

# 人脸微表情识别综述

徐峰<sup>1,2</sup> 张军平<sup>1,2</sup>

**摘要** 人脸表情是人际交往的重要渠道, 识别人脸表情可促进对人心理状态和情感的理解. 不同于常规的人脸表情, 微表情是一种特殊的面部微小动作, 可以作为判断人主观情绪的重要依据, 在公共安防和心理治疗领域有广泛的应用价值. 由于微表情具有动作幅度小、持续时间短的特点, 对微表情的人工识别需要专业的培训, 且识别正确率较低. 近年来不少研究人员开始利用计算机视觉技术自动识别微表情, 极大地提高了微表情的应用可行性. 本文综述人脸微表情识别的定义和研究现状, 总结微表情识别中的一些关键技术, 探讨潜在的问题和可能的研究方向.

**关键词** 微表情识别, 表情识别, 情绪识别, 计算机视觉, 面部动作编码系统

**引用格式** 徐峰, 张军平. 人脸微表情识别综述. 自动化学报, 2017, 43(3): 333–348

**DOI** 10.16383/j.aas.2017.c160398

## Facial Microexpression Recognition: A Survey

XU Feng<sup>1,2</sup> ZHANG Jun-Ping<sup>1,2</sup>

**Abstract** Facial expression is an important channel in social interaction. Reading facial expression can improve understanding of psychological condition and emotional status. Different from normal expressions, microexpression is a special kind of subtle facial action. It serves as a vital clue for affective estimation, and has broad applications in public security and psychological treatment. Recognizing microexpression requires professional training for human due to its short duration and subtle movement. So far a low recognition accuracy has been reported in the literature. In recent years, researchers have been studying microexpression recognition based on computer vision, which can largely improve the feasibility of such recognition. In this article, we introduce problem definition and current research status of microexpression, survey several representative techniques in this topic, and discuss some underlying issues and potential research directions.

**Key words** Microexpression recognition, expression recognition, emotion recognition, computer vision, facial action code system (FACS)

**Citation** Xu Feng, Zhang Jun-Ping. Facial microexpression recognition: a survey. *Acta Automatica Sinica*, 2017, 43(3): 333–348

表情是人类情绪的直观反应, 表情识别一直是计算机视觉的重要研究课题之一. 在过去几十年中, 研究人员已经在各类表情识别问题中取得了重要的成果<sup>[1–6]</sup>.

近年来, 针对自发式的表情 (Spontaneous expression) 的识别逐渐成为新的研究热点<sup>[7–8]</sup>. 与普通表情不同, 自发式的表情无法伪造和抑制, 可以反映人的真实情绪.

微表情 (Microexpression) 是一种自发式的表

情, 在人试图掩盖内在情绪时产生, 既无法伪造也无法抑制<sup>[9]</sup>. 与表情研究相比, 微表情发现的历史较短. 1966 年 Haggard 等第一次提出了微表情的概念<sup>[10]</sup>, 此后 Ekman 等报道了关于微表情的案例<sup>[11]</sup>. 在一段心理医生与抑郁症病人的对话中<sup>[11]</sup>, 经常微笑的病人偶尔有几帧非常痛苦的表情. 研究人员将这种人在经历强烈情绪时产生的快速、无意识的自发式面部动作称为微表情.

微表情在情绪识别任务上的可靠度很高, 对表情情感识别任务有潜在的利用价值, 如婚姻关系预测<sup>[12]</sup>、交流谈判<sup>[13]</sup>、教学评估<sup>[14–15]</sup> 等. 除了用于情感分析, 研究人员观察到了有意说谎时产生的微表情<sup>[16]</sup>, 经过微表情识别训练, 普通人识别说谎的能力得到提高<sup>[17]</sup>.

微表情的持续时间非常短, 研究表明微表情仅持续 1/25 s ~ 1/3 s<sup>[18]</sup>, 且动作幅度非常小, 不会同时在上半脸和下半脸出现<sup>[16, 18]</sup>, 因此正确观测并且识别有着相当的难度.

收稿日期 2016-05-15 录用日期 2016-07-28  
Manuscript received May 15, 2016; accepted July 28, 2016  
国家自然科学基金 (61673118, 61273299), 浦江人才计划 (16PJJD09) 资助  
Supported by National Natural Science Foundation of China (61673118, 61273299) and Shanghai Pujiang Program (16PJJD09)  
本文责任编辑 赖剑煌  
Recommended by Associate Editor LAI Jian-Huang  
1. 上海智能信息处理重点实验室 上海 200433 2. 复旦大学计算机科学技术学院 上海 200433  
1. Shanghai Key Laboratory of Intelligent Information Processing, Shanghai 200433 2. School of Computer Science, Fudan University, Shanghai 200433

尽管已有了专业的培训工具<sup>[19]</sup>,但依靠人力识别的准确率并不高,已知的文献报道中只有 47%<sup>[20]</sup>.另外,依靠人力识别受限于专业培训和时间成本,难以进行大规模推广.因此,近年来依赖计算机实现人脸微表情自动识别的需求越来越高.

利用计算机识别微表情具有独特的优势. 1) 无论多迅速的运动,只要客观上被摄像机捕捉,计算机就能获取相应的信息并进行处理,因此研究人员引入高速摄像机用于微表情的捕捉. 2) 只要能训练出高效、稳定的模型,计算机就能够以低廉的成本处理大规模的微表情识别任务,这显然超过专业人员人工识别微表情的效率.

然而面部产生微表情时也存在无关的变化,正确分离无用信息,提取微表情相关的重要信息,是用计算机有效识别微表情的关键所在.

目前的研究可以分为两类研究方向,一种是通过构建计算机视觉特征,寻找高效的表达形式描述微表情并进行模型学习;另一种是针对微表情本身的特点,寻找合适的学习算法进行模型的构建.

这两种研究方式并不是互斥的,一个现实的微表情识别系统往往需要两部分协同工作才能有效地完成识别任务.而一些机器学习的技术也被应用到求解特征表达的过程之中.目前这两类工作都在现有数据集上取得了一定效果.

当前国内关于微表情方面的研究综述主要集中在心理学方面,例如中国科学院的吴奇等的工作综述了微表情的潜在应用<sup>[21]</sup>.关于微表情自动识别的综述仅有山东大学的贾晔焯等较早期的工作<sup>[22]</sup>,将微表情识别方法分为应变模式法和利用机器学习的方法,前者计算面部皮肤的应变响应模式并确定阈值作为识别的依据,后者使用特征提取与模式分类的框架进行识别.

近年来涌现了大量针对微表情中不同细分问题的新研究,已经超出了以往的分类范畴.与贾晔焯等的综述<sup>[22]</sup>相比,本文侧重按微表情识别中的不同问题进行分类,介绍近年来大量新的研究工作.本文组织结构如下:第 1 节给出微表情识别的具体问题定义;第 2 节介绍微表情的数据集;第 3 节从不同的问

题定义出发介绍当前微表情识别的方法;第 4 节和第 5 节讨论微表情识别潜在的问题及未来值得研究的方向.

## 1 问题定义

微表情识别是一个较为宽泛的问题,实际包含了多个具体的目标.

对于一段给定的面部图像序列,首先要鉴别其中是否包含微表情.这包括两种场景,一种是将微表情从无表情(即肌肉不运动时静止的面部状态)的序列中区分出来;另一种是区别微表情和普通的表情.这两者都称为微表情的检测任务.

与普通表情一样,微表情蕴含了人类的情绪.鉴别微表情包含的情绪是最常见的任务.在一段已知是微表情的人脸序列中,识别其中表达的情绪,称为微表情的分类任务.

通常所说的微表情识别包含了上述两类任务.其基于的应用场景假设是,从一段人脸图像序列中检测微表情的片段,再对检测出的微表情的情绪进行分类.

