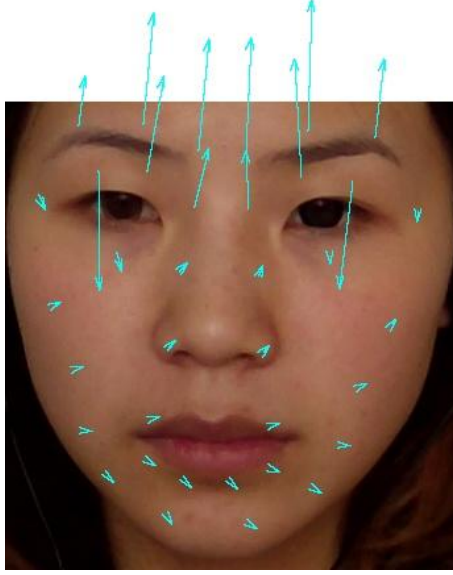


图 2.10 微表情“surprise”的加权平均运动特征

视频序列通过加权平均后的微表情特征可以用如下公式表示：

$$\bar{\Psi} = [\bar{V}_1, \bar{V}_2, \dots, \bar{V}_{36}] \quad (2-12)$$

即，使用运动矢量 \bar{V} 的对偶特征 \bar{U} 可以表示为：

$$\bar{\Psi} = [\bar{U}_1, \bar{U}_2, \dots, \bar{U}_{36}] = \left[\left(\frac{\bar{\rho}_1}{\bar{\theta}_1} \right) \left(\frac{\bar{\rho}_2}{\bar{\theta}_2} \right) \dots \left(\frac{\bar{\rho}_{36}}{\bar{\theta}_{36}} \right) \right] \quad (2-13)$$

然后，进一步将微表情运动特征的强度进行归一化：

$$\tilde{\rho}_k = \frac{\bar{\rho}_k}{\max\{\bar{\rho}_r, r=1, 2, \dots, 36\}} \quad (2-15)$$

从而得到微表情序列特征归一化后的向量形式：

$$\tilde{\Psi} = \left[\left(\frac{\tilde{\rho}_1}{\bar{\theta}_1} \right) \left(\frac{\tilde{\rho}_2}{\bar{\theta}_2} \right) \dots \left(\frac{\tilde{\rho}_{36}}{\bar{\theta}_{36}} \right) \right] \quad (2-16)$$

2.4 测试评估

在前文中，我们通过方向直方图的方法获取了基于运动主方向的微表情特征，

并使用特征向量 $\bar{\Psi}$ 进行表示。为了在本小节，我们将在 CASME, CASME2 和 SMIC 这三个微表情数据集上对我们提出的微表情运动特征进行实验评估，并对其中的结果进行分析。

2.4.1 测试方法

为了便于使用 SVM 分类器对本文提出的基于运动主方向的微表情特征进行实验评估，我们首先将微表情运动特征分成两个部分，强度特征 P 和方向特征 Q 。其中，长度和方向特征可以使用如下公式表示：

$$P = (\tilde{\rho}_1, \tilde{\rho}_2, \dots, \tilde{\rho}_{36}) \text{ and } Q = (\cos(\theta), \sin(\theta)) \quad (2-17)$$

其中 $\theta = (\bar{\theta}_1, \bar{\theta}_2, \dots, \bar{\theta}_{36})$ 。同时为了评估在微表情运动特征中强度和方向对识别率的影响程度，我们在式 (2-17) 的基础上引入参数 $\lambda \in [0,1]$ ，用以表征强度特征和方向特征在微表情识别中的重要程度，因此，微表情序列的特征向量改写成如下形式：

$$\bar{\Psi} = (\lambda P, (1 - \lambda)Q) \quad (2-18)$$

对于一个给定的微表情数据集，我们使用上文所述的方法提取其运动主方向特征，并将特征向量改写为式 (2-18) 的形式作为训练和预测的样本。同时使用基于 POLY 核函数的 SVM^[31]分类器对样本进行训练和分类。其中 POLY 核函数形式及他的参数如下：

$$\mathcal{K}(x_i, x_j) = (\gamma x_i^T x_j + coef)^{degree} \quad (2-19)$$

在实验中，我们固定其中的参数为： $\gamma = 0.1, degree = 4, coef = 1$ 。

在训练和测试过程中，我们使用 Leave-One-Video-Out 的实验方法，在三个数据集的原始帧率上进行评估。Leave-One-Video-Out 意思是将数据集中的一个小微表情数据取出来作为测试数据，其他微表情数据作为训练数据而得到的实验结果。在实验结果的展示里，除了展示参数 λ 的影响之外，我们还给出了各个类别之间的混淆矩阵，用以表征数据集中各个类别之间的相似关系。最后我们将对这些结果进行简单的分析。

2.4.2 测试数据

为了对本文提出的基于运动主方向的微表情特征进行较为全面的测验和评估，我们选择在 CASME^[32]，CASME2^[33]和 SMIC^[20]这三个数据集上进行测试。这三个数据集中的微表情都是在被试者无意识情况下产生的，是数据库的建立者通过使

用高速摄像机进行录制和分类而建立起来的。而且，这三个数据集是世界上少数几个非模拟产生的微表情数据库。在本小节，我们将对这三个数据集进行简单的介绍。

表 2.2 CASME 数据集上的微表情分类

数据集	Positive	Negative	Surprise	Others
数据集 A	6	5	7	78
数据集 B	5	49	14	33
CASME	11	54	21	111

CASME 数据集是中科院心理所的 Yan^[32]等人建立的一个全部由被试者自发产生的微表情组成的数据集。它包含 A、B 两个子集，是分别使用不同的采集环境得到的微表情数据集。其中数据集 A 包含 7 个被试者共 96 个微表情数据，这些微表情数据是使用分辨率为 1280×720 ，帧率为 60fps 的数码相机在自然光照射的条件下采集的。数据集 B 包含 12 个被试者共 101 个微表情数据，这些微表情是使用分辨率为 640×480 ，帧率为 60fps 的数码相机在 LED 光照条件下采集的。在这些数据采集完毕之后，研究者将这些微表情数据根据 FACS 以及被试者的实验反馈，将包含微表情的视频序列单独分离开来，并将其分为四大类：Positive, Negative, Surprise 以及 Others。如表 2.2 所示，为数据集 A 和数据集 B 中的微表情数据在四个类别上的分布情况。



图 2.11 CASME2 微表情采集方式，摘自[33]

从表 2.2 中可以看出，CASME 数据集上的数据分类非常不平均，其中数据集 A 上类别 Others 占了总数的 80% 多，数据集 B 上的类别 Negative 和 Others 两类的

个数也比 Positive 和 Surprise 多得多。事实上, CASME 数据集上极不均匀的分布结果会对微表情的训练造成负面影响。

CASME2 数据集是由 Yan^[33]等人在 CASME 数据集的基础上建立的又一个自发的微表情数据集。如图 2.11 所示,在 CASME2 数据集的采集过程中, Yan 等人提供了光照条件更好,速率更高的数码摄像机,对被试者的实验数据进行采集。CASME2 最终选取了 26 个被试者共 247 个微表情序列作为有效数据,并对其进行分类编码。这些微表情序列采集自分辨率为 640×480 , 帧率为 200fps 的数码相机。同样的, CASME2 上的数据集被分为了 4 大类: Positive, Negative, Surprise 以及 Others。在对微表情的分类方法上,除了根据 FACS 和被试者反馈以外,研究者通过邀请两位微表情学者独立地对这些采集到的视频序列进行分类,并对两种结果进行整合而得到最终的微表情分类结果。因此,相比于 CASME 数据集而言, CASME2 的分类结果可信度更高。

表 2.3 CASME2 上的微表情分类结果

数据集	Positive	Negative	Surprise	Others
CASME	33	60	25	129

如表 2.3 所示,为 CASME2 中的微表情数据在四个类别上的分布情况。从表 2.3 中可以看出, CASME2 存在着与 CASME 一样的数据分布不平均的问题。尤其是其中的类别 Others 类别,占了总数的一半还多。于此相对的是 Surprise 类别,所占的个数过少。这样的极不平均分布情况,同样会对为微表情的训练造成影响。

SMIC 数据集是由 Zhao^[20]等人在研究使用 LBP-TOP 算子作为微表情特征的时候,建立的一个用于测试算法效果的数据集。SMIC 是世界上第一个采集自被试者自发产生的微表情。在微表情采集的过程中,被试者被安排在一个布置得如同审讯室一样的房间内,通过给他们观看各种极端情况的视频序列,对被试者的感官产生冲击,从而获得被试者在感官受到冲击的情况下的面部特征。同时为了获得微表情,而避免产生普通的面部表情,在实验中,被试者被要求尽量压抑自己的情感,使得实验者可以抓取被试者在压抑自己真实情感过程中伴随而生的微表情特征。事实上, CASME 和 CASME2 这两个之后建立的数据库在数据采集的过程中也借鉴了 Zhao 等人的方法。

最终 Zhao 等人从录制的视频序列中选取了 6 个被试者共 152 个视频序列(其中,微表情和非微表情的视频序列各占 76 个),整合到 SMIC 数据集中。这些视频数据采集自分辨率为 640×480 , 帧率为 100fps 的数码相机。除了对微表情和表情进行分类外, Zhao 等人选出 76 个微表情数据中的 35 个数据,并使用 Positive

(17个)和 Negative (18个) 进行分类。最后,他们使用 LBP-TOP 特征对这两种分类结果进行实验评估。

根据这三个微表情数据集以及它们的分类结果,我们将对我们提出的基于运动主方向的微表情特征进行简单的实验评估,并对评估结果进行分析。

2.4.3 在 CASME 上的测试结果和分析

据上文所述, CASME 是由两个采集自不同实验环境的数据集组合而成的,为了加以区别,我们对这两个数据集分别进行了评估实验。

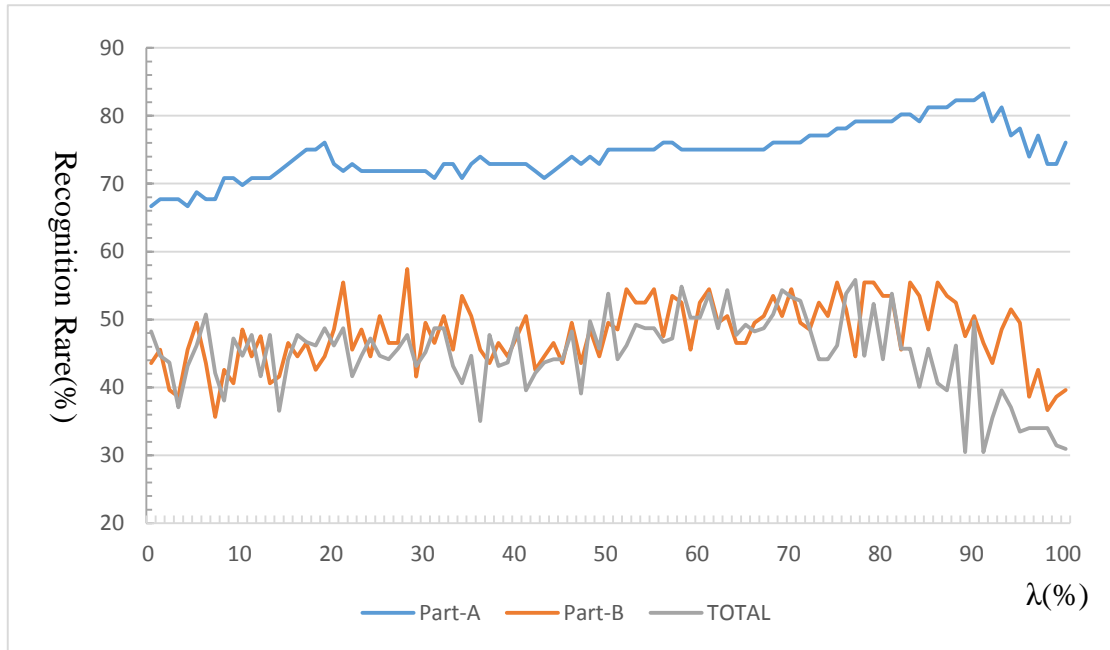


图 2.12 CASME 数据集上的识别率折线图

如图 2.12 所示,对于式 (2-18),我们以 1%的间隔在区域 $[0, 100\%]$ 上对参数 λ 进行采样,并使用 SVM 分类器对采样后的特征向量 $\bar{\Psi} = (\lambda P, (1 - \lambda)Q)$ 进行训练和分类。如图 2.12 所示,三条折线都表示了数据集 CASME 上的识别率,其中蓝色折线代表微表情特征在数据集 A 上的识别率;黄色折线代表了微表情特征在数据集 B 上的识别率,灰色折线代表了微表情特征在整个 CASME 数据集上的识别率。

从折线图的走向可以看出,随着参数 λ 取值的增大,三条折线都呈现上升的趋势,并在 $\lambda = 80\% \sim 90\%$ 左右达到峰值,然后逐渐递减。通过这张折线图的走向,我们可以知道,在 CASME 数据集的整体趋势上,当 λ 取值范围在 $80\% \sim 90\%$ 左右的时候,微表情特征取得最好的识别率。其中,在数据集 A 上,当参数 $\lambda = 91\%$

时，获得最高的识别率为 83.33%；在数据集 B 上，当参数 $\lambda = 28\%$ 时，获得最高的识别率 57.43%，当参数 $\lambda = 83\%$ 时，获得识别率的第二峰值 55.45%；在整个数据集上，当参数 $\lambda = 77\%$ 时，获得最高的识别率 55.84 %。

表 2.4 CASME 上的混淆矩阵

数据集	类别	预测结果			
		Positive	Negative	Surprise	Others
Part-A $\lambda = 91\%$	Positive	1	0	0	5
	Negative	1	0	0	4
	Surprise	0	0	3	4
	Others	1	0	1	76
		Positive	Negative	Surprise	Others
Part-B $\lambda = 83\%$	Positive	0	1	1	3
	Negative	0	32	4	13
	Surprise	0	8	5	1
	Others	0	13	1	19
		Positive	Negative	Surprise	Others
Total $\lambda = 77\%$	Positive	3	1	2	5
	Negative	1	30	3	20
	Surprise	2	5	9	5
	Others	7	31	5	68

为了获得 CASME 上四大类别之间的关系，除了给出识别率折线图之外，我们分别给出数据集 A, 数据集 B 和整个数据集在识别率达到峰值情况下的混淆矩阵，如表 2.4 所示。