除了上述两种任务,微表情识别还有更加精细的划分.与普通表情相似,微表情也可以分为两段,并以三个时间点分割.起始点(Onset)是指微表情出现的瞬间;顶点(Apex)是指微表情幅度最大的瞬间;终止点(Offset)是指微表情消失的瞬间.阶段划分就是从一段人脸图像序列中检测微表情并对三个特殊的时间点进行标定.

面部动作编码系统(Facial action coding system, FACS)是一种表情识别领域常用的编码标准,包含一组预先定义的编码表,每个编码称为一个动作单元(Action unit, AU).每个动作单元表示一个特定的面部局部动作,例如动作单元 2 表示眉毛外端上扬(Outer brow raiser),动作单元 3 表示眉毛下垂(Brow lowerer)等.微表情可以用 FACS 进行编码,即对微表情进行动作单元识别.

图 1 展示了微表情识别的两类通用框架.在图 1(a)中,算法只需要考虑已经分割好的面部图像序

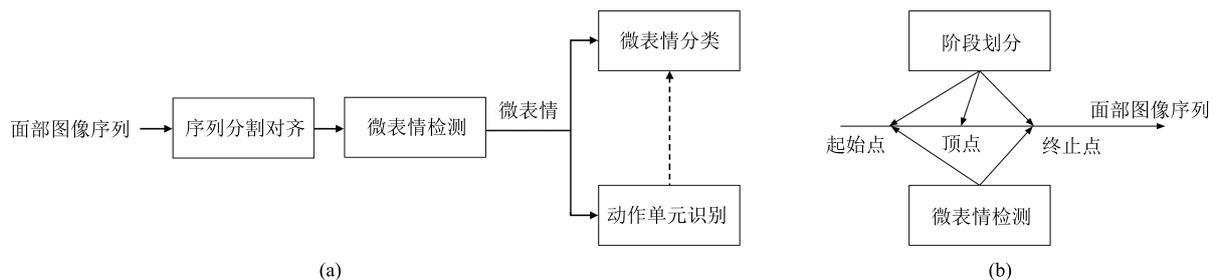


图 1 微表情识别中的具体任务

Fig. 1 Specific tasks in microexpression recognition

列, 检测其中属于微表情的序列, 并对微表情序列进一步分类或者识别其中的动作单元; 在图 1 (b) 中, 算法在长视频里检测寻找微表情并进行阶段划分, 标定起始点、顶点、终止点, 这也可以看作是检测问题的一种拓展.

## 2 微表情数据集

微表情数据采集困难, 且非专业人员很难鉴定微表情. 因此微表情数据集的采集和选择非常重要. 目前已知的微表情数据集有: 芬兰 Oulu 大学的 SMIC (Spontaneous microexpression corpus)<sup>[23]</sup> 和 SMIC 2<sup>[24]</sup>、中国科学院的 CASME (Chinese Academy of Sciences microexpression)<sup>[25]</sup> 和 CASME II<sup>[26]</sup>、美国南佛罗里达大学的 USF-HD<sup>[27]</sup> 和日本筑波大学的 Polikovsky dataset<sup>[28]</sup>. 其中 SMIC 2 包含三个子集 HS、VIS 和 NIR, 其区别是拍摄设备不同, 分别是高速摄像机、普通相机和近红外摄像机.

由于微表情的特殊性, 一个数据集有 3 项值得注意的要素, 分别是帧率、诱导方式和标注方式.

### 2.1 帧率

微表情的持续时间通常只有  $1/25\text{s} \sim 1/3\text{s}$ , 而普通摄像机的帧率是 25 帧/秒, 因此有可能仅捕捉到非常少的几帧图像, 难以进一步处理. 为此, 一些数据集引入高速摄像机拍摄微表情. 例如 SMIC 使用 100 帧/秒的摄像机, CASME 使用 60 帧/秒的摄像机, 捕捉更多的面部图像. 现有文献中帧率最高的是 Polikovsky 等使用的数据集<sup>[28]</sup> 和 CASME II<sup>[26]</sup>, 两者都采用 200 帧/秒的高速摄像机. 但帧率并不是越高越好, 过高的帧率会导致快门进光量减少, 降低图像质量.

也有一些例外, 如 SMIC 2/VIS、SMIC 2/NIR 和 USF-HD, 这些数据集是用来衡量算法在普通帧率下识别微表情的性能的.

### 2.2 诱导方式

微表情有特定的发生场景, 是人在试图掩盖自己情绪时产生的微小面部动作. 严格地说, 人主观模拟的微小表情不能称为微表情. 因此诱导方法决定了微表情数据集的可靠程度.

在 SMIC 中<sup>[23]</sup>, 被试者被要求观看能够引起情绪波动的视频, 并尽力不流露出内在情绪; 而记录者则被要求在不观看视频的情况下猜测被试者的情绪. 为了激励被试者尽力抑制表情, 如果被试者的情绪被记录者发现, 则被试者需要填一份冗长的问卷作为惩罚. 这样的机制能够确保微表情的可靠性.

CASME<sup>[25]</sup> 采用了类似的机制确保数据集的可靠性. 同样以观看视频诱发被试者的情绪, 如果被试者成功抑制了自己的情绪没有被记录者发现, 可以获得一份现金奖励.

在另外一些数据集中, 没有类似的机制来保证数据的可靠性, 只是要求被试者观看微表情的视频图像资料, 并试图模仿微表情, 因此可能并不能称为严格意义上的微表情.

### 2.3 标注方式

根据问题定义的不同, 微表情数据集的标注方法各有不同. 在多数数据集中, 微表情识别通常被拆分成两项子任务, 即检测与分类. 给定一段图像序列, 检测旨在鉴定该序列是否包含微表情; 分类是在该序列是微表情的假定下, 对微表情进行进一步的情绪分类.

在进一步的分析中, 与普通表情类似, 微表情可以用情绪和 FACS<sup>[29]</sup> 两种方法进行标记. 微表情对应的情绪可以根据粒度的大小分为不同的类别, 例如积极情绪和消极情绪, 或愉快、惊讶、厌恶、恐惧和悲伤等.

SMIC 和 SMIC 2 都使用了情绪分类进行标注. CASME 和 CASME II 对每个表情既标注了情绪, 也标注了 FACS 编码.

表 1 总结了上述微表情数据集. 图 2 展示了一些微表情的示例.

表 1 现有微表情数据集

Table 1 Existing datasets of microexpressions

数据集	帧率	# 被试	# 微表情	# 非微表情	诱导方式	标注方法
SMIC	100	6	76	76	自发	情绪
SMIC 2/HS	100	20	164	164	自发	情绪
SMIC 2/VIS	25	10	71	71	自发	情绪
SMIC 2/NIR	25	10	71	71	自发	情绪
CASME	60	35	195	/	自发	情绪/FACS
CASME II	200	35	247	/	自发	情绪/FACS
USF-HD	29.7	/	100	181	模仿	微/非微表情
Polikovsky	200	10	/	/	模仿	FACS



图 2 微表情数据集示例

Fig. 2 Examples of microexpression datasets

### 3 微表情的识别方法

在过去几年中,涌现了不少微表情识别的研究工作.大多数工作同时研究微表情的检测与分类两个问题,也有一些工作针对特点问题,例如特定情绪的微表情的检测、微表情不同阶段的划分等.本节根据不同的问题定义,介绍现有的微表情识别方法.

#### 3.1 通用的微表情检测与分类方法

给定一段面部图像序列,检测任务要求判断该序列是否包含微表情,而分类任务则在假定包含微表情的情况下判定微表情的情绪种类.在机器学习的意义上,这两个问题都属于分类问题.而实际上,的确有一些方法用相同的框架处理微表情的检测和分类任务,这些方法抽取相同的特征,用相同的算法针对两个任务训练两个独立的模型,分别用于微表情的检测和分类.

#### 3.1.1 基于 LBP-TOP 的识别方法

Pfister 等的工作<sup>[23]</sup>是最早的自动识别微表情的尝试之一.该方法极具代表性,为之后的微表情识别工作提供了可靠的验证平台和对比标杆.

该方法首先使用一个 68 点的主观形状模型 (Active shape model, ASM)<sup>[30]</sup>定位人脸的关键点.以得到的关键点为基础,使用局部加权平均算法 (Local weighted mean, LWM)<sup>[31]</sup>计算每个序列第一帧中面部图像与模型面部图像的形变关系,并将该形变作用于对应序列的每一帧图像.这在一定程度上消除了不同人脸、不同序列在无表情状态下的差异.

由于输入图像序列的帧数量一般是不同的,因此需要统一对齐到相同帧数上.这项工作中使用时域插值模型 (Temporal interpolation model, TIM)<sup>[32]</sup>进行该插帧任务.将每帧图像视作高维空间中的一个点,而图像通常是高度冗余的数据,因此

存在一个对应的低维流形映射. 这种插值算法把图像序列映射到低维流形上进行插值, 再重新映射到原空间, 就得到了插值后的视频.

对于上述得到的相同帧数的对齐后的人脸图像序列, 需要提取其进一步的特征. 这项工作中使用了局部二值模式 (Local binary pattern, LBP)<sup>[33]</sup> 的一种拓展方法.

局部二值模式试图编码图像中局部像素的共生 (Co-occurrence) 模式. 以最简单的局部二值模式为例, 考虑一个像素与周围 8 个相邻像素的大小关系, 周围像素值比中心像素值大或者相等的记 1, 周围比中心小的记 0, 连接后得到一个二进制数用于表征局部像素共生模式. 图 3 展示了一个  $3 \times 3$  的图像块, 图 4 展示了该图像块上的局部二值模式的计算. 将相邻像素值减去中心像素值, 对得到的差值进一步处理, 将  $\geq 0$  的差值记 1,  $< 0$  的差值记 0, 得到 8 位二进制数 (00111010)<sub>2</sub>.

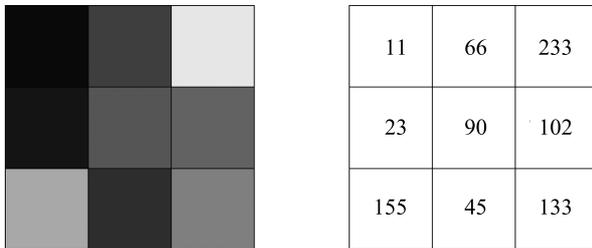


图 3 一个  $3 \times 3$  的图像块及其对应的像素值  
Fig. 3 A  $3 \times 3$  image patch and the corresponding pixel values

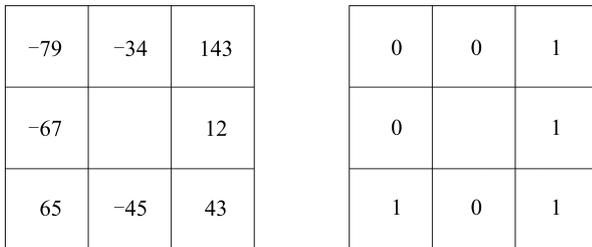


图 4 局部二值模式计算过程  
Fig. 4 Calculation process of local binary pattern

更一般地, LBP 算子可以定制两个参数, 以一个像素为中心, 作半径为  $R$  的圆, 在圆周上均匀采  $P$  个点, 计算每个点所在位置的像素与中心像素的关系, 即得到一个  $P$  位二进制数.

在微表情的识别中, 为了编码时空的共生模式, 该工作使用了 LBP-TOP (Local binary pattern on three orthogonal plane) 算子<sup>[34]</sup>, 对视频  $XY$  平面、 $XT$  平面、 $YT$  平面分别抽取 LBP 特征. 具体地, 设定三个时空轴 ( $X, Y, T$ ) 上的半径  $R_X, R_Y, R_T$

和三个时空平面上的采样数量  $P_{XY}, P_{XT}, P_{YT}$ , 在每个时空平面上作对应半径决定的椭圆, 并均匀采点, 计算该平面上的局部二值模式, 最后拼接得到最终的特征表达. 图 5 展示了一个 LBP-TOP 特征抽取的例子, 其中,  $R_X = R_Y = 3, R_T = 1, P_{XY} = 20, P_{XT} = P_{YT} = 8$ . 阴影部分是参与计算的像素.

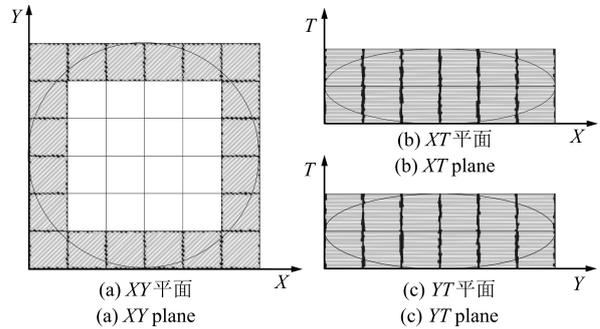


图 5 LBP-TOP 示例<sup>[34]</sup>

Fig. 5 Illustration of LBP-TOP<sup>[34]</sup>

最后, 在 LBP-TOP 特征的基础上使用支持向量机 (Support vector machine, SVM)、随机森林 (Random forest, RF) 和多核学习 (Multiple kernel learning, MKL) 等算法进行检测和分类.

该工作的算法设计较简单, 利用了常规表情分析中的很多技术, 作为微表情识别的初期尝试, 获得了不错的效果. 一个重要的优点在于预处理十分精细, 为适应微表情这一特征领域做了尝试, 也为后来的工作奠定了基础和比较的参考. 这种方法各组成部分的代码都是公开的<sup>1</sup>.

### 3.1.2 基于 STCLQP 的识别方法

完备局部量化模式 (Completed local quantized pattern, CLQP)<sup>[35]</sup> 是 LBP 的一项改进工作. 与 LBP 只编码局部像素的灰度值大小关系不同, 完备局部量化模式将中心像素与周围像素的局部共生模式分解成符号正负和幅值大小, 并加入中心像素的梯度信息, 分别用二进制数进行编码. 在构建统计直方图的阶段, 为了降低特征的维度, 完备局部量化模式并不统计所有可能的二进制编码, 而是考虑最常出现的二进制模式, 引入了向量量化的技术, 可以指定量化过程中的中心数量 (编码本中词的数量), 得到指定维度的直方图作为特征. STCLQP (Spatial temporal completed local quantized pattern)<sup>[36]</sup> 是 CLQP 在三维时空的扩展, 计算方式与 LBP-TOP 类似, 即在  $XY$  平面、 $XT$  平面和  $YT$  平面分别抽取 CLQP 特征, 并进行拼接, 作为 STCLQP 特征.

<sup>1</sup><http://www.cse.oulu.fi/CMV/Downloads/LBP Matlab>

将 STCLQP 用于微表情识别的优点在于考虑更多信息,但这不可避免地引入了高维度,使用向量量化的技术一定程度上可以缓解这一问题.

### 3.1.3 基于 LBP-SIP 的识别方法

与基于 LBP-TOP 改进的微表情识别工作不同,六交点局部二值模式 (Local binary pattern with six intersection points, LBP-SIP)<sup>[37]</sup> 从另一个角度拓展了 LBP 特征用于微表情识别. 考虑 LBP-TOP 的一种特例,即  $R = 1, P = 4$ , 此时 LBP-TOP 计算三个平面上各 4 个像素与中心像素的关系,需要用 12 位的二进制数表示. 针对于此, LBP-SIP 将中心点同平面上的四个点用作空间纹理描述,前后两帧的中心点用作时间纹理的描述. 这样仅需要  $4 + 2 = 6$  位二进制数进行描述. 图 6 展示了 LBP-SIP 的计算示例,其中阴影部分是参与计算的像素.