从数据集 A 的混淆矩阵中可以看出，由于数据集 A 中 Others 类别的样本数超过总样本数的 80%，使得即使是所有类别都被识别为“Others”，其获得的准确率也能超过 80%，因此，数据集 A 上的识别结果并不能代表微表情特征的好坏。而从数据集 B 的混淆矩阵上可以看出，由于其中 Positive 样本过少，对数据集 B 的四分类问题变成了三分类问题，其中没有任何一个微表情数据被识别为了 Positive 类别。因此，类别 Positive 在数据集 B 上呈现出无效属性。在 A 和 B 的混合数据集上，由于“Positive”和“Surprise”类别训练样本的增加，增强了这两个类别的区分度，使得这两个类别的识别率都有所上升。

通过对 CASME 数据集上的测试结果，我们了解到了使用基于运动主方向的

微表情特征在 CASME 上的效果，并可以使用使得数据集识别率达到峰值的参数 λ 的取值范围，来表征一个微表情数据集的性质。同时通过分析数据集在峰值附近的混淆矩阵，我们发现由于 CASME 四大类样本分布不均匀而导致 SVM 训练结果的差异，同时这种差异也可以用于表征一个微表情数据集的质量。

2.4.4 在 CASME2 上的测试结果和分析

CASME2 数据集是在 CASME 的基础上通过改进数据采集和分类方式制作完成的。相比于 CASME 而言，CASME2 数据集中的人脸有效区域更大。但正是由于在数据集建立时候对这个因素过于重视，使得其中很多视频序列的人脸区域并没有完全在摄像机的镜头之内，或者人脸边缘与图像边缘非常小。这样采集得到的人脸图像，直接导致了 CLM 搜索算法失效。如图 2.13 所示，人脸的下巴位置并没有完全覆盖在镜头之内，CLM 搜索算法无法在这样的图像上获取人脸的特征点。基于以上原因，我们在这个数据集中选取了 173 个拥有比较好的 CLM 特征点的数据样本进行实验。这 173 个样本在四种微表情分类上与原数据集有着相似的分布，其中，Positive 含有 20 个样本，Negative 包含 46 个样本，surprise 包含 20 个样本，Others 包含 87 个数据样本。



图 2.13 CASME2 数据集中的无效人脸

如图 2.14 所示，是我们以 1% 为间隔在区域 $[0, 100\%]$ 上对参数 λ 进行采样，获得的微表情运动特征在 CASME2 上的识别率折线图。

从折线图的走向可以看出，随着参数 λ 取值的增大，识别率折线图在总体上呈现平稳上升的趋势，然后在参数 $\lambda = 60\% \sim 70\%$ 左右达到峰值，随后平稳递减到 $\lambda =$

80%附近,之后开始迅速下降。其中当参数 $\lambda = 64\%$ 时,微表情特征获得最高的识别率 67.63 %。

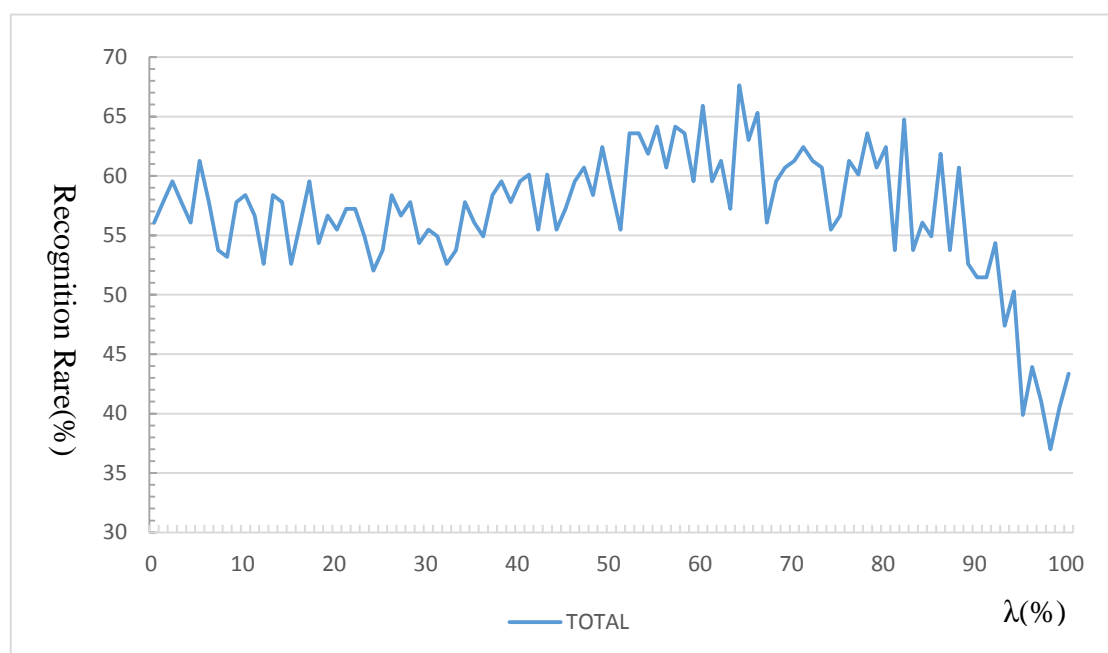


图 2.14 CASME2 数据集上的识别率随曲线

与 CASME 数据集相比, CASME2 数据集的识别率折线在参数 $\lambda = 60\% \sim 70\%$ 之间获得峰值,这意味着在 CASME2 中,方向特征所提供的分类信息比 CASME 数据集更多。另外,在横坐标区间 $[0, 10\%]$ 上,即微表情强度特征可以忽略的情况下,仅仅凭借微表情的方向特征其识别率就可以接近 60%;而在横坐标区间 $[90\%, 100\%]$ 上,由于忽略方向特征的关系,使得微表情的识别率迅速下降,这两个事实明确地告诉了我们微表情方向特征的在识别实验中的重要性。事实上, CASME 上的识别率折线图也存在这种效应,只是这种效应没有 CASME2 实验图像明显。

表 2.5 CASME2 数据集上的混淆矩阵

数据集	类别	预测结果			
		Positive	Negative	Surprise	Others
Total $\lambda = 64\%$	Positive	11	1	2	6
	Negative	3	26	0	17
	Surprise	1	0	15	4
	Others	3	17	2	65

同样,为了评估 CASME2 上的 173 个微表情数据在四种微表情分类上的的区

分程度，我们给出在识别率达到峰值时候的混淆矩阵，如表 2.5 所示。

从表 2.5 的混淆矩阵中可以看出，CASME2 上的四种微表情之间的区分度要比 CASME 上要好的多，其中 Positive 和 Surprise 这两种微表情与其他类别的微表情区分度比较明显；而 Negative 和 Others 这两个类别区分度不是特别明显。

通过对 CASME2 数据集上的测试结果，我们发现了微表情运动特征中的方向特征在微表情类内区分中有着非常显著的作用。尤其是在折线图 2.14 上，在忽略方向特征的前提下，微表情识别率在 CASME2 上呈现迅速下降的特性，有力验证了这一结论。

2.4.5 在 SMIC 上的测试结果和分析

SMIC 是 Pfister^[20]等人为了研究使用 LBP-TOP 特征应用于微表情识别中而建立的一个微表情数据集。在他们的实验中，他们使用两种分类方式对 SMIC 进行分类：一种是 Micro/Notmicro 的分类方式，另一种是 Positive/Negative 分类方式。其中 Micro/Notmicro 是区分微表情和非微表情之间的数据集，Positive/Negative 是区分微表情类内关系的数据集。在本小节中，我们在这两种分类方式上分别使用 SVM 分类器对本文提出的微表情运动特征进行实验评估。

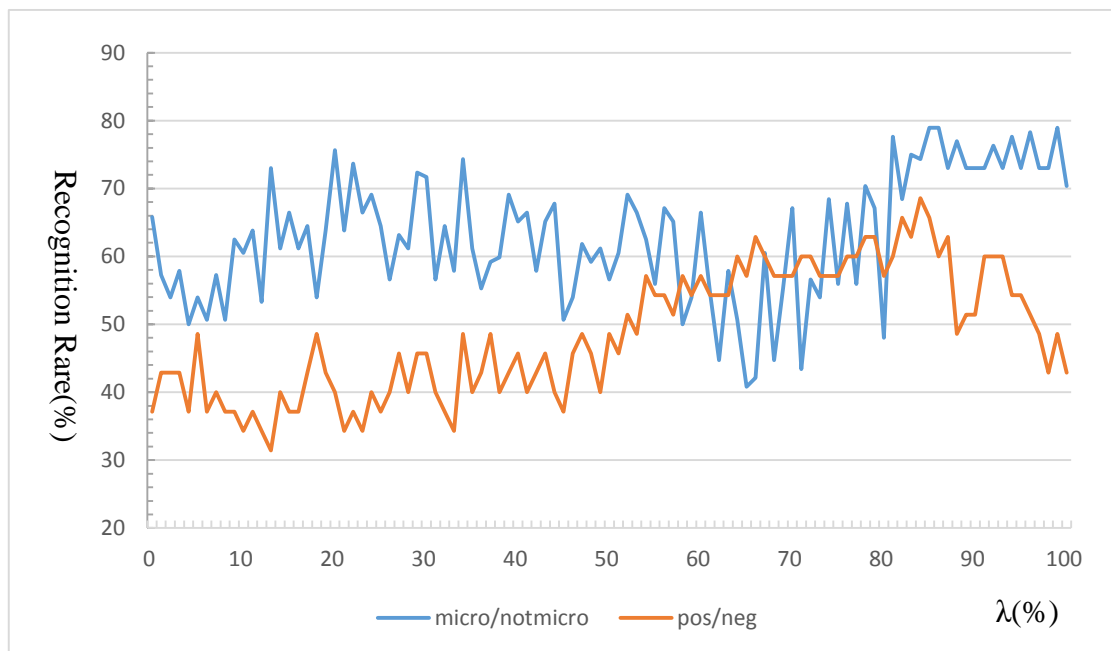


图 2.15 SMIC 数据集上的识别率曲线

如图 2.15 所示，是我们以 1% 为间隔在区域[0, 100%]上对参数 λ 进行采样，获得的在数据集 SMIC 上的两种分类结果。其中蓝色的折线段表示 Micro/Notmicro 数

据集的认可率，黄色的折线段表示 Positive/Negative 数据集的认可率。

从这两条折线图的走向可以看出，随着参数 λ 取值的增大，这两条折线图的走向呈现出相互交错的趋势。而这种趋势表明了微表情与非微表情的区分方式，与微表情内部的区分方式侧重的不同。对于 Micro/Notmicro 数据集而言，他的认可率折线图显示了当微表情特征的强度特征的权值越大，微表情和非微表情之间的区分度更高，亦即在区分微表情和非微表情之间，使用强度特征就可以获得相当令人满意的效果。而对 Positive/Negative 这个表征微表情内部区分的分类方式而言，它的认可率折线图与 CASME2 的认可率折线有着非常相似的走向，这种相似的走向也表明微表情特征中，方向特征在微表情的类内区分中有着非常显著的作用。

同时，从折线图上可以看出，当参数 $\lambda > 85\%$ 以后，Micro/Notmicro 一直保持着相当高的认可率，其中最高的认可率为 78.95 %；当参数 $\lambda = 84\%$ 时，即在数据集 Positive/Negative 获得峰值的区间内，它获得了最高的认可率 68.57 %。同样的，我们在这两个数据集取得峰值的情况下，分别给出他们的混淆矩阵，混淆矩阵如表 2.6 所示。

表 2.6 SMIC 数据集上的混淆矩阵

数据集	类别	预测结果	
Micro/Not micro $\lambda = 86\%$	Not micro	Notmicro	Micro
		59	17
	Micro	15	61
Positive/Negative $\lambda = 84\%$	Negative	Negative	Positive
		14	4
	Positive	7	10

通过对 SMIC 数据集上的测试结果，我们发现了微表情和非微表情的认可率折线和微表情之间的认可率折线图走向相迥异的特点。这一特点表明了，使用运动特征中的强度特征就可以很好的区分微表情和非微表情。这一结论，与微表情运动幅度小有着直接的关系。而 SMIC 数据集中微表示表情类内的 Positive/Negative 的认可率折线图，有着与 CASME2 中微表情认可率的折线图相一致的走向。这两个微表情数据集中相似的认可率走向，共同印证了微表情特征的方向特征对区分微表情有着非常重要的作用。

2.5 本章小结

在本章中我们提出了一种基于运动主方向特征的微表情特征的提取方法，并

使用 SVM 和 Leave-One-Video-Out 的训练和识别方法对这种特征提取方法在 CASME, CASME2 和 SMIC 上进行了评估测试。

这种方法通过将提取人脸 ROI 区域的方向直方图, 获取该区域的运动特征。同时, 使用加权平均的方法, 在降低微表情特征维数的同时, 能够保存微表情的特征信息。与传统的运用于表情和微表情识别中的 LBP-TOP 特征相比, 它的计算更加方便, 形成的特征向量维度更加少。

在最后的评估测试中, 我们在三个采集自被试者自发产生的微表情库上得出的识别率都要比人的裸眼识别率高。于此同时, 根据微表情的识别率曲线, 我们发现在微表情的类内识别中, 微表情的方向特征对微表情的识别率有着非常重要的影响, 而在微表情和非微表情的区分中, 仅仅使用微表情的强度特征就已经可以获得满意的识别率。

第3章 基于光流域矫正的微表情运动特征的提取方法

在上一章中，我们提出了一种基于运动主方向的微表情特征的提取方法，并分别在 CASME, CASME2 和 SMIC 数据集上对上述特征进行了评估测验。尽管测验的结果相较裸眼对微表情识别率有所提高，但其提高的还不是很明显。经过对微表情数据的进一步测验，我们发现在微表情的持续时间中，时常伴随着被试者很轻微的头部运动。然而，由于微表情运动幅度小，即使是幅度很小的头部运动，如旋转和平移，都会影响微表情特征的采集。如图 3.1 所示，为一个表示“紧张”的微表情在轻微的头部运动的情况下所采集得到的特征序列。在图 3.1 所示的微表情序列中，由于存在被试者头部的轻微右移（相对于观察者），使得提取到的微表情特征在整体呈现出右移的趋势。这样的结果导致了，即使在如被试者面部的下巴处没有任何运动的面部区域，使用第二章的特征提取方式也能够获得相当强度的微表情特征。