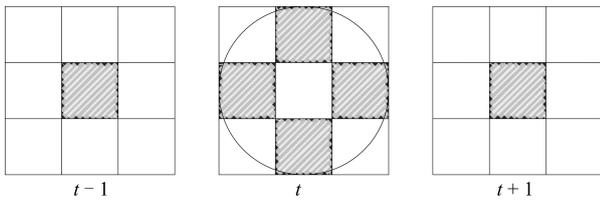


图 6 LBP-SIP 示例

Fig. 6 Illustration of LBP-SIP

这项工作的主要改进就是降低了特征的维度,提高了特征抽取的效率. 据报道,其处理速度是 LBP-TOP 的 2.8 倍<sup>[37]</sup>. 在留一人验证的实验设置下, LBP-SIP 在 5 类 CASME II 数据集和 SMIC 数据集上的最高分类准确率分别为 66.40% 和 64.02%; 在 CPU 为 Core i7、内存为 8GB 的实验环境下, CASME II 中序列的 LBP-TOP 平均抽取时间是 18.289s, LBP-SIP 抽取时间为 15.888s. 前者的识别时间为 0.584s, 后者为 0.208s.

### 3.1.4 基于 Delaunay 时域编码的识别方法

Lu 等<sup>[38]</sup> 提出了基于 Delaunay 三角化的时域编码模型 (Delaunay-based temporal coding model, DTCM). 利用主观表观模型 (Active appearance model, AAM)<sup>[39]</sup> 对人脸图像序列进行标定. 由于微表情自身的变动幅度很小, 仅用关键点不能很好地描述表情变化, 因此利用特征点将序列图像归一化, 得到特征点位置固定的人脸图像序列. Delaunay 三角化<sup>[40]</sup> 可以根据给定的特征点, 把人脸分割为一系列三角形区域. 因为特征点已经过归一化, 因此每个三角区域的大小和形状是相同的, 具有相同的像素数量. 通过对比相同区域随着时间的变化, 可以刻画微表情的动态过程. 具体地, 对第  $i$  帧的第  $j$  个三角区域, 计算如下特征:

$$x_{i,j} = \frac{\sum_{k=1}^m (T_{i,j,k} - T_{i-1,j,k})}{m} \quad (1)$$

若以  $T_{i,j}$  表示序列中第  $i$  帧的第  $j$  个三角区域的特征向量,  $T_{i,j,k}$  是该向量的第  $k$  个像素,  $m$  用于归一化计算.  $x_{i,j}$  计算每个相同三角区域在相邻帧中特征向量之差的累加值. 显然, 正值的  $x_{i,j}$  表示连续帧相同区域灰度增强, 负值表示灰度减弱, 这可能是表情变化或整体照明的变化引起的. 因此, 为了进一步编码该特征, 需要选定合理的阈值:

$$\tau_i^+ = \frac{a}{R} \sum_{r=1}^R \text{mean}\{x_{i,j,r}^+\}$$

$$\tau_i^- = \frac{a}{R} \sum_{r=1}^R \text{mean}\{x_{i,j,r}^-\} \quad (2)$$

其中,  $x_{i,j,r}^+, x_{i,j,r}^-$  表示第  $r$  段序列的第  $i$  帧的第  $j$  个区域的特征值, 加号上标表示正值的特征, 减号上标表示负值的特征;  $R$  是序列数量,  $a$  是可调参数.  $\tau_i^+, \tau_i^-$  表示基于序列特征定义的阈值. 对于第  $i$  帧的所有三角区域的特征, 特征大于  $\tau_i^+$  的区域用 1 表示, 特征小于  $\tau_i^-$  的区域用 -1 表示, 两者之间的用 0 表示. 这种方式可以用很低的维数表示每帧的表情变化, 再利用支持向量机或者随机森林就可以很好地进行识别.

这种方法引入了主观表观模型 (AAM) 和 Delaunay 三角化对人脸进行区域的划分, 只编码含有重要信息的区域, 理论上识别能力很强. 但是在所有图像上使用 AAM 进行标注可能面临标注结果不稳定的问题, 从而破坏后续处理的有效性.

### 3.1.5 基于时空梯度特征的识别方法

Polikovskiy 等<sup>[28]</sup> 使用一种梯度特征描述面部的时空局部动态. 1) 使用主观形状模型定位人脸关键点, 根据常用 FACS 编码把人脸分割为 12 个区域; 2) 对每个区域的图像进行一些预处理, 包括归一化和平滑处理. 上述过程得到预处理完毕的 12 个时空块, 每个块对应一个面部局部区域.

在一个微表情序列中, 可以在每个像素点计算三个方向 ( $X, Y, T$ ) 上的梯度. 计算 12 个区域的每一帧中所有像素的梯度并量化, 就能构建该区域中的梯度直方图, 并作为该微表情图像序列的特征.

在获得该特征后, 假定每一帧微表情图像只包含一个动作单元 (Action unit, AU), 即 FACS 的具体编码, 则该动作单元就可以作为图像的标注. 在所有图像的梯度直方图特征空间上使用  $k$ -均值算法进行聚类, 类别数量设置为所有图像中出现过的动作单元的数量. 对每一个聚类簇, 将多数特征对应的

动作单元作为类簇的真实标注。

对一个新的需要测试的微表情图像序列, 依照上述方式提取每帧的梯度直方图特征, 用上述类簇判定每一帧的动作单元, 再使用加权投票的方式决定该微表情序列的动作单元和对应情绪。

该工作的特征比较简单, 是平面梯度直方图的一种拓展. 模型构建过程采用了较为复杂的过程, 可以看作一种用  $k$ -均值算法辅助构造的  $k$  近邻模型, 该算法一定程度上对标注的正确性比较鲁棒, 对少量的错误标注不敏感. 同时也存在局限, 虽然微表情牵涉的面部肌肉数量较少, 但假设只有一个动作单元仍有可能不成立. 此时模型的识别结果是次优的。

### 3.1.6 基于 Gabor 特征的识别方法

Wu 等<sup>[41]</sup> 通过 Gabor 特征描述每一帧的面部图像, 并用 GentleSVM 识别微表情. 首先在视频的每一帧中抽取 Gabor 特征:

$$\Psi_{u,v}(\mathbf{z}) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\frac{\|k_{u,v}\|^2 \|\mathbf{z}\|^2}{2\sigma^2}} \left( e^{ik_{u,v}\mathbf{z}} - e^{-\frac{\sigma^2}{2}} \right) \quad (3)$$

其中,  $\mathbf{z}$  是一个二维向量, 表示点坐标,  $k_{u,v}$  和  $\sigma$  都是特征参数.  $\Psi_{u,v}(\mathbf{z})$  表示抽取的 Gabor 特征。

对于一段新的视频, 用上述训练得到的模型判定每一帧的情绪, 计算最长的连续具有相同情绪的帧序列. 通过视频的帧率可以计算每一段情绪的持续时间, 根据微表情的定义, 持续时间在  $1/25 \sim 1/5$  s 之间的是微表情片段, 长于  $1/5$  s 的是常规表情. 丢弃常规表情后对微表情进行进一步的分类。

这一方法的优点在于自动完成了图像序列的分割, 相比使用滑动窗口截取视频段再分类的方法降低了计算代价; 另外 Gabor 特征也具有较强的描述能力, 在很多人脸相关的问题中已经得到过验证. 但是也存在一些问题, 首先模型需要训练基于图像的微表情模型, 而目前的微表情数据集是针对视频的, 如果使用常规表情数据集, 则模型正确性可能较差; 在帧分类阶段, 部分帧可能被分错, 从而导致一个普通表情被分为多段, 则较短的时间会致使模型错判为微表情。

Zhang 等<sup>[42]</sup> 也采用 Gabor 作为微表情的特征, 采用 CASME II 作为数据集, 建立以图像为样本的模型. 另外使用 PCA 和 LDA 用作维数约简和判别分析。

### 3.1.7 基于颜色空间变换的特征增强

Wang 等<sup>[43]</sup> 探索了颜色空间对后续特征抽取的影响, 提出了张量独立颜色空间 (Tensor independent color space, TICS). 在人脸面部的图像数据中, 用 RGB 编码的图像的三个通道分量是高度相关的, 也就是三通道之间的互信息量接近零. 因

此在这样的三通道图像中进一步抽取特征 (例如常用的 LBP-TOP), 则很有可能得到几乎一致的特征表达, 并不能给识别带来提升. 在另一项工作<sup>[44]</sup> 中, 尝试了 CIELab 和 CIELuv 两种颜色空间, 这两种颜色空间在人肤色相关应用中有着较好的应用. 实验证明颜色空间的转换带来了识别效果的提升. 为了进一步利用这种效应, 他们试图通过算法寻找最优的颜色空间变换. 首先将图像序列看作 4 阶张量  $\mathcal{X} \in \mathbf{R}^{I_1 \times I_2 \times I_3 \times I_4}$ , 其中  $I_1, I_2$  是图像的尺度,  $I_3$  是图像序列的帧数,  $I_4$  是颜色通道的数量. 则问题转换为寻找一个张量在第 4 阶上的投影  $U_4^T$ , 使得变换  $\mathcal{Y}_i = \mathcal{X}_i \times_4 U_4^T$  后得到最优的颜色空间. 这一问题可以通过独立成分分析完成。

在此基础上, 比较了 RGB 空间上的 LBP-TOP 算子和优化后的颜色空间上的 LBP-TOP 算子, 证明颜色空间的优化给识别效果带来了提高。

### 3.1.8 基于 STLBP-IP 的识别方法

面部图像积分图 (Integral projection) 是计算机视觉中的一种常用技术, 通过对面部图像的像素进行横向和纵向的累加, 得到与宽度和高度相同维度的特征向量. 可以进行面部器官的定位, 人眼状态的识别等。

Huang 等<sup>[45]</sup> 拓展了积分图技术用于微表情的识别, 提出了时空局部二值模式积分图 (Spatiotemporal local binary pattern with integral projection, STLBP-IP). 首先将一段图像序列中所有帧减去某一无表情的图像, 得到一段差异图像的序列. 对每一帧图像进行横向和纵向的累加求和, 得到两个方向的积分向量  $\mathbf{V}_t$  和  $\mathbf{H}_t$ , 其中  $t$  是帧序号. 在一维的积分图上计算一维局部二值模式 (1DLBP)<sup>[46]</sup>

$$1DLBP_{t,W} = \sum_P \delta(S_t(z_p) - S_t(z_c)) 2^P \quad (4)$$

其中, 参数  $W$  为线性掩模, 作用类似于平面 LBP 中的半径长度;  $\delta$  是狄拉克函数;  $z_c$  是掩模中心像素,  $S_t(z_c)$  是其对应的值;  $z_p$  是  $z_c$  的相邻像素. 在每一帧的累加积分图上计算 1DLBP, 得到特征  $f_{XY}$ , 这是对平面图像的描述。

为了加入随时间变化的信息, 对每一帧的水平积分向量  $\mathbf{H}_t$  进行拼接, 得到  $h \times T$  的积分图图像, 其中  $h$  是图像高度,  $T$  是图像帧数. 对该图像进行归一化后计算平面 LBP 特征, 得到特征  $f_{YT}$ 。

结合  $f_{XY}$  和  $f_{YT}$  可以共同描述微表情序列, 再使用支持向量机就完成了微表情的检测和分类任务。

### 3.1.9 基于 FDM 的识别方法

Xu 等<sup>[47]</sup> 以光流场为描述微表情运动模式的基础特征, 并进一步提取更精简表达的形式, 提出了面部动力谱特征 (Facial dynamics map, FDM)。

首先抽取两帧之间的稠密光流场. 在稠密光流场 ( $U_t, V_t$ ) 的基础上, 进一步去除面部平移造成的误差. 具体地, 定义目标

$$\begin{aligned}\Delta u_t &= \arg \max_u \Phi(U_t + u \times I) \\ \Delta v_t &= \arg \max_v \Phi(V_t + v \times I)\end{aligned}\quad (5)$$

其中,  $U_t$  和  $V_t$  分别是第  $t$  帧图像和第  $t+1$  帧图像的光流场的水平和垂直分量,  $I$  是所有元素为 1 的矩阵,  $\Phi$  统计矩阵中 0 元素的数量. 即寻找水平和垂直方向上的修正量, 使得修正后的光流场中大多数元素为 0, 这是基于微表情中面部绝大部分区域是静止的这样的假设. 这样的预处理能力在像素级别上对面部进行精细化的对齐.

由于面部肌肉尺度的限制, 在局部空间和时间中, 运动向量应当是趋同的. 因此把抽取出的光流场进一步分割成小的时空立方体, 在每个立方体中使用一种迭代的算法抽取时空立方体的主方向. 用  $w_{i,j}$  表示立方体  $i, j$  坐标的光流运动向量, 则主方向的目标

$$\xi^* = \arg \max_{|\xi|=1} \sum_{i,j} \langle \xi, w_{i,j} \rangle \quad (6)$$

其中,  $w_{i,j}$  取自  $\tau$  个候选方向向量  $\{w_{i,j}^t\}_{t=1}^\tau$ ,  $\tau$  是当前时空立方体的帧数.  $w_{i,j}^t$  是上述修正后的光流场 ( $U_t, V_t$ ) 的  $(i, j)$  坐标上的运动向量.  $\tau$  表示该时空立方体包含的帧数. 上述目标选取了一个时空块中最主要的运动模式. 但上述目标方程的解需要遍历所有可能性, 计算代价过大. 因此采用了一种迭代的方式求解问题, 实验证明迭代收敛非常快.

该方法基于光流场的计算, 能较好地反应微表情的运动模式, 且计算的面部动力谱特征容易可视化, 可以对微表情的深层理解起一定的辅助作用. 作者提供了核心算法的代码<sup>2</sup>. 其瓶颈在于稠密光流场的计算时间比较长, 不适合做实时、大规模的微表情识别.

### 3.1.10 基于 MDMO 的识别方法

Liu 等<sup>[48]</sup> 在视频序列中抽取主方向, 并进一步计算面部分块中的平均光流特征, 提出了主方向平均光流特征 (Main directional mean optical flow feature, MDMO).

在进一步提取基于光流场的特征之前, 首先对面部图像帧进行操作. 先利用 DRMF 模型<sup>[49]</sup> 定位每一帧的人脸关键点, 然后对第 2 帧起的每一帧光流场进行修正, 寻找一个仿射变换矩阵, 使得每一帧的面部特征点在该矩阵变换下与第 1 帧的面部关键点差异最小化.