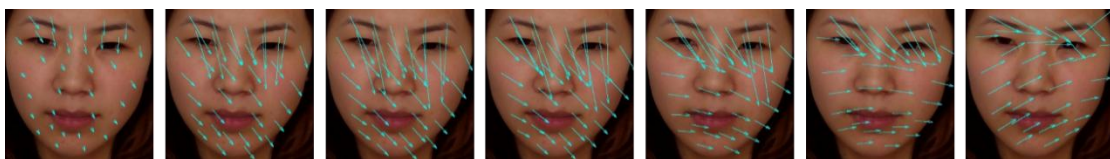


图 3.1 微表情“紧张”的运动特征序列

为了解决上述问题，本文提出了一种基于光流域矫正的微表情运动特征的提取方法。该方法在第二章提出的基于运动主方向的微表情特征提取的基础上，通过在光流域对微表情序列的运动特征使用仿射变换进行矫正，去除了被试者的头部运动对微表情特征造成的影响。

3.1 仿射变换简介

由于微表情持续时间短，运动幅度小，任何来自于被试者头部的微小干扰（如头部的移动和旋转）都可能影响微表情特征的提取。为了尽可能减小这种发生在微表情采集过程中的干扰，我们使用仿射变换对微表情的运动特征进行矫正。

仿射变换是图像领域中一种常用的图形变换方式，它表示了一个图像在二维平面中任意的拉升、扭曲和旋转变换的组合。如图 3.2 所示，其中图形 a 是原始图形，图形 b 是图形 a 通过仿射变换后所有可以获得的图形形状。

由于一个任意的仿射变换可以表示为该图像左乘一个矩阵再加上一个向量的

形式，所以在计算中，人们通常使用基于 2×3 矩阵的变换表示仿射变换，其定义如下：

$$A = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \quad B = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \quad T = [A \ B] \quad X = \begin{bmatrix} x \\ y \end{bmatrix} \quad X' = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3-1)$$

$$Y = AX + B \xleftrightarrow[X \text{ 扩展成 } X']{ } Y' = TX' \quad (3-2)$$

其中 T 表示仿射矩阵， X 表示原始图形， X' 表示 X 的扩展， Y 表示经仿射变换之后的图形， Y' 为 Y 在第三维上的扩展。由公式可以知道，图像的仿射变换可以使用仿射矩阵左乘原始图像取得。

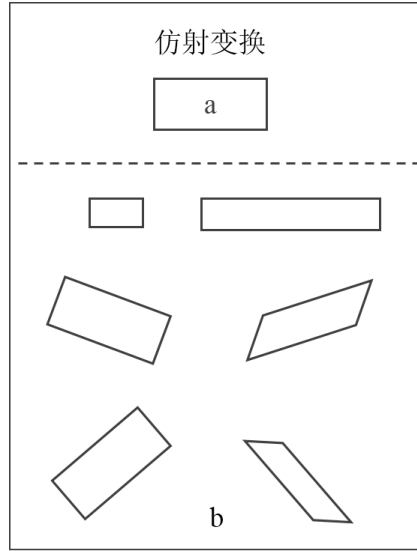


图 3.2 仿射变换示意图

3.2 基于仿射变换的微表情特征的矫正方法

通常在图像领域中，人们比较熟悉的操作是对原始图像进行仿射变换并得到新的图像，然后对新的图像进行下一步的操作。如 Pfister^[20]等人在使用 LBP-TOP 作为特征对微表情进行检测的时候，他们在数据预处理环节，就选择一个标准脸模板对图像中所有的人脸进行矫正，并对矫正后的图像进行下一步的特征提取工作。这是因为，相比于对图像的运动特征进行矫正，在图像域进行矫正显得更为直观。然而，由于微表情运动持续时间中，被试者头部运动的幅度非常小，即使是一个微表情序列发生的第一帧和最后一帧其运动的幅度也不过是几个像素点，甚至仅仅一个像素点而已。而基于图像域的矫正，在图像的保存过程中，存在着一个

像素点的截断误差,这种误差使得微表情数据在图像域的矫正操作并不明显。这种在图像域的操作,反而可能破坏图像的固有结构,影响对人脸面部特征的提取。

在本文中,为了体现这种差异,我们使用两种不同的方法对微表情的运动特征进行矫正。1) 基于图像域的矫正方法。基于图像域的矫正方法是首先将整个视频序列根据其第一帧进行仿射矫正,然后获取图像的运动特征。2) 基于光流域的矫正。基于光流域的矫正方法是首先获取整个视频序列的运动特征,然后对运动特征进行仿射矫正。

3.2.1 基于图像域的微表情运动特征的矫正方法

在本文中,我们首先讲述使用图像域矫正的方法对微表情的运动特征进行矫正的方法。根据2.1节所述,对于一个给定的视频序列 $\Gamma = \{F_1, F_2, \dots, F_n\}$ 中的任意一帧,我们可以使用CLM搜索算法获取图像上表征人脸位置的CLM特征点。在这里,我们将66个CLM特征点中的位于人脸轮廓上的17个特征点作为仿射矫正的参照点,如图3.3所示。



图 3.3 位于人脸轮廓上的 17 个 CLM 参照点

对于给定的视频序列上的第 i 帧,我们首先根据仿射变换规则,将这17个参照点在第三维上进行扩展,并使用矩阵 P^i 表示,如式(3-3)所示。

$$P^i = \begin{bmatrix} x_1^i & \dots & x_{17}^i \\ y_1^i & \dots & y_{17}^i \\ 1 & \dots & 1 \end{bmatrix} \quad (3-3)$$

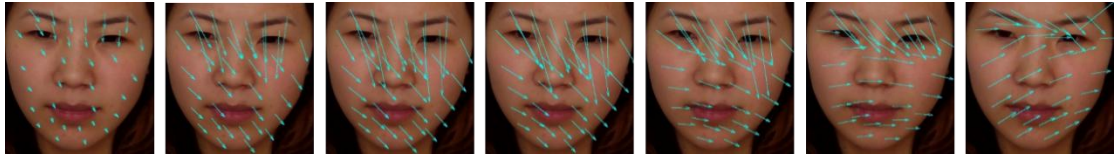
我们使用符号 T^i 表示第 i 帧的仿射变换矩阵,并使用如式(3-4)所示的优化函数来求解这个仿射矩阵:

$$\min \|T^i P^i - P^1\|_2 \quad (3-4)$$

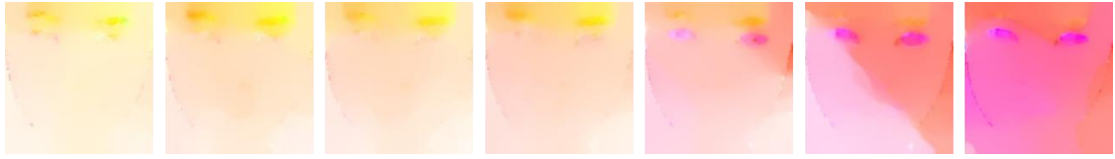
其中 P^1 表示在第一帧上的参照点的位置。优化函数表征了从第 i 帧向第一帧矫正的方法。由于在式 (3-4) 中, P^i 是一个 17×2 的数组, 而 T^i 是一个 2×3 的数组。



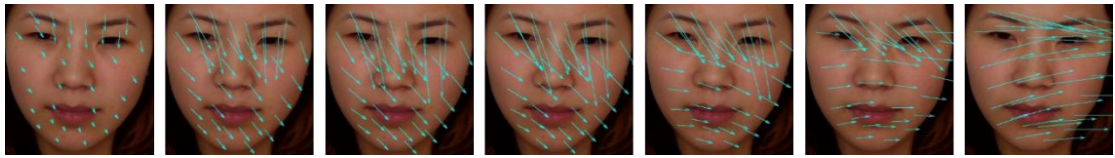
(a)



(b)



(c)



(d)

图 3.4 使用图像域矫正的光流图和视频序列的运动特征。(a)视频序列的光流图; (b)原始序列的运动特征; (c)图像域矫正后的光流图; (d)图像域校正后视频序列的光流图。

为了获取式 (3-4) 的近似解, 在本文计算中, 我们使用矩阵的 SVD 分解对这个优化函数进行求解, 求解过程如下:

$$\min \|T^i P^i - P^1\|_2 \xrightarrow{\text{ideal}} T^i P^i = P^1 \quad (3.5)$$

令矩阵 P^i 的 SVD 分解为:

$$P^i = U \Sigma V^*, \quad (P^i)^+ = V \Sigma^+ U^* \quad (3-6)$$

其中 $(P^i)^+$ 是矩阵 P^i 的伪逆，那么可以求得矩阵 T^i 的形式：

$$T^i \approx P^1 (P^i)^+ = P^1 V \Sigma^+ U^* \quad (3-7)$$

因此，对每一个给定视频序列 $\Gamma = \{F_1, F_2, \dots, F_n\}$ ，可以用图像左乘矩阵 T 的方式得到视频序列的变换图片。然后根据第二章的方法，计算这个视频序列的微表情特征。即得到通过图像域矫正之后的视频序列的运动特征。如图 3.4 所示，为一个表征“紧张”的微表情序列的原始序列和经图像域矫正后的光流图像和运动特征。

根据图 3.4 可以得知，通过图像域矫正的方法并不能解决因被试者轻微头动对微表情特征提取所造成的影响，反而因为图像域的截断误差，使得图像结构受到了破坏，图像的运动特征也出现了明显的断层，如图 3.4 (c) 所示。

3.2.2 基于光流域的微表情运动特征的矫正方法

上一小节介绍了从图像域对微表情运动特征进行矫正的方法。但是从图 3.3 的光流图像和运动特征图像表明了通过图像域对运动特征的矫正效果并不明显。其原因有两个：

- 1) 由于视频序列持续时间短，面部的运动幅度小，使得基于图像域的矫正效果并不明显。
- 2) 由于图像的仿射变换的截断误差和微表情的运动幅度处在一个量级，使得截断误差破坏了图像的原始结构，容易对后续的特征提取工作产生影响。

为了解决上述问题，我们提出了一种基于光流域的微表情运动特征的矫正方法。该方法通过对图像的运动特征进行矫正，避免了直接操作原始图像所带来的截断误差。该方法如下：

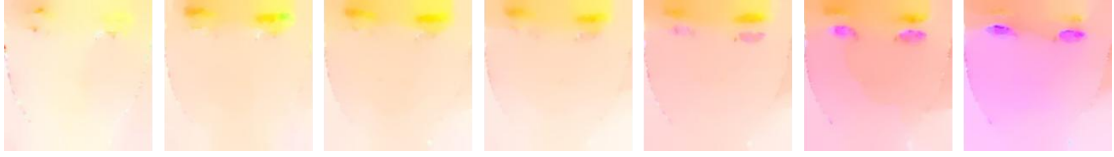
对于一个给定的视频序列 $\Gamma = \{F_1, F_2, \dots, F_n\}$ ，我们首先使用 Brox^[29] 算法获取从视频的第一帧到第 i 帧图像的运动特征 M^i ，其运动特征可以用如下矩阵表示，如式 (3-8) 所示。

$$M^i = \begin{bmatrix} V_{0,0}^i & \cdots & V_{w-1,0}^i \\ \vdots & \ddots & \vdots \\ V_{0,h-1}^i & \cdots & V_{w-1,h-1}^i \end{bmatrix} \quad (3-8)$$

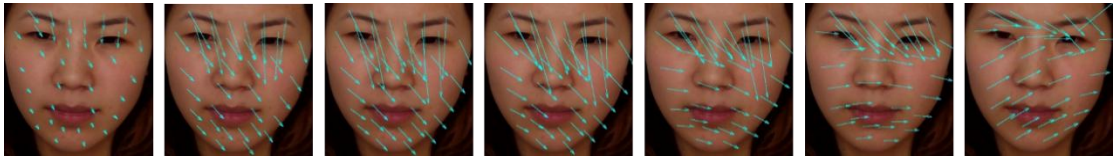
其中 $V_{x,y}^i = [u_{x,y}^i, v_{x,y}^i]^T$ 表示像素点 (x, y) 的运动特征。

然后我们使用 CLM 搜索算法计算视频序列的第一帧的 66 个特征点，并将其中心位于人脸轮廓上的 17 个特征点作为参照点。将这 17 个参照点在第三维上进行扩展之后，我们得到了它的矩阵表达形式：

$$P^1 = \begin{bmatrix} x_1^1 & \dots & x_{17}^1 \\ y_1^1 & \dots & y_{17}^1 \\ 1 & \dots & 1 \end{bmatrix} \quad (3-9)$$



(a)



(b)



(c)



(d)

图 3.5 使用图像域矫正的光流图和视频序列的运动特征。(a)视频序列的光流图；(b)原始序列的运动特征；(c)光流域矫正后的光流图；(d)光流域校正后视频序列的光流图。

那么视频序列中第 i 帧上的 17 个参照点的位置可以通过 P^1 和以及第 1 帧到第 i 帧的光流 M^i 给出。令 dP^i 为这 17 个参照点在第 i 帧上的光流，如式 (3-10) 所示：

$$dP^i = \begin{bmatrix} u_{x_1^1, y_1^1}^i & \dots & u_{x_{17}^1, y_{17}^1}^i \\ v_{x_1^1, y_1^1}^i & \dots & v_{x_{17}^1, y_{17}^1}^i \\ 0 & \dots & 0 \end{bmatrix} \quad (3-10)$$

则有：

$$P^i = P^1 + dP^i \quad (3-11)$$

其中 P^i 为第 i 帧上参照点的位置矩阵。通过式(3-4)~式(3-7)可以得到第 i 帧的仿射变换矩阵 T^i 。

使用获得的仿射变换矩阵 T^i ，我们将视频序列中每一帧的运动特征进行矫正，并获得的矫正之后的运动特征 \tilde{M}^i ：

$$\tilde{M}^i = \begin{bmatrix} \tilde{V}_{0,0}^i & \cdots & \tilde{V}_{w-1,0}^i \\ \vdots & \ddots & \vdots \\ \tilde{V}_{0,h-1}^i & \cdots & \tilde{V}_{w-1,h-1}^i \end{bmatrix} \quad (3-12)$$

其中 $\tilde{V}_{x,y}^i = [\tilde{u}_{x,y}^i, \tilde{v}_{x,y}^i]^T$ 为矫正后的运动特征，它使用变换矩阵 T^i 得到，其计算方式如下：