在特征抽取中, 定义了一种基于关键点的人脸分块规则, 将面部分割为互不重叠的 36 个区域. 同时, 逐帧抽取光流场, 然后在每一分块中提取主方向. 与 Xu 等的工作<sup>[47]</sup> 不同, 他们计算每个分区中最类似的光流场运动向量的平均值, 并作为该区域的运动特征. 具体地, 他们在每个分块中计算 HOOF (Histogram of oriented optical flow) 特征<sup>[50]</sup>, 将所有光流方向向量量化到 8 个区间, 然后作统计直方图. 并基于此计算:

$$\bar{u}_i^k = \frac{1}{|B_{\max}|} \sum_{u_i^k(p) \in B_{\max}} u_i^k(p) \quad (7)$$

其中,  $p$  表示一确定坐标,  $u_i^k(p)$  表示第  $k$  帧中第  $i$  个区域中坐标为  $p$  的点的方向向量,  $B_{\max}$  是光流统计直方图中数量最多的区间对应的方向向量集合,  $|B_{\max}|$  是其元素数量.  $\bar{u}_i^k$  是数量最多的方向的向量的平均. 由此, 可以对每个区域计算一个二维的方向向量, 并进行归一化的运算. 最后, 把方向特征分解为幅度和方向都是 36 维的两部分, 获得最终的 72 维向量.

得到的向量可以用支持向量机建模, 用于处理微表情的检测和识别任务.

### 3.1.11 基于判别式张量子空间分析的识别方法

Wang 等<sup>[51]</sup> 将微表情看作三维时空中的张量, 通过判别式的子空间学习方法 (Deterministic tensor subspace analysis) 学习最优的特征表达, 然后使用极限学习机 (Extreme learning machine, ELM)<sup>[52]</sup> 进行模式分类.

将微表情图像序列看作三阶张量  $\mathcal{X} \in \mathbf{R}^{I_1 \times I_2 \times I_3}$ , 其中  $I_1 \times I_2$  是图像的尺度,  $I_3$  是帧数. 由于微表情序列可能有不同的帧数或图像尺度, 因此需要先使用插值算法得到维度相同的张量. 如果对这些张量进行投影, 得到

$$\mathcal{Y}_i = \mathcal{X}_i \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T \quad (8)$$

投影的目的是使得变换后的张量  $\mathcal{Y}_i$  具有尽量大的类间距离和尽量小的类内距离, 具体地, 优化如下目标函数:

$$\max_{U_1, U_2, U_3} \frac{\sum_{i,j} \|\bar{\mathcal{Y}}_i - \bar{\mathcal{Y}}_j\|_F^2 B_{ij}}{\sum_c \sum_{i,j} \|\mathcal{Y}_i^c - \mathcal{Y}_j^c\|_F^2 W_{ij}^c} \quad (9)$$

其中, 分子是所有  $\mathcal{Y}$  (变换后的样本) 的类间距离, 分母是各个类的类内距离.  $C$  是类别数量,  $\bar{\mathcal{Y}}_i$  表示第  $i$  个类的算术平均;  $N_c$  是第  $c$  个类的样本数量;  $B_{ij} = \exp(-\|\bar{\mathcal{X}}_i - \bar{\mathcal{X}}_j\|_F^2)$ ,  $W_{ij}^c = \exp(-\|\mathcal{X}_i^c - \mathcal{X}_j^c\|_F^2)$ .

<sup>2</sup><http://www.iipl.fudan.edu.cn/%7Ezhangjp/sourcecode/fdm.py>

上述优化问题可以转变成广义特征值问题, 并进行迭代求解. 求得变换矩阵  $U_1, U_2, U_3$ , 对于新的测试样本, 使用求得的矩阵对其进行变换, 得到新的特征表达后使用极限学习机进行模式分类.

### 3.1.12 基于稀疏张量典型相关性分析的识别方法

Wang 等<sup>[53]</sup> 通过张量表示微表情序列及其 LBP 特征, 并在张量上进行典型相关性分析 (Sparse tensor canonical correlation analysis), 学习微表情序列本身与其 LBP 特征的关系. 将两种变换后的表达作为最终特征, 并利用最近邻算法进行分类, 性能取得了进一步的提升.

该方法融合了计算机视觉的特征和比较强的理论算法, 比单一的特征工程更着重于实际问题.

### 3.1.13 基于 MMPTR 的识别方法

Ben 等<sup>[54]</sup> 把微表情图像序列看作三阶张量, 并寻找最优投影矩阵. 试图优化类间拉普拉斯散度 (Laplacian scatter)<sup>[54]</sup>  $\phi_b^{(n)}$  与类内拉普拉斯散度  $\phi_w^{(n)}$  之差.

对于新的样本, 利用学习得到的投影矩阵对其进行变换, 得到新的张量表达形式, 然后用两种方式对其进行分类.

在第一种分类方式中, 寻找与新样本欧氏距离最接近的已有标签样本, 并将已有标签样本的标签作为新样本的分类. 即张量空间中的最近邻方法.

在第二种分类方法中, 先将变换后的张量向量化, 再在已有标签样本中进行最近邻搜索.

### 3.1.14 基于 RPCA 的识别方法

Wang 等<sup>[55]</sup> 通过 RPCA 将微表情分解成静态面部图像和动态微表情过程. 假设有一段微表情序列  $V \in \mathbf{R}^{h \times w \times f}$ , 其中  $h$  和  $w$  是图像的高和宽,  $f$  是视频的帧数. 由于微表情的动作幅度很小, 可以分解为没有表情的部分和微表情变化的部分. 用  $D \in \mathbf{R}^{hw \times f}$  表达图像序列, 即  $D$  有  $h \times w$  行和  $f$  列. 假设  $D = A + E$ , 目标优化

$$\begin{aligned} \min_{A,E} \text{rank}(A) + \|E\|_0 \\ \text{s.t. } D = A + E \end{aligned} \quad (10)$$

即最小化  $A$  的秩与  $E$  的 0 范数之和; 这是非凸的优化问题, 可以转换为最小化  $A$  的 \* 范数与  $E$  的 1 范数之和

$$\begin{aligned} \min_{A,E} \|A\|_* + \lambda \|E\|_1 \\ \text{s.t. } D = A + E \end{aligned} \quad (11)$$

通过拉格朗日法可以解该问题. 上述过程中得到的  $E$  在这里可以表征微表情的动态. 在此基础上, 使用一种改进的算子局部时空方向特征 (Local

spatiotemporal directional feature, LSDF)<sup>[55]</sup> 提取微表情动态的特征, 并用支持向量机进行分类.

## 3.2 针对微表情分类的方法

针对微表情的分类任务提出的算法, 其中不少可能也可以应用到检测任务中, 但并没有经过实验的进一步验证.

### 3.2.1 基于 CBP-TOP 的分类方法

中心化二值模式 (Centralized binary pattern, CBP)<sup>[56]</sup> 是一种针对局部二值模式的改进. 它的计算方式与 LBP 类似, 以当前像素为中心点, 作半径为  $R$  的圆, 在圆周上取均匀分布的  $P$  个点, 得到对应位置上的  $P$  个像素. 与 LBP 不同的是, 像素的取值是中心点与周围相邻点的平均值之差, 因此对应二进制编码长度大约是 LBP 的一半, 直方图维度更低. 另外编码过程中增加了一个阈值作为参数, 如果该像素值之差的绝对值超过预设阈值则记作 1, 否则记 0, 得到  $P/2 + 1$  位的二进制数. 图 7 展示了一个 CBP 计算的例子, 其中  $R = 1, P = 8$ , 其对应的图像块是图 3 中的  $3 \times 3$  像素矩阵, 其阈值参数为 50, 得到的二进制数是  $(01110)_2$ .

	21	78		0	1
	-6	-79		0	1
		122			1

图 7 中心化二值模式计算过程

Fig. 7 Calculation process of centralized binary pattern

CBP 相对于 LBP 的改进之处在于: 1) 可以获得更低维度的特征; 2) 考虑了中心点与周围像素的关系, 增加了一些信息量.