令符号 $p_{x,y}^i = [x, y, 1]^T$ 为图像上任意一点的扩展，符号 $dp_{x,y}^i = [u_{x,y}^i, v_{x,y}^i, 0]^T$ 为点 (x, y) 的运动特征的扩展。那么位于像素点 (x, y) 的矫正后的运动特征 $\tilde{V}_{x,y}^i$ 可以表示为：

$$\tilde{V}_{x,y}^i = T^i(p_{x,y}^i + dp_{x,y}^i) - V_{x,y}^i \quad (3-13)$$



图 3.6 代表“紧张”的微表情通过加权平均后得到的特征值

根据获得的经光流域矫正后的运动特征 \tilde{M}^i ，我们可以通过第二章的方法，获取视频序列的微表情特征。如图 3.5 所示，为一个表征“紧张”的微表情序列的原始序列和经光流域矫正后的光流图像和运动特征。

通过 3.5 (c) 所示的光流图像，能够很明显的看出微表情序列在眉弓上有强烈的运动特征，而图 3.5 (d) 则通过 ROI 的运动特征将这些地方的运动强度量化了出来。通过对比图 3.4 和图 3.5，我们可以得到如下结论：基于光流域的微表情矫正方法比基于图像域的矫正方法，在去除因被试者轻微头部运动带来的噪声方面，

有着更显著的效果。图 3.6 表示了图 3.5 的微表情序列通过光流域矫正后的加权平均特征。可以看出，表示“紧张”的微表情序列在使用基于光流域的矫正之后，基本上去除了因被试者头部运动带来的影响。如 3.6 图中下巴和脸颊的位置，几乎看不到有面部肌肉的运动痕迹。

在本小节的最后，我们将第二章和第三章的算法进行整合，给出基于光流域矫正的运动特征的提取方法的算法小结，如算法 1 所示：

算法 1 基于运动主方向的微表情特征提取方法。给定微表情数据集 $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ ，其中 Ω 表示整个微表情数据集， ω_i 表示其中第 i 个微表情视频序列， n 表示微表情数据集的大小。

```

1:   I. 数据预处理
2:   for each  $\omega_i, i \in n$  do
3:       for each  $f_{i,j} \in \omega_i$ ，其中  $f_{i,j}$  表示视频帧 do
4:           使用CLM搜索算法获取图像中表征人脸位置的66个特征点；
5:           根据CLM特征点，分离有效的人脸区域；
6:           根据FACS将人脸分成36个ROI区域；
7:       end for
8:   end for
9:   II. 特征提取
10:  for each  $\omega_i, i \in n$ , do
11:      for each  $f_{i,j} \in \omega_i$ ， do
12:          使用稠密光流算法，获取图像的运动特征；
13:          使用光流域矫正的方法，对运动特征进行矫正；
14:          使用矫正的图像的运动特征，获取基于运动主方向的ROI区域的运动特征
15:      end for
16:      使用加权平均的方法，对视频的特征序列进行归一化；
17:      对微表情特征的长度特征进行归一化；
18:      使用参数 $\lambda$ ，改写微表情特征；
19:  end for

```

3.3 测试评估

为了对本文提出的基于光流域矫正的微表情运动特征进行测试和评估，我们依然选择在 CASME, CASME2 和 SMIC 这三个数据集上对矫正后的微表情特征进

行简单的测试和分析。

3.3.1 测试方法

如 2.4.1 节的实验方法一致，对于一个矫正的微表情数据集，我们使用式 (2-18) 表征矫正后的微表情特征中强度特征和方向特征的权重，并使用 POLY 核函数的 SVM 分类器对样本进行训练和分类。其中为了保持实验方法的一致，POLY 核函数的参数仍旧固定为： $\gamma = 0.1, degree = 4, coef = 1$ 。

在训练和测试过程中，我们依然使用 Leave-One-Video-Out 的实验方法，将式 (2-18) 所示的矫正后的特征向量在三个数据集的原始帧率上进行评估。为了将基于图像域的矫正方法和基于光流域的矫正方法进行比较，在实验结果的展示里，我们将同时展示通过这两种矫正方法的微表情识别率。同时为了观察每个数据集中各个类别之间的区分度，我们将分别给出他们获得最高的识别率时的混淆矩阵，用以表征数据集中各个类别之间的区分度，并对这些结果加以分析。

3.3.2 测试数据

为了将实验结果与未矫正的微表情特征进行比较，在本章使用的实验数据如 2.4.2 节的实验数据保持一致，并会对相同数据集上的不同的识别率进行简单的比较和分析。

3.3.3 在 CASME 上的测试结果和分析

为了与 2.4.3 小节的未矫正的微表情运动特征进行比较，我们同样在 CASME 的数据集 A 和数据集 B 上完成了评估实验。

如图 3.7 所示，是我们在数据集 CASME 上分别用基于光流域矫正和图像域矫正的微表情特征作为训练样本和测试样本所得到的识别率折线图。其中，实线表示的是基于光流域矫正的微表情识别率，虚线表示的是基于图像域矫正的微表情识别率，他们的横坐标间隔都为 $\Delta\lambda = 1\%$ 。

从图 3.7 可以看出，在数据集 CASME 的每个数据集（A，B，Total）上，基于光流域矫正的微表情的识别率在折线图的任何一个时刻都要比基于图像域矫正的识别率高。而且，就整个折线图的整个变化趋势来看，经过光流域矫正的微表情特征的折线图都要比未矫正和使用图像域矫正方法的折线图平稳。这种平稳的性质，表征了 SVM 分类器对参数 λ 在小范围内扰动的不敏感，也即表明了使用光流域矫正的方法使得微表情的识别率越接近真实值。

从实线的折线图的走向可以看出，随着参数 λ 的增大，三条折线呈现平稳上升

的过程，并在 $\lambda = 80\% \sim 90\%$ 左右达到峰值。尤其是对于数据集 B 而言，这种识别率递增的现象更为明显，这种现象意味在微表情识别中，强度和方向的比例越接近最优比例，所获得的识别率越高。而当参数 $\lambda > 90\%$ 之后，识别率曲线呈直线下降趋势，这种结果意味着，微表情的方向特征对于它的识别率有着比强度特征更为显著的作用。其中，在数据集 A 上，当参数 $\lambda = 63\%$ 时，基于光流域矫正的微表情特征，在这个点获得最高的识别率 85.42%；在数据集 B 上，当参数 $\lambda = 88\%$ 时，其获得最高的识别率 69.31%；在整个数据集上，当参数 $\lambda = 80\%$ 时，获得最高的识别率 65.48 %。这些数值都比第二章使用未矫正的光流域矫正的实验结果好。

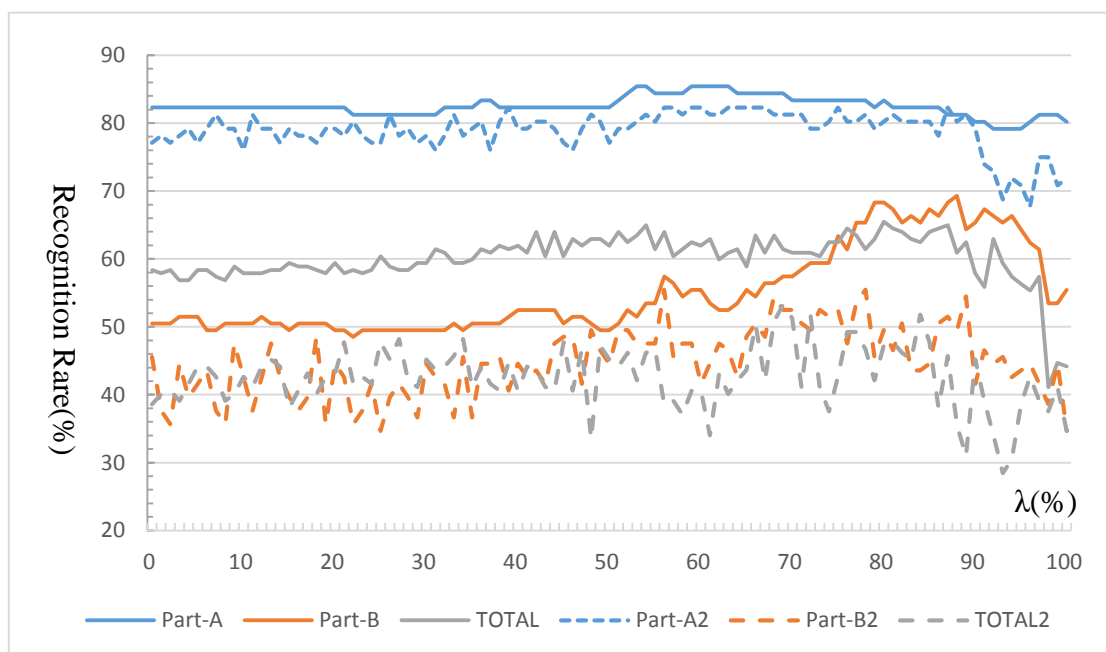


图 3.7 CASME 数据集上的识别率随曲线

为了给出数据集中，四种微表情分类之间的关系，我们给出基于光流域矫正的实验结果的混淆矩阵，如表 3.1 所示。表 3.1 所示的混淆矩阵与表 2.4 所示的混淆矩阵有相似的关系。其中数据集 A 上的主要原因依旧是因为类别 Others 占的样本数过多，使得最终的结果并没有太大的参考价值。数据集 B 和数据集 Total 这两个数据集除了在 Negative 和 Others 两个类别之间体现了比未矫正的微表情特征更加明显的分类之外，也存在着这两个类别过多的问题。不过总体上，使用光流域矫正后的微表情特征之后，CASME 上的 4 个微表情类别的区分度更加明显了。

通过对 CASME 上的测试结果，我们了解到了使用光流域矫正的微表情的运动特征的识别效果要比未矫正和使用图像域矫正方法的效果好得多。同时使用光流域矫正的方法，使得识别率折线图随参数 λ 的变化变得更加平稳，这表明了 SVM

对参数 λ 的微小扰动不敏感，这一现象从另一方面验证了使用光流域矫正的方法使得微表情的识别率越接近真实值。另外，图像上三条曲线在参数 λ 越接近 1，下降的越迅速的现象，支持了 2.4.4 的结论，即微表情的方向特征在微表情识别过程中有着显著的作用。

表 3.1 光流域矫正实验的混淆矩阵

数据集	类别	预测结果			
		Positive	Negative	Surprise	Others
Part-A $\lambda = 63\%$	Positive	1	0	0	5
	Negative	1	0	0	4
	Surprise	0	0	6	1
	Others	2	0	1	75
Part-B $\lambda = 88\%$	Positive	0	0	3	2
	Negative	2	42	3	4
	Surprise	1	5	6	2
	Others	1	9	1	22
Total $\lambda = 80\%$	Positive	1	4	2	4
	Negative	3	29	3	19
	Surprise	1	2	12	6
	Others	6	16	2	87

3.3.4 在 CASME2 上的测试结果和分析

为了在 CASME2 上对矫正后的微表情特征和未矫正的微表情特征进行比较，我们使用了与 2.4.4 相同的数据集进行实验。在实验结果中，我们同时给出了基于图像域矫正和基于光流域矫正的折线图，用于对这两种矫正方式进行比较。

如图 3.8 所示为矫正后的微表情特征在 CASME2 上的实验结果，其中横坐标为式(2-18)中的 λ ，表示强度和方向的权值，纵坐标表示的是识别率。图像上，虚线表示的是使用图像域矫正的识别率折线图，实线表示使用光流域矫正的识别率折线图。从图 3.8 上可以清楚的看出，在数据集 CASME2 上，通过光流域矫正的方法对于任何一个参数 λ 取值的识别率都比基于图像域矫正的识别率高 10% 左右。同时，基于光流域矫正的折线图表现出了更好的稳定性，这个性质与 3.3.3 小节中的

结果一致，都表明了使用光流域矫正的方法更接近理想的微表情特征。

同时，随着参数 λ 的上升，基于光流域矫正的折线图在整体上呈现出平稳上升的趋势，并在区间[70%,90%]之间达到峰值，在参数 λ 超过 90%之后，识别率成迅速下降趋势，这个现象也和 CASME 上的折线图特征一致，表明了微表情的方向特征比强度特征对微表情识别有着更为显著的作用。其中，在基于光流域矫正的折线图像上，当参数 $\lambda = 73\%$ 时，获得最高的识别率为 71.68%，比未使用光流域矫正的实验结果好。

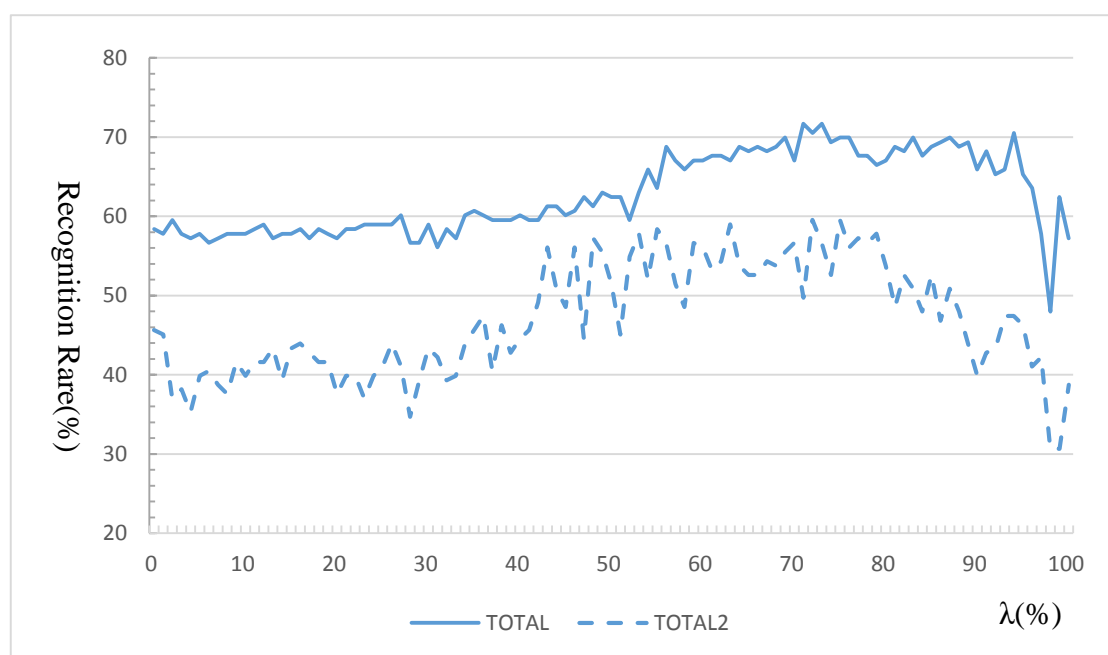


图 3.8 CASME2 数据集上的识别率随曲线

同样的为了获得 CASME2 数据集上各个分类之间的区分度，我们给出基于光流域矫正的实验结果的混淆矩阵，如表 3.2 所示。从 CASME2 的混淆矩阵可以看出，通过光流域矫正，CASME2 中的四种微表情有了更加明显的区分度。