Guo 等<sup>[57]</sup> 提出用 CBP-TOP 算子代替 LBP-TOP, 实验取得了更好的效果. CBP-TOP 是 CBP 特征在三维时空的拓展, 即在  $XY$  平面、 $XT$  平面、 $YT$  平面分别计算 CBP 特征, 然后拼接得到 CBP-TOP 特征. 使用极限学习机对得到的特征进行模式分类, 处理微表情的分类问题.

### 3.2.2 基于 Riesz 小波变换的识别方法

Oh 等<sup>[58]</sup> 使用 Riesz 小波函数将图像帧进行小波变换:

$$\begin{aligned} f_{\text{mono},k}(x) = \\ (\psi_k * f(x), R_1\{\psi_k * f\}(x), R_2\{\psi_k * f\}(x)) \end{aligned} \quad (12)$$

其中,  $f(x)$  是输入图像,  $R_1, R_2$  是 Riesz 操作符,  $\psi_k$  是小波函数,  $k$  是小波尺度.

上述过程可以表达为

$$r = I + iR_x + jR_y \quad (13)$$

其中,  $I = \psi_k * f(x)$ ,  $R_x = R_1\{\psi_k * f\}(x)$ ,  $R_y = R_2\{\psi_k * f\}(x)$ ,  $*$  表示卷积操作. 由此可以计算变换后图像的幅值  $A_r$ 、方向  $\phi_r$  和相位  $\theta_r$ .

$$\begin{aligned} A_r &= \sqrt{I^2 + R_x^2 + R_y^2} \\ \phi_r &= \arctan \frac{\sqrt{R_x^2 + R_y^2}}{I} \\ \theta_r &= \arctan \frac{R_y}{R_x} \end{aligned} \quad (14)$$

在上述三种中间特征的基础上, 分别构建三者的统计直方图, 就可以得到最终的特征. 由于三者并不是来自同一数据分布, 因此使用了多核学习进行微表情的建模.

### 3.2.3 基于运动模式放大的分类方法

微表情识别的两大难点在于持续时间和动作幅度小. 前者已经通过高帧率摄像机得到一定的解决, 后者依靠精细的预处理和高分辨率的特征表达来解决. Li 等<sup>[59]</sup> 使用欧拉视频放大 (Eulerian video magnification, EVM)<sup>[60]</sup> 技术将微小的动作幅度增强放大, 再通过一些常规的识别技术对放大后的微表情进行识别.

欧拉视频放大技术是一种计算机图形学技术, 用于视频中微小变化的捕捉和放大. 其原理如下: 一段视频可以通过拉普拉斯变换得到其频域表示, 不同尺度的运动对应着不同频域的频率分量. 因此通过对某些频率波段施加带通的放大滤波器就可以增强对应尺度的运动.

具体地, 使用截断频率是 [0.3, 4] Hz 的无限脉冲响应 (Infinite impulse response, IIR) 滤波器对视频进行处理, 对微表情进行定向地放大. 然后尝试了 LBP-TOP、HOG-TOP、HIGO-TOP 三种特征提取微表情的表示. 需要注意的是, 这里处理的视频需要是微表情数据, 因此这种方法针对解决的是微表情分类的问题.

Chavali 等<sup>[61]</sup> 也使用了这种动作放大技术.

### 3.2.4 基于特定点跟踪的特定动作单元识别

Yao 等<sup>[62]</sup> 试图解决特定 AU 的识别, 即动作单元 12 (Lip corner puller) 和动作单元 16 (Lip corner depressor) 的识别. 这两种动作单元对应着愉悦和厌恶两种重要的情绪, 因此有一定的应用价值. 他们的方法首先使用一套精细定义的规则定位人脸,

再基于人脸寻找嘴角的位置. 用 TLD (Tracking-learning-detection) 跟踪器<sup>[63]</sup> 追踪嘴角的位置, 识别嘴角的运动模式, 就可以区分这两种动作单元.

## 3.3 针对微表情检测的方法

微表情的检测任务在逻辑上要比分类任务更优先, 只有通过检测任务筛选的片段才会进行进一步的分类. 因此检测算法的质量直接关系到后续处理的有效性. 在检测任务中, 可以再细分为两种. 第一种是一个简单的二分类问题, 给定一段较短的面部图像序列, 算法只要判断该序列是否是一个微表情即可. 这之前介绍的方法类似; 第二种则更适用于实际应用, 给定一段长视频, 算法需要从中找出微表情开始和结束的时间点. 微表情的阶段分割任务可以看作是后者的进一步工作, 但目前针对该问题的的工作很少.

### 3.3.1 基于几何形变建模的检测方法

Xia 等<sup>[64]</sup> 通过对几何形变进行建模, 解决微表情检测的问题. 对于一段人脸图像序列, 首先使用 STASM (Active shape models with SIFT descriptor)<sup>[65]</sup> 进行关键点的定位, 得到每帧的面部形状. 为了消除头部移动造成的变化, 将每帧与第 1 帧进行 Procruste 变换<sup>[30]</sup>, 即通过平移、缩放、旋转最小化与第 1 帧之间的误差, 完成这一处理后的差别就是面部形变造成的.

对每一帧的特征点, 计算两部分特征, 即

$$\begin{aligned} \mathbf{v}_i^b &= (\mathbf{z}_i - \mathbf{z}_b) \\ \mathbf{v}_i^d &= (\mathbf{z}_i - \mathbf{z}_a) \end{aligned} \quad (15)$$

其中,  $\mathbf{z}_i$  是第  $i$  帧的面部形状向量,  $\mathbf{z}_b$  是基线面部形状向量,  $\mathbf{z}_a$  是时间区间为  $[i - L/2, i + L/2]$  的帧中的面部形状向量的算术平均. 这三者的维度都是面部特征点数量的两倍, 对应每个特征点的二位平面坐标. 这两部分特征分别表达了瞬时表情自身的绝对动态及其在一段序列中的相对动态. 将两种特征融合得到

$$v_i = \frac{\beta v_i^b}{\sigma(v^b)} + \frac{(1 - \beta) v_i^d}{\sigma(v^d)} \quad (16)$$

其中,  $\sigma$  表示在整段序列之中取方差,  $\beta$  和  $1 - \beta$  分别是两部分的权重. 该式计算的  $v_i$  表示两部分差异特征的加权平均.

随后通过随机过程对每一帧是否含有微表情进行建模:

$$p_{t+1}(i) = \alpha \sum_{j \in \Omega_{f_i}} p_t(j) \Phi(i, j) + (1 - \alpha) p_0(i) \quad (17)$$

其中,  $\Phi(i, j)$  是基于上述特征计算的转移概率, 由训练数据集估算得到;  $\Omega_{f_i}$  是第  $i$  帧的前后近邻;  $p_t(i)$

表示第  $t$  轮计算中第  $i$  帧含有微表情的概率, 迭代计算 20 次后根据概率取阈值判定是否含有微表情.

### 3.3.2 基于特征差异的检测方法

计算面部图像连续帧的计算机视觉特征, 考察短时间内的特征变化, 常规表情与微表情的变化应当是不同的. Moilanen 等<sup>[66]</sup> 基于这样的原则使用手工设计的特征和准则处理微表情的检测问题.

把面部图像分割成  $6 \times 6 = 36$  个图像块, 在每个图像块中计算 LBP 直方图特征. 为了衡量一帧图像在一段图像序列中的突变程度, 计算与当前帧 (Current frame, CF) 各相差  $k$  帧的前后两帧 (分别为 Tail frame (TF) 和 Head frame (HF)) 的特征平均值, 通过 CF 的特征与该平均值的卡方距离衡量当前帧在视频中的变化程度

$$\chi(\mathbf{P}, \mathbf{Q}) = \sum_i \frac{(P_i - Q_i)^2}{P_i + Q_i} \quad (18)$$

其中,  $\mathbf{P}$  和  $\mathbf{Q}$  是两个待计算差异的直方图向量,  $i$  是直方图区间序号.