表 3.2 CASME2 数据集上的混淆矩阵

数据集	类别	预测结果			
		Positive	Negative	Surprise	Others
$\lambda = 64\%$	Positive	12	1	1	6
	Negative	1	26	0	19
	Surprise	0	0	16	4
	Others	2	14	1	70

通过对基于光流域矫正的微表情特征提取方法在 CASME2 数据集上的测试结果，我们加强了这种结论：1) 基于光流域矫正的方法比未使用矫正和使用图像域矫正的方法的效果要好得多；2) 使用光流域矫正使得识别率曲线更加平稳，是获得的微表情特征更加接近真实的微表情特征；3) 折线图在参数 λ 接近 100%的时候迅速下降，体现了微表情识别中，方向特征的有着显著的重要性。

3.3.5 在 SMIC 上的测试结果和分析

同样的我们在 SMIC 上也对基于光流域矫正和基于图像域矫正的微表情特征进行简单的测试和分析。

如图 3.9 所示为在数据集 SMIC 上使用两种矫正方式对微表情特征进行矫正后的实验结果，其中虚线表示使用图像域矫正的方法得到的识别率，实线表示使用光流域矫正的方法得到的识别率，蓝色的折线段表示 Micro/Notmicro 数据集的识别率，黄色的折线段表示 Positive/Negative 数据集的识别率。从图 3.9 所示的结果来看，我们依然能得到如下的结论：1) 基于光流域的矫正方法比基于未矫正以及图像域矫正的方法效果好得多，2) 基于光流域矫正的识别率比其他两种方法有着更好的稳定性。

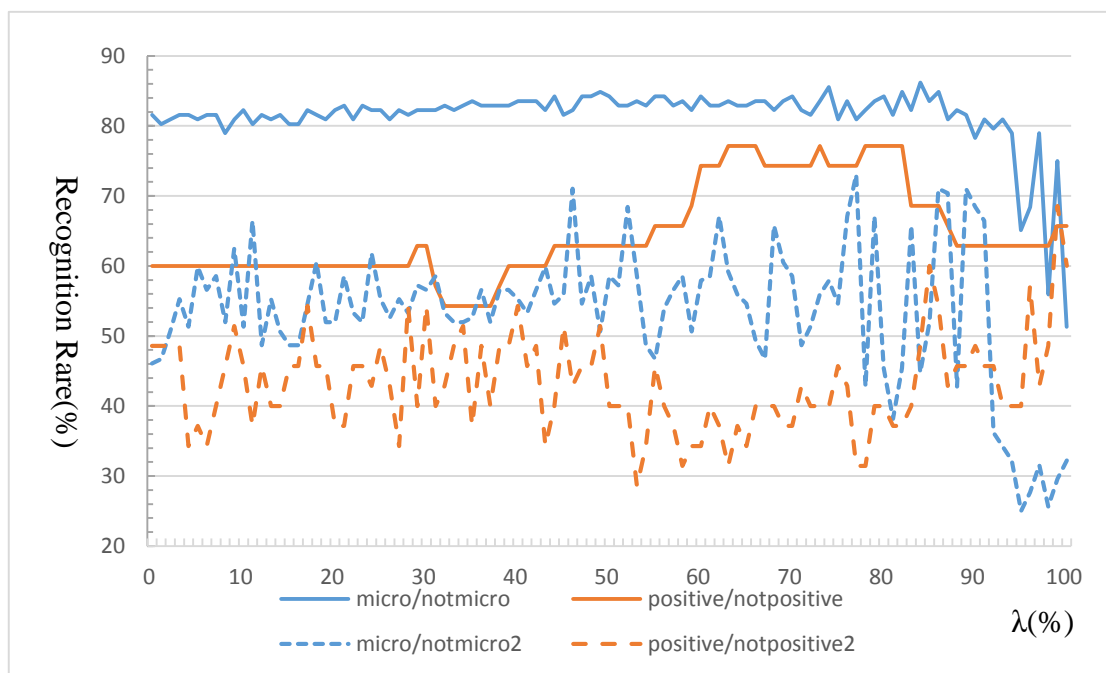


图 3.9 SMIC 数据集上的识别率曲线

根据图 3.9，我们可以得知，随着参数 λ 的增加，基于光流域矫正的方法总体上呈现稳定上升的趋势，并在区域[60%，80%]之间达到峰值，当参数 $\lambda > 80\%$ 之后，

识别率变得不稳定,并开始迅速下降。通过这个现象,我们可以入之前一样得出如下结论:1)微表情特征中的方向特征在微表情识别中有着有着比强度特征更为显著的作用。2)当参数 $\lambda = 74\%$ 时,在“Micro/Notmicro”数据集上获得最高的识别率为85.53%;当参数 $\lambda = 73\%$ 时,获得最高的识别率为77.13%。这两个值都比未使用光流域矫正的实验结果好,最后我们给出此时的混淆矩阵,如表3.3所示。

表 3.3 SMIC 数据集上的混淆矩阵

数据集	类别	预测结果	
Micro/Not micro $\lambda = 74\%$		Notmicro	Micro
	Not micro	67	9
	Micro	13	63
Positive/Negative $\lambda = 73\%$		Negative	Positive
	Negative	15	3
	Positive	5	12

从表3.3中的数据,我们可以得到同2.4.5小节更为明显的结论:微表情和非微表情之间有着比微表情的类内更加明显的区分度。

然而同为SMIC数据集上的识别率,图2.15和图3.9上的微表情识别率折线走向却不一致。其中,在2.4.5小节的实验中,类别Micro/Notmicro的识别率折线图随着参数 λ 的增大而增大,而在本小节中,这个用以界定微表情和非微表情类别的识别率走向却和微表情类内的走向有着更强的一致性。这表明了,在使用更接近微表情的真实特征作为训练样本后,微表情的方向特征在界定微表情和非微表情中也存在着非常显著的作用。

3.4 本章小结

在本章中我们提出了一种基于光流域矫正的微表情运动特征的提取方法,并使用SVM和Leave-One-Video-Out的训练和识别方法对这种特征提取方法在CASME, CASME2和SMIC上进行了评估测试。

这种方法通过选用视频序列中第一帧的人脸作为模板脸,第一帧人脸轮廓上的点作为模板点,通过光流域矫正的方法,对图像的运动特征进行矫正,继而通过第二章提取运动主方向特征的方法,获得矫正后的微表情运动特征。这种方法解决了因在采集过程中被试者头部的轻微运动而导致识别率低下的问题。

在最后的评估测试中,我们在三个采集自被试者自发产生的微表情库上,分别将基于光流域矫正的方法与未使用矫正和使用图像域矫正的方法进行比较。得出

了如下结论：1) 基于光流域矫正的特征提取方法，比未使用矫正和使用图像域矫正的方法要好。2) 微表情运动特征中的方向特征不管是在微表情类内的识别中，还是在微表情与非微表情的界定中都有着显著的作用。3) 基于光流域矫正的方法比未矫正和使用图像域矫正的方法有着更好的稳定性。

第4章 实验与结论

如第二章和第三章所述，本文提出了一种经光流域矫正的基于图像运动主方向的微表情特征的提取方法。这种方法通过使用方向直方图统计的方法，获取人脸 ROI 区域的主方向特征；通过加权平均的方法，将微表情特征归一化；通过光流域矫正的方法，消除了因被试者轻微头动带来的影响。

在此基础上，我们又给出了一种微表情识别算法，用以衡量微表情的强度特征和方向特征在微表情识别中的重要程度。通过第二章和第三章的测试方法，我们使用 Leave-One-Video-Out 的方法分别在三个微表情数据集的原始帧率上对本文提出的特征提取方法进行测试评估，并获得了较高的识别率。

在本章中，我们将对本文提出的微表情识别方法进行更加深入的实验测试，并与目前世界上在微表情识别方面效果最好的 LBP-TOP 特征进行比较。

4.1 LBP-TOP 特征简介

LBP-TOP (LBP-TOP(Local Binary Patterns on Three Orthogonal Planes) [21]是由 LBP[34]特征发展而来的，1996 年 Ojala 等人首先提出了 LBP 这种表征图像纹理特征的描述符，它通过计算图像上任意一个像素点与其周围的像素点之间的关系，而获得这个像素点的纹理特征。

作为一种常用的表示图像纹理特征的描述符，LBP 特征的运算非常简单。如图 4.1 所示，对于图像上任意一个像素 x ，以这个像素点为中心， R 为半径，取一个圆；在圆上等间隔的取 P 个点，用以描述该点的纹理特征；若这些点的灰度值比像素点 x 的大，那么这个位置就用 1 代替，反之用 0 代替；将着周围 P 个像素点的 01 值串成一个二进制串表示像素点 x 的纹理特征。

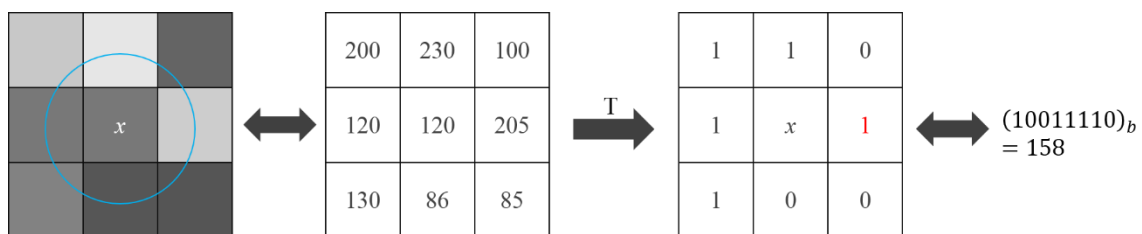


图 4.1 LBP 特征计算过程，其中 $R=1$ ， $P=8$

然而 LBP 特征只能表示二维的静态图像的纹理特征，为了表示三维图像序列

中的任意一个像素点在三维上的纹理特征,研究者将 LBP 运用到动态纹理特征上,并使用 VLBP 表示^[21]。如图 4.2 所示,表示的是 VLBP 特征的计算方式: 对于一个视频序列中的任意一个像素点 x , 以这个像素点为中心, R 为半径, 取一个圆; 在圆上等间隔的取 P 个点, 用以描述该点在同一帧上的纹理特征; 同时再从位于它之前 L 帧和之后 L 帧的图像上各取 $P+1$ 个点, 用于表示像素点 x 在时域上的纹理特征; 最后使用这 $3P+2$ 个点描述像素点 x 在视频序列中的纹理特征。同 LBP 特征计算一样, 若这 $3P+2$ 个点的灰度值比像素点 x 的大, 那么这个位置就用 1 代替, 反之用 0 代替; 将这周围 $3P+2$ 个像素点的 01 值串成一个二进制串表示像素点 x 的纹理特征。

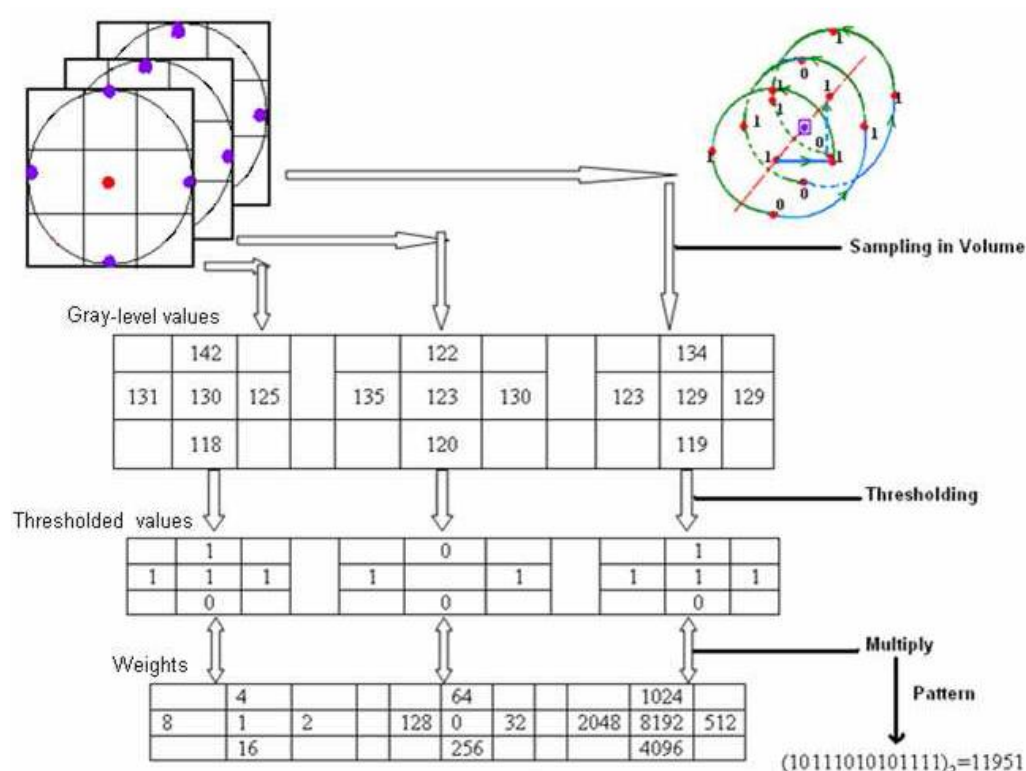


图 4.2 VLBP 描述符的计算方式, 其中 $L=1$, $R=1$, $P=4$, 摘自[21]

尽管使用 VLBP 能够非常清楚的表征像素点在视频序列上的纹理特征, 但是使用 VLBP 方法最终形成的特征向量的维度与 p 是成指数增长的。这就意味着, 如果想获得图像上更多的纹理信息, 那么我们会选择更大的参数 P 的值, 这就意味特征向量的维度越长, 计算越复杂。这一性质使得 VLBP 很难应用到实际的纹理识别中。为了解决特征向量的维度随着 p 增大的问题, Zhao 等人提出了 LBP-TOP 算子^[21]。LBP-TOP 算子采用 XY, XT, YT 三个在中心点垂直的平面来代替 VLBP 中三个平行的平面, 使得在选取相同的参数 p 的同时, 降低了纹理特征的维

度，LBP-TOP 特征的计算方式如图 4.3 所示。

LBP-TOP 特征是表情识别领域中非常经典的方法。2011 年，Pfister^[20]等人将 LBP-TOP 引入到微表情的检测和识别中，并在 SMIC 数据集上进行测试，取得了非常好的实验结果。

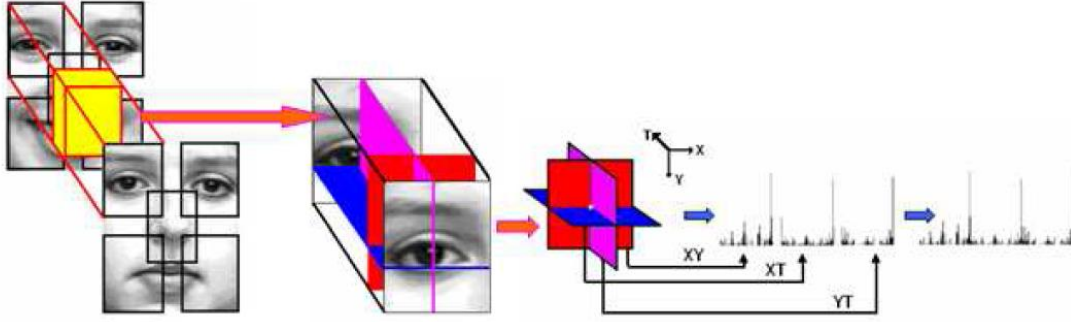


图 4.3 LBP-TOP 特征计算过程

之后，2014 年，Wang^[22]等人将 LBP-TOP 特征应用到张量独立的颜色空间（Tensor Independent Color Space, TICS），并与 LBP-TOP 特征与传统的 RGB 空间和灰度空间的实验结果进行比较。同样取得了不错的实验结果。

4.2 在原始帧率的数据集上的对比实验

4.2.1 实验方法

在第二章和第三章的叙述中，我们采用 Leave-One-Video-Out 的方法对本文提出的微表情特征提取方法在 CASME, CASME2 和 SMIC 三个数据集上进行了测试评估，均获得了非常好的识别率。

为了对本文的方法作进一步的评估，我们采用 Leave-One-Subject-Out 的方法重新在这三个数据集上进行实验评估。Leave-One-Subject-Out 的意思是将数据集中某一个被试者的所有微表情数据作为测试数据，其他被试者的微表情数据作为训练数据完成的实验。通过 Leave-One-Subject-Out 方法得出的识别率，可以评估基于光流域矫正的微表情运动特征对人物的敏感度。

同样我们使用式 (2-18) 的特征表达式，并使用 POLY 核函数的 SVM 分类器作为训练和识别工具，其中 POLY 核函数的参数依然为： $\gamma = 0.1, degree = 4, coef = 1$ 。

同时，为了与 LBP-TOP 特征进行比较，我们在相同的实验环境下，即相同的数据集，人脸分割方式，ROI 分割方式，和相同的实验方式，分别在三个数据集上

完成对比实验。

4.2.2 在 CASME 上的对比结果

由第三章的实验结果和分析我们知道，由于 CASME 中的数据集 A 的四种微表情的分类分布极不平均（其中 Others 占到了总体样本数的 80% 以上，导致在数据集 A 上的实验结果没有明显的意义。因此，在本次实验中，我们仅仅在 CASME 的数据集 B 上进行对比实验。

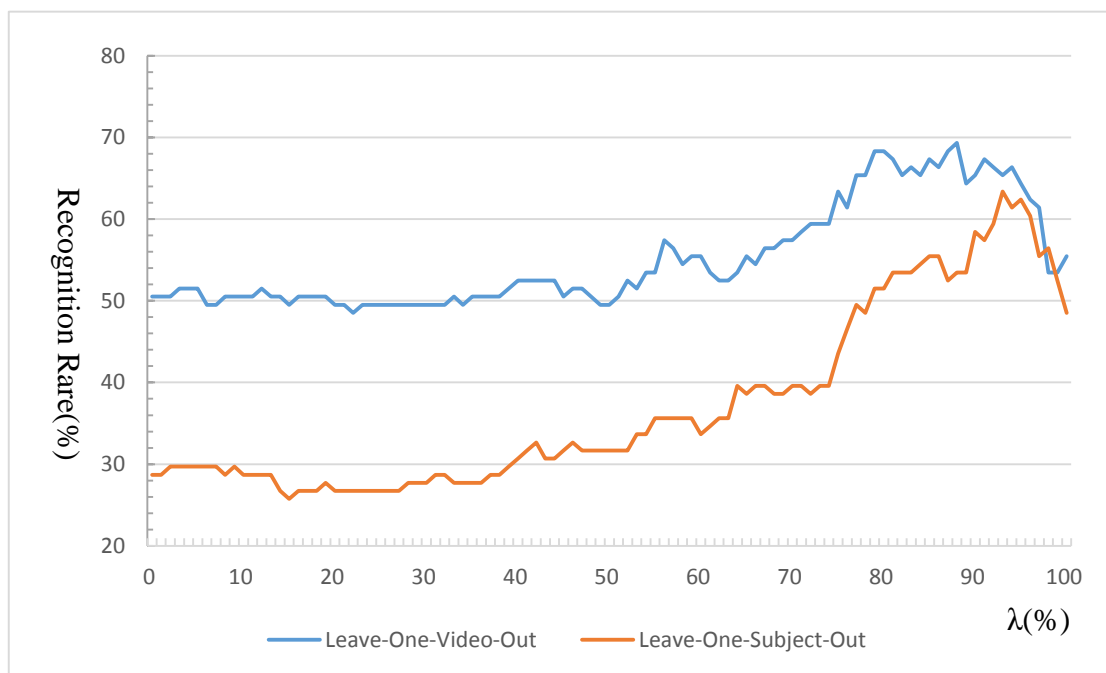


图 4.4 使用本文的特征提取方法在 CASME 的数据集 B 上的识别率

如图 4.4 所示，为我们分别使用 Leave-One-Video-Out 以及 Leave-One-Subject-Out 两种方法在 CASME 的数据集 B 上的实验结果。其中蓝色的折线图为 Leave-One-Video-Out 的实验结果，黄色的折线图为 Leave-One-Subject-Out 的实验结果。

从图 4.4 可以看出，两种方法的识别率折线图在总体上都呈现出比较稳定的特性，没有剧烈的波动，体现了 SVM 对参数 λ 的不敏感，从侧面印证了基于光流域矫正的微表情更加接近真实的微表情特征。

从折线图的走向来看，两条曲线都是首先呈平稳上升趋势，然后在区间 [80%, 90%] 之间获得峰值，然后开始迅速下降。这种结果与第三章在 CASME 上的实验结果保持一致。从数值上看，Leave-One-Subject-Out 的方法在整体上要比 Leave-One-Video-Out 的结果要差，但是在达到峰值的区域，Leave-One-Subject-Out 方法的识别率慢慢接近了 Leave-One-Video-Out 的识别率。其中，在参数 $\lambda = 93\%$ ，Leave-

One-Subject-Out 的方法获得峰值识别率 63.37%。因此，我们可以得出如下结论：使用本文的微表情特征的提取方法，在 CASME 的数据集 B 上，对人物存在一定的敏感性，但敏感性不高。

表 4.1 使用 LBP-TOP 特征在 CASME 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	58.42%	58.42%
	R=1, L=2	60.40%	58.42%
	R=2, L=1	56.44%	60.40%
	R=2, L=2	61.39%	57.43%
Leave-One-Subject-Out	R=1, L=1	29.70%	33.63%
	R=1, L=2	31.68%	39.60%
	R=2, L=1	28.71%	28.71%
	R=2, L=2	29.70%	29.70%

如表 4.1 所示，是使用 LBP-TOP 特征在 CASME 的数据集 B 上的识别率，其中，参数 P 表示取像素点周围的 P 个点作为描述改点纹理的特征的参照点；参数 R 表示取参照点时的空域半径；参数 L 表示取参照点时的时域半径；这些参数的意义如 4.1 节所述。

为了根据表 4.1 可知，LBP-TOP 特征在 Leave-One-Video-Out 和 Leave-One-Subject-Out 下的最高识别率为 61.39% 和 39.60%。这两个数据要比本文的微表情识别方法低得多。其中，使用 Leave-One-Subject-Out 方法的识别率比 Leave-One-Video-Out 低的多的结果表明人物对 LBP-TOP 的影响很大。

4.2.3 在 CASME2 上的对比结果

同样的，我们在 CASME2 上也给出本文的方法和 LBP-TOP 特征的对比实验结果，其中我们仍与第三章一样，取 CASME2 数据集中 173 个样本为代表进行实验。

如图 4.5 所示，为我们分别使用 Leave-One-Video-Out 以及 Leave-One-Subject-Out 两种方法在 CASME2 数据集上的实验结果。其中蓝色的折线图为 Leave-One-Video-Out 的实验结果，黄色的折线图为 Leave-One-Subject-Out 的实验结果。

从图 4.5 可以看出，两种方法的识别率与图 4.4 的折线图一样，在总体上都呈现出比较稳定的特性，体现了 SVM 分类器参数 λ 不敏感的特性，从侧面印证了基于光流域矫正的微表情特征更加接近真实的微表情特征。

从折线图的走向来看，两条折线图的走向基本保持一致，在参数 $\lambda < 90\%$ 的时

候呈缓慢上升趋势，之后都略有下降。从数值上看，Leave-One-Subject-Out 的方法在整体上比 Leave-One-Video-Out 的结果差了 10% 左右。其中，在参数 $\lambda = 95\%$ ，Leave-One-Subject-Out 的方法获得峰值识别率 61.85%。这个结果说明了，在 CASME2 上，本文提出的方法对人物的敏感性比 CASME 更强。

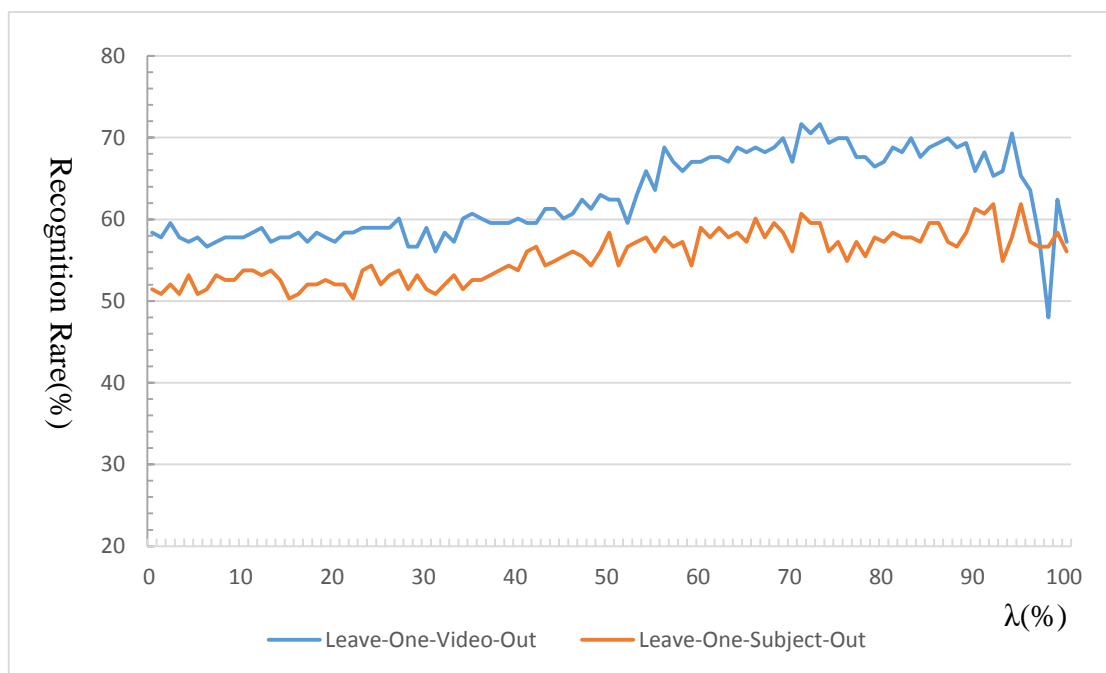


图 4.5 使用本文的特征提取方法在 CASME2 数据集上的识别率

相应的，我们给出了使用 LBP-TOP 特征在 CASME2 的数据集上的识别率，如表 4.2 所示：

表 4.2 使用 LBP-TOP 特征在 CASME2 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	54.91%	56.07%
	R=1, L=2	60.12%	55.49%
	R=2, L=1	60.69%	56.07%
	R=2, L=2	63.58%	57.80%
Leave-One-Subject-Out	R=1, L=1	45.66%	49.13%
	R=1, L=2	46.24%	48.55%
	R=2, L=1	43.35%	49.71%
	R=2, L=2	46.24%	52.02%

根据表 4.2，我们发现，在数据集 CASME2 上，不管是使用 Leave-One-Video-

Out 还是 Leave-One-Subject-Out 方法, LBP-TOP 特征的识别率也都比本文的方法低。同时根据两种实验方法的识别率的比较, 我们依然可以得出人物对使用 LBP-TOP 特征的微表情识别方法存在着比较大的影响。

4.2.4 在 SMIC 上的对比结果

SMIC 是 Pfister 等人为了测试 LBP-TOP 的性能而建立的一个微表情库, 在本小节是实验中, 为了实验数据的准确性我们同时给出了使用 Leave-One-Video-Out 和使用 Leave-One-Subject-Out 在 LBP-TOP 特征以及本文的特征提取方法的实验结果, 并进行分析比较。

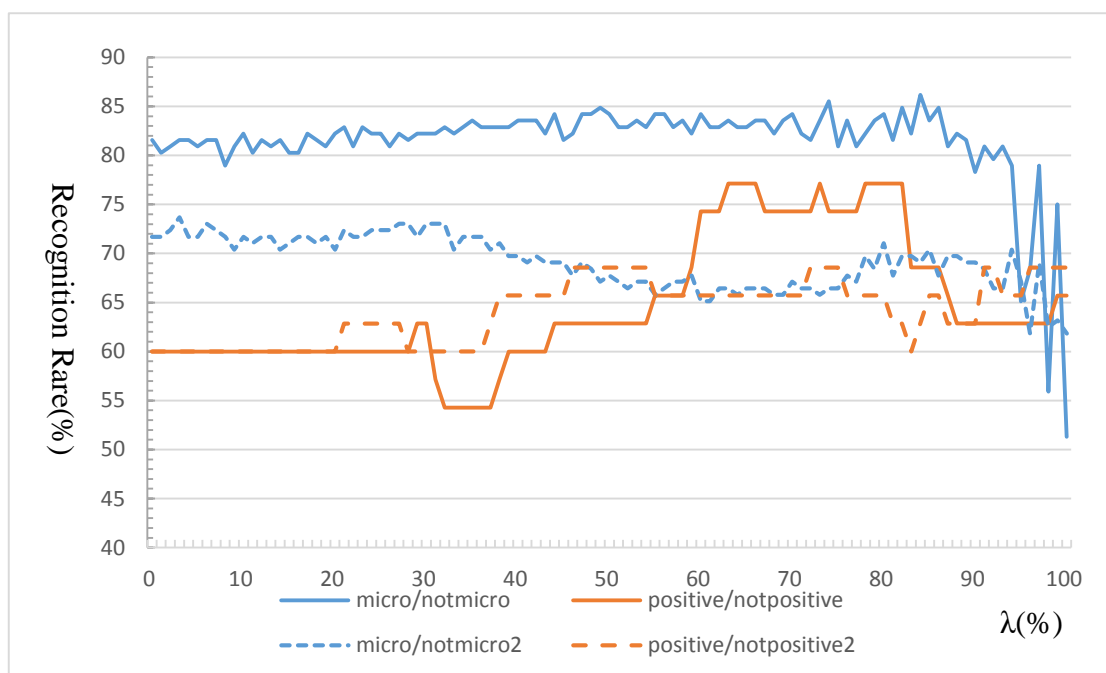


图 4.6 使用本文的特征提取方法在 SMIC 数据集上的识别率

如图 4.6 所示, 为我们分别使用 Leave-One-Video-Out 以及 Leave-One-Subject-Out 两种方法在 SMIC 数据集上的实验结果。其中实线图为 Leave-One-Video-Out 的实验结果, 虚线图为 Leave-One-Subject-Out 的实验结果。

从图 4.5 可以看出, 两种方法的识别率与 CASME 数据集一样, 在总体上都呈现出比较稳定的特性, 没有因为参数 λ 的改变而出现剧烈波动的情况, 体现了 SVM 对参数 λ 的微小扰动的不敏感, 从侧面印证了基于光流域矫正的微表情更加接近真实的微表情特征。

从折线图的走向来看, 使用 Leave-One-Subject-Out 实验方法的两条折线图呈交错状态, 这与第二章的在 SMIC 数据集上的折线图走向一致, 表明了微表情和

非微表情之间的界定中，微表情的强度特征更为重要；而在微表情类间的识别中，微表情的方向特征更为重要。从数值上看，Leave-One-Subject-Out 的方法在整体上比 Leave-One-Video-Out 的结果差了 10% 左右。其中 Leave-One-Subject-Out 的方法在 Micro/Notmicro 分类和 Positive/Negative 上获得的最高的识别率分别为：73.68% 和 68.57%。这个结果比 Leave-One-Video-Out 方法的结果低很多，说明了在 SMIC 上本文提出的方法对人物也存在很强的敏感性。

相应的，如表 4.3 和表 4.4 所示，我们给出了使用 LBP-TOP 特征在 SMIC 的数据集上的识别率。其中表 4.3 为 LBP-TOP 特征在 Micro/Notmicro 分类集上的识别率，表 4.4 为 LBP-TOP 特征在 Positive/Negative 分类集上的识别率。

表 4.3 使用 LBP-TOP 特征在 Micro/Notmicro 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	86.84%	78.95%
	R=1, L=2	86.18%	76.97%
	R=2, L=1	91.45%	86.84%
	R=2, L=2	87.50%	84.21%
Leave-One-Subject-Out	R=1, L=1	58.55%	60.53%
	R=1, L=2	57.24%	65.79%
	R=2, L=1	51.97%	62.50%
	R=2, L=2	50.66%	59.87%

表 4.4 使用 LBP-TOP 特征在 Positive/Negative 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	34.29%	40.00%
	R=1, L=2	40.00%	40.00%
	R=2, L=1	37.14%	45.71%
	R=2, L=2	37.14%	48.57%
Leave-One-Subject-Out	R=1, L=1	42.86%	48.57%
	R=1, L=2	48.57%	48.57%
	R=2, L=1	51.43%	54.29%
	R=2, L=2	51.43%	57.14%

从这两张使用 LBP-TOP 特征在 SMIC 数据集上的识别率结果可以发现：1) LBP-TOP 特征在使用 Leave-One-Video-Out 的方法下在区分微表情和非微表情方面存在着很大的优势，它最高识别率为 91.45%；但是在 Leave-One-Subject-Out 的方法下识别率就差了很多，仅为 65.79%，这表明了 LBP-TOP 特征对人物有着很强

的敏感性。2) LBP-TOP 特征不适合用于微表情的类内区分。不管是使用 Leave-One-Video-Out 还是 Leave-One-Subject-Out 的方法, Positive/Negative 的识别率都非常低, 仅为 48.57%, 57.41%。

4.3 在帧率为 25fps 的数据集上的对比实验

4.3.1 实验方法

在上一节中, 我们在微表情数据集的原始帧率上对本文提出的特征提取方式与 LBP-TOP 特征进行了比较。为了获取当帧率下降时, 微表情识别率和特征提取方式之间的关系, 我们将在帧率为 25fps 的数据集上进行测试。帧率为 25fps 的数据集是通过对三个微表情数据集 CASME, CASME2 和 SMIC 上的视频序列进行采样而完成的。然后我们分别对这三个采样后的数据集进行对比实验。

在采样后的数据集上, 我们仍然采用 Leave-One-Video-Out 以及 Leave-One-Subject-Out 这两种方法进行实验, 并与未采样的实验数据进行比较, 从而获得微表情特征与视频帧率的敏感性。

同时在分类器与核函数的选择上, 我们也与上一小节保持一致。

4.3.2 在 CASME 上的对比结果

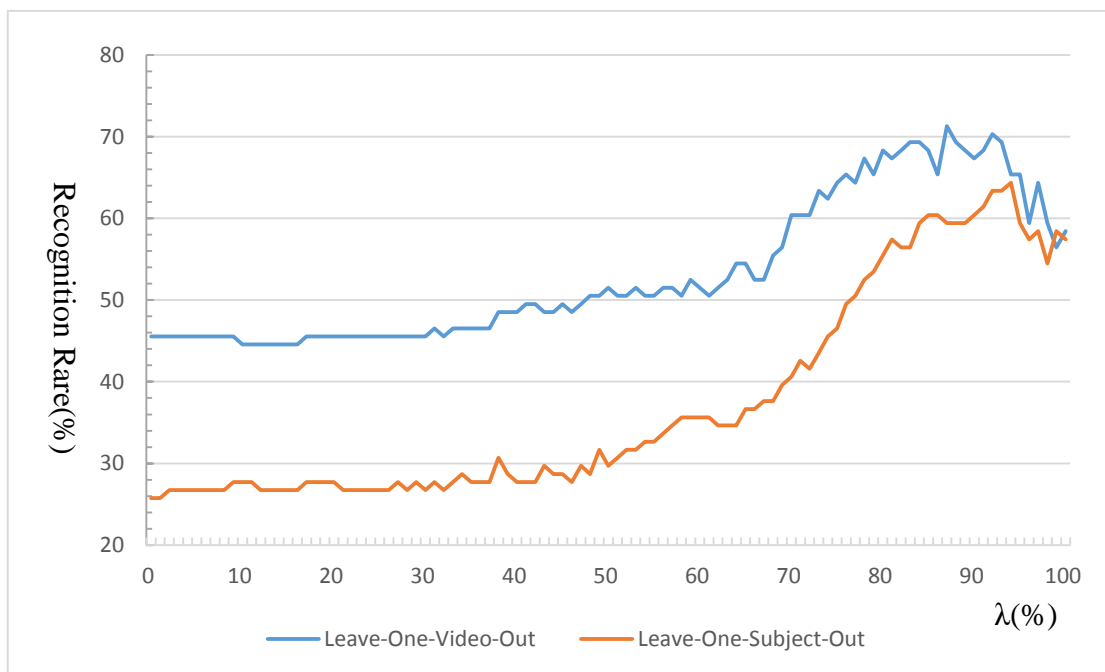


图 4.7 使用本文的特征提取方法在 25fps 的 CASME 的数据集 B 上的识别率

为了与 4.2.2 小节的实验数据保持一致，在本小节，我们仍然选择在数据集 B 上完成我们的对比实验。同时，我们给出使用 LBP-TOP 在帧率降至 25fps 的数据集 B 上的识别率，如表 4.5 所示。从表格上可以看出，LBP-TOP 特征在低帧率的情况下识别率也比本文的识别方法低，而且从 Leave-One-Subject-Out 方法上识别率比 Leave-One-Video-Out 方法低得多的情况下，依然可以得出 LBP-TOP 方法对人物非常敏感这一结论。

表 4.5 使用 LBP-TOP 特征在帧率为 25fps 的 CASME 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	58.42%	54.46%
	R=1, L=2	64.36%	56.44%
	R=2, L=1	57.43%	54.46%
	R=2, L=2	60.40%	54.46%
Leave-One-Subject-Out	R=1, L=1	33.66%	34.65%
	R=1, L=2	32.67%	32.67%
	R=2, L=1	31.68%	32.67%
	R=2, L=2	31.68%	32.67%

4.3.3 在 CASME2 上的对比结果

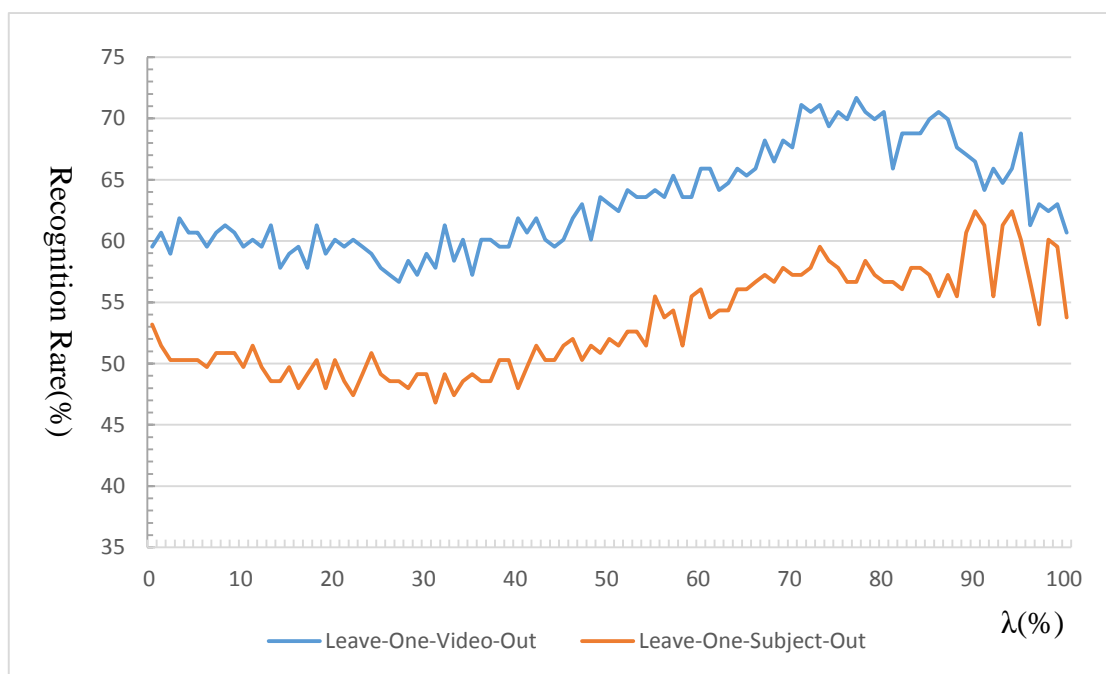


图 4.8 使用本文的特征提取方法在 CASME2 数据集上的识别率

同样的，我们在 CASME2 的帧率下降为 25fps 的数据集上也给出本文的方法和 LBP-TOP 特征的对比实验结果。

如图 4.8 所示，为我们分别使用 Leave-One-Video-Out 以及 Leave-One-Subject-Out 两种方法在 CASME2 数据集上的实验结果。其中蓝色的折线图为 Leave-One-Video-Out 的实验结果，黄色的折线图为 Leave-One-Subject-Out 的实验结果。

同样我们从图 4.5 和图 4.8 可以看出，使用本文提出的特征提取方法，在帧率为 25fps 的 CASME2 上的微表情识别率与原帧率上的识别率折线图拥有者高度一致的稳定性和走向。其中，Leave-One-Video-Out 和 Leave-One-Subject-Out 分别获得最高的识别率为 71.68% 和 62.43%，与原帧率上的识别率没有太大的差别。图 4.5 和图 4.8 的这两组数据同样表明了，在 CASME2 上本文提出的微表情特征提取方法呈现出与帧率不敏感的特性。

同时，我们给出使用 LBP-TOP 在帧率 25fps 的 CASME2 上的识别率，如表 4.6 所示。从 LBP-TOP 在 CASME2 上的识别率可以看出，使用 LBP-TOP 的特征在低帧率上的结果反而比原始帧率的数据集上的实验结果好。但是，它在低帧率下依然存在着对人物敏感的特性。

表 4.6 使用 LBP-TOP 特征在帧率为 25fps 的 CASME2 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	72.83%	69.36%
	R=1, L=2	74.57%	67.05%
	R=2, L=1	72.83%	72.25%
	R=2, L=2	73.99%	72.25%
Leave-One-Subject-Out	R=1, L=1	47.98%	51.455
	R=1, L=2	51.45%	53.76%
	R=2, L=1	49.71%	53.18%
	R=2, L=2	52.02%	54.91%

4.3.4 在 SMIC 上的对比结果

如图 4.9 所示，为我们分别使用 Leave-One-Video-Out 以及 Leave-One-Subject-Out 两种方法在 CASME2 数据集上的实验结果。其中实线图为 Leave-One-Video-Out 的实验结果，虚线图为 Leave-One-Subject-Out 的实验结果。同样的，通过图 4.6 和图 4.9 两个图像上的折线图对比，我们仍然能得到如下结论：在 SMIC 上本文提出的微表情特征提取方法呈现出与帧率不敏感的特性。其中，在“micro/notmicro”数据集上，Leave-One-Video-Out 和 Leave-One-Subject-Out 分别获得最高的识别率为

84.87%和 71.71%；在“positive/negative”数据集上，这一数据为：80%和 71.43%。

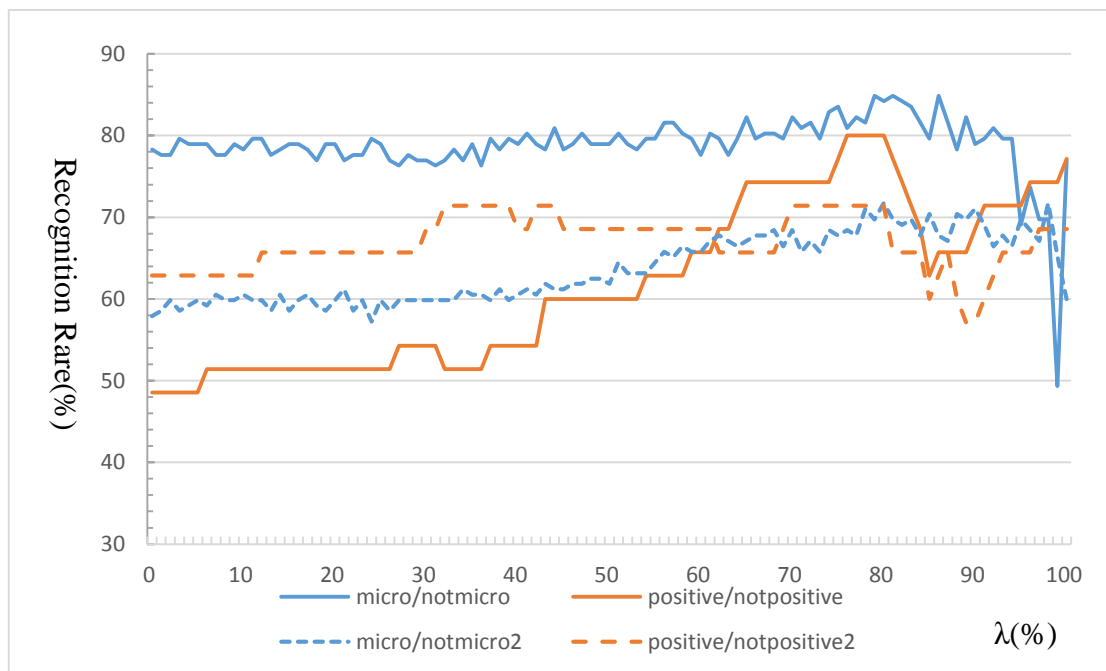


图 4.9 使用本文的特征提取方法在 SMIC 数据集上的识别率

相应的，如表 4.7 和表 4.8 所示，我们给出了使用 LBP-TOP 特征在在帧率为 25fps 的 SMIC 的数据集上的识别率。从表 4.7 和 4.8 的结果可以看出，LBP-TOP 特征在 SMIC 上对帧率不敏感，但是仍然存在着对人物敏感的特点。

表 4.7 使用 LBP-TOP 特征在帧率为 25fps 的 Micro/Notmicro 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	91.45%	88.82%
	R=1, L=2	89.47%	88.16%
	R=2, L=1	93.42%	91.45%
	R=2, L=2	93.42%	89.47%
Leave-One-Subject-Out	R=1, L=1	53.29%	55.26%
	R=1, L=2	53.95%	57.24%
	R=2, L=1	61.84%	56.58%
	R=2, L=2	56.58%	57.89%

4.4 本章小结

在本章中，我们对第 2 章和第 3 章提出的微表情识别算法在 CASME, CASME2

和 SMIC 上进行了更加细致的实验。在本次实验中，我们在 Leave-One-Video-Out 的基础上，使用 Leave-One-Subject-Out 的方法对基于光流域矫正的微表情运动特征的识别方法进行评估。评估显示，本文提出的微表情算法对人物有一定的敏感性，但敏感性不强烈。

同时，我们将本文提出的方法在帧率降至 25fps 的三个数据集上再次完成了上述实验，实验表明，本文提出的微表情识别方法，对数据集的帧率不敏感。

表 4.8 使用 LBP-TOP 特征在在帧率为 25fps 的 Positive/Negative 上的识别率

Method	Interval	P=4	P=8
Leave-One-Video-Out	R=1, L=1	45.71%	40.00%
	R=1, L=2	51.43%	40.00%
	R=2, L=1	57.14%	51.43%
	R=2, L=2	54.29%	42.86%
Leave-One-Subject-Out	R=1, L=1	54.29%	42.86%
	R=1, L=2	48.57%	45.71%
	R=2, L=1	51.43%	57.14%
	R=2, L=2	51.43%	54.29%

为了将本文的方法与其他方法进行比较，我们在每次实验中，都完成了与 LBP-TOP 特征的对比实验，对比实验显示：1) 本文提出的方法在使用 Leave-One-Video-Out 的实验方法下，它的实验结果整体上要稍好于 LBP-TOP 方法；2) 本文提出的方法在使用 Leave-One-Subject-Out 的实验方法下，它的实验结果比 LBP-TOP 方法好得多，表明了本文方法对人物的敏感性要远小于 LBP-TOP 方法；3) 本文的提出的方法和 LBP-TOP 方法，都表现出了与帧率无关的特性。

第 5 章 总结与展望

微表情是一种快速的不易被人察觉的面部表情，它在生活中的很多领域都有着非常广泛的应用前景。然而由于微表情运动持续时间短和运动幅度小，许多应用在表情识别中的研究方法都不能直接运用到微表情识别中去。同时，在之前的微表情自动检测和识别的研究工作中，研究者往往忽视微表情的方向特征在识别中的作用，从而并没有获得令人满意的微表情识别率。为了解决上述问题，本文提出了一种基于运动主方向特征的微表情识别方法。

同时，在微表情的识别研究过程中，我们发现在微表情的持续时间内，存在着被试者的轻微头动这一常见的现象。为了解决因被试者轻微头动对微表情识别造成的影响，本文提出了一种基于光流域的微表情运动特征的矫正方法。与本文提到的从图像域进行矫正的方法相比，该方法避免了直接对原始图像进行矫正的操作，使得获得的矫正效果更好。

最后，为了评估微表情特征中强度特征和方向特征在识别中的作用，我们引入了权值参数，对微表情特征向量进行了改写。同时，我们使用 **Leave-One-Video-Out** 和 **Leave-One-Subject-Out** 的方法，分别在 **CASME**，**CASME2** 和 **SMIC** 这三个数据集上对本文获取的微表情特征进行评估和分析。结果表明，在微表情的识别实验中，方向特征有着比强度特征更为显著的作用。

除了对微表情的强度特征和方向特征进行评估外，我们还把本文的方法与 **LBP-TOP** 特征进行比较，比较的结果表明，本文的方法在三个数据集上的效果都要比 **LBP-TOP** 的方法好。同时，为了评估帧率对本文方法的影响，我们在三个数据集的帧率降至 **25fps**，并在新的数据集上对本文的方法进行评估，评估结果显示，本文的方法对数据集的帧率不敏感。

参考文献

- [1] Russell, James A, Jos é Miguel Fern ández-Dols. The psychology of facial expression. Cambridge university press, 1997.
- [2] Weinberger S. Airport security: intent to deceive? Nature, 2010, 465(7297): 412-415.
- [3] Ekman P, Friesen W V. Nonverbal leakage and clues to deception. Psychiatry, 1969, 32(1): 88-106.
- [4] Ekman P. Lie catching and microexpressions. The philosophy of deception, 2009: 118-133.
- [5] Haggard E A, Isaacs K S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. Methods of research in psychotherapy. Springer US, 1966: 154-165.
- [6] Coolican H. Research methods and statistics in psychology. Hodder & Stoughton Educational, 1990.
- [7] Ekman P, Friesen W V. Nonverbal Behavior and Psycho Pathology. 1974.
- [8] Ekman P, O'Sullivan M. Who can catch a liar? American psychologist, 1991, 46(9): 913.
- [9] Frank M G, Ekman P. The ability to detect deceit generalizes across different types of high-stake lies. Journal of personality and social psychology, 1997, 72(6): 1429.
- [10] Matsumoto D, LeRoux J, Wilson-Cohn C, et al. A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). Journal of Nonverbal Behavior, 2000, 24(3): 179-209.
- [11] Ekman P. MicroExpression training tool (METT). San Francisco: University of California, 2002.
- [12] Ekman P. Lie catching and microexpressions. The philosophy of deception, 2009: 118-133.
- [13] Russell T A, Chu E, Phillips M L. A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro - expression training tool. British Journal of Clinical Psychology, 2006, 45(4): 579-583.
- [14] Russell T A, Green M J, Simpson I, et al. Remediation of facial emotion perception in schizophrenia: concomitant changes in visual attention. Schizophrenia research, 2008, 103(1): 248-256.
- [15] Fellner A N, Matthews G, Funke G J, et al. The effects of emotional intelligence on visual search of emotional stimuli and emotion identification. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Sage Publications, 2007, 51(14): 845-849.

-
- [16] Frank M G, Herbasz M, Sinuk K, et al. I see how you feel: Training laypeople and professionals to recognize fleeting emotions. The Annual Meeting of the International Communication Association. Sheraton New York, New York City. 2009.
 - [17] Polikovskiy S, Kameda Y, Ohta Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. 2009.
 - [18] Shreve M, Godavarthy S, Goldgof D, et al. Macro-and micro-expression spotting in long videos using spatio-temporal strain. Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. IEEE, 2011: 51-56.
 - [19] Shreve M, Godavarthy S, Manohar V, et al. Towards macro-and micro-expression spotting in video using strain patterns. Applications of Computer Vision (WACV), 2009 Workshop on. IEEE, 2009: 1-6.
 - [20] Pfister T, Li X, Zhao G, et al. Recognising spontaneous facial micro-expressions. Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 1449-1456.
 - [21] Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007, 29(6): 915-928.
 - [22] Wang S J, Yan W J, Li X, et al. Micro-expression Recognition Using Dynamic Textures on Tensor Independent Color Space. Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014: 4678-4683.
 - [23] Cristinacce D, Cootes T F. Feature Detection and Tracking with Constrained Local Models. BMVC. 2006, 2(5): 6.
 - [24] Cootes T F, Edwards G J, Taylor C J. Active appearance models. IEEE Transactions on pattern analysis and machine intelligence, 2001, 23(6): 681-685.
 - [25] Cootes T F, Taylor C J, Cooper D H, et al. Active shape models-their training and application. Computer vision and image understanding, 1995, 61(1): 38-59.
 - [26] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, 1: I-511-I-518 vol. 1.
 - [27] P. Ekman and W.V. Friesen, Facial Action Coding System. Palo Alto, Calif.: Consulting Psychologists Press, Inc., 1978.
 - [28] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision. IJCAI. 1981, 81: 674-679.
 - [29] Brox T, Bruhn A, Papenberg N, et al. High accuracy optical flow estimation based on a theory for warping. Computer Vision-ECCV 2004. Springer Berlin Heidelberg, 2004: 25-36.
 - [30] Chaudhry R, Ravichandran A, Hager G, et al. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human

-
- actions. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 1932-1939.
- [31] Chang C C, Lin C J. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.
- [32] Yan W J, Wu Q, Liu Y J, et al. Casme database: a dataset of spontaneous micro-expressions collected from neutralized faces. Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013: 1-7.
- [33] Yan W J, Li X, Wang S J, et al. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PloS one, 2014, 9(1): e86041.
- [34] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 1996, 29(1): 51-59.

致 谢

在这三年多的科研生活里，我从各位老师和同学那儿学到了很多专业知识和研究技能。在此，我要感谢曾经给予我帮助的老师 and 同学。

首先我要衷心感谢我的指导老师刘永进老师。刘老师是一个学识渊博，工作负责的老师，在这三年的科研生活中，不管是在学术指导还是在日常生活上，他都给予了我很多的关怀和帮助。他认真的工作态度和渊博的学术知识都是值得我们敬佩和学习的。

其次，我要感谢中科院软件所的马翠霞老师。马老师是一个非常勤奋和细心的人。在每天的科研工作里，我都能感受到马老师诲人不倦，孜孜以求的态度。在我刚刚踏入科研工作的时间里，她时常指导我解决在具体工作中遇到问题的一些棘手的问题。

然后，我要感谢中科院心理所的傅小兰老师，王甦菁博士和颜文靖博士。他们在心理学领域给予了我很多的帮助，让我顺利地完成了我的硕士毕业论文。

最后，我要感谢与我一起工作生活的实验室伙伴们。在忙碌而充实的科研生活中，有了你们的关心和帮助，我才能在这条路上走得如此顺利。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____日 期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

1989 年 3 月 23 日出生于江苏省无锡市。

2005 年 9 月至 2008 年 7 月就读于江苏省南菁高级中学。

2008 年 9 月考入清华大学计算机科学与技术系。

2012 年 7 月本科毕业并获得工学学士学位。

2012 年 9 月免试进入清华大学计算机科学与技术系攻读工学硕士至今。

发表的学术论文

- [1] Zhang Jinkai, Ma Cuixia, Liu Yongjin, et al. Collaborative Interaction for Videos on Mobile Devices Based on Sketch Gestures. Journal of Computer Science and Technology, 2013, 28:810-817.

研究成果

- [1] 刘永进, 张金凯, 张艳秋, 付秋芳, 傅小兰。基于草图手势及弧形摘要时间轴的视频交互方法和系统。申请号: 201310048730.4, 申请日: 2013-02-06。
- [2] 刘永进, 张金凯, 余旻婧, 吴陈铭。一种用于 SLA3D 打印机的平面分割方法及装置。申请号: 2015032300586890, 申请日: 2015-03-26。
- [3] 刘永进, 张金凯, 吴陈铭。SLA 型 3D 打印机控制系统软件。软件著作权号: 2015SR051323, 发表日: 2015.1.1, 著作权人为: 清华大学。