对于每一帧, 以差异值最大的三分之一的图像块的差异和作为该帧的进一步特征  $F_t$  ( $t$  是帧序号), 并考虑前后帧之间的关系, 计算

$$C_t = F_t - \frac{1}{2}(F_{t+k} + F_{t-k}) \quad (19)$$

即对差异序列  $F_t$  进行时间平滑. 经过上述过程计算得到的特征  $C_t$  可以表征人脸图像序列的变化程度. 在此基础上, 通过手工选择阈值  $T = C_{\text{mean}} + p \times (C_{\text{max}} - C_{\text{mean}})$  就可以判定微表情的出现, 其中  $p$  是一个  $[0, 1]$  之间的百分数,  $C_{\text{max}}$  和  $C_{\text{mean}}$  分别表示  $C_t$  序列的最大值和平均值.

这种方法比较巧妙, 实验证明具有效果, 但是设计思路较为复杂, 同时需要手工设定参数, 在实际应用中可能需要重新设定参数、阈值等.

### 3.3.3 基于光流场积分的阶段分割方法

Patel 等<sup>[67]</sup> 试图对一段微表情图像序列标定其起始点 (Onset)、顶点 (Apex) 和终止点 (Offset). 首先用 DRMF (Discriminative response map fitting) 模型<sup>[49]</sup> 定位人脸的关键点, 再基于 FACS 的规则将关键点进行分组, 即相同面部器官上的关键点归为同一组, 例如左眉毛、右眉毛、嘴部等. 再抽取逐帧之间的光流场, 则每个特征点都分配到对应的运动向量, 每一分组的平均运动可以由其中点的运动向量取算术平均获得. 对每一分组, 计算其随着时间的运动幅值的累加值, 通过寻找累加值的顶点, 则可以得到其对应的动作单元的顶点. 在顶点的基础上, 寻找起始点可以看作一个优化问题: 在起止点之前的帧到顶点所在帧, 其运动幅值累加值在逐渐

增长; 而起始点所在帧开始, 到顶点所在帧, 其运动幅值是逐渐减小的. 通过这一准则, 可以寻找到合理的起始点位置. 终止点位置的寻找基于这样的规则: 顶点之后第一个与起始点的点积为负的帧即是终止点.

### 3.3.4 基于特征差异的微表情顶点定位

Yan 等<sup>[68]</sup> 利用特征差异定位微表情顶点, 具体使用了两种特征.

在基于受限局部模型 (Constrained local model, CLM)<sup>[69]</sup> 的方法中, 利用该模型定位人脸的 66 个关键点. 从第 2 帧起计算每一帧的特征向量相对第 1 帧特征向量的累计偏差, 寻找该偏差的峰值, 则对应为微表情的顶点.

在基于局部二值模式的方法中, 先利用 CLM 定位人脸关键点, 并划分出若干关键区域, 计算关键区域中的局部二值模式直方图作为每帧的特征向量. 类似地, 计算每帧的特征向量与第 1 帧的相关性

$$d = \frac{\sum_{i=1}^{nBins} h_{1i} \times h_{2i}}{\sqrt{\sum_{i=1}^{nBins} h_{1i}^2 \times \sum_{i=1}^{nBins} h_{2i}^2}} \quad (20)$$

其中,  $\mathbf{h}_1$  和  $\mathbf{h}_2$  分别是两个不同的直方图向量,  $i$  是直方图区间序号,  $nBins$  是直方图长度,  $d$  是两个直方图向量的相关性, 其峰值对应微表情的顶点.

这种方法标定的顶点与人识别结果对比, 取得了较好的效果.

### 3.3.5 基于 Strain Tensor 的检测方法

Shreve 等<sup>[27, 70]</sup> 通过计算运动强度来检测微表情.

首先使用主观形状模型定位面部特征点并切割出小的矩形区域. 然后, 计算每个区域中的光流场, 并以此为基础计算张力张量 (Strain tensor), 用于度量一块区域中的人脸部位相对于时间的变化强度. 具体地, 首先计算两帧图像之间的光流场, 得到每个点的运动向量  $[u, v]^T$ , 并计算该点的有限张力张量 (Finite strain tensor)

$$\epsilon = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{1}{2} \left( \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \\ \frac{1}{2} \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) & \frac{\partial v}{\partial y} \end{bmatrix} \quad (21)$$

其中,  $\mathbf{u}$ ,  $\mathbf{v}$  是每个点的运动向量.  $\frac{\partial u}{\partial x}$ ,  $\frac{\partial u}{\partial y}$ ,  $\frac{\partial v}{\partial x}$ ,  $\frac{\partial v}{\partial y}$  分别表示水平和垂直方向上的运动向量对水平和垂直方向的偏导. 上述矩阵的元素之和称为张力幅值, 用于表征该点的运动强度. 因此, 可以通过一块区域中所有像素位置上的张力幅值来考察面部的局部动作.

这种方法可以检查微表情是否存在,也可以用来区分常规表情与微表情.然而这种方法没有使用机器学习技术,需要通过人工选定阈值.在训练数据较大的情况下不适用,而当训练数据较小时,阈值选择的有效性又很难保障.

Liong 等<sup>[71]</sup>使用类似的方法,加入时间平滑的技巧.即对每一帧的每个像素,计算前后若干帧在该点的张力张量的平均值.将由此得到的特征用支持向量机建模,进行微表情的检测和分类.

## 4 对比与总结

### 4.1 实验结果对比

为了给出现有方法的直观比较,表 2 选取了在公共数据集 CASME 和 CASME II 上验证过的若干方法的结果作为对比.由于数据集的类别间分布并不均衡,现有工作通常选择其中的几类进行验证;不少工作选择“留一人验证”的方式,即每次选出一名被试的样本作为测试集,其余用作训练.也有部分工作选择了其他的验证方式.表 2 注明了每个结果使用的数据集类别数量,以及相应的验证方式.其中 LBP-TOP、STCLQP、LBP-SIP 的结果取自文献<sup>[72]</sup>;DTSA 的验证方式是在每个类别中随机选 15 个样本作为训练集,其余作为测试集;MMPTR 的验证方式是随机选 15 个样本作为训练集,其余作为测试集;其余方法使用留一人验证.

表 2 现有微表情识别方法的识别准确率 (%) 对比  
Table 2 Recognition accuracy (%) of existing approaches on common datasets

方法	CASME	CASME II
LBP-TOP <sup>[23]</sup>	37.43 (4 类)	46.46 (5 类)
STCLQP <sup>[36]</sup>	57.31 (4 类)	58.39 (5 类)
LBP-SIP <sup>[37]</sup>	36.84 (4 类)	46.56 (5 类)
DTCM <sup>[38]</sup>	64.95 (4 类)	N/A
TICS <sup>[43-44]</sup>	61.86 (4 类)	62.30 (4 类)
STLBP-IP <sup>[45]</sup>	N/A	59.51 (5 类)
FDM <sup>[47]</sup>	56.14 (5 类)	45.93 (5 类)
MDMO <sup>[48]</sup>	64.07 (4 类)	57.16 (4 类)
DTSA <sup>[51]</sup>	46.90 (5 类)	N/A
MMPTR <sup>[54]</sup>	N/A	80.2 (4 类)
RPCA + LSDF <sup>[55]</sup>	N/A	65.45 (4 类)
Riesz 小波 <sup>[58]</sup>	N/A	46.15 (4 类)
EVM <sup>[59]</sup>	N/A	67.21 (4 类)

### 4.2 现有方法总结

Pfister 等<sup>[23]</sup>是微表情识别最早的尝试者之一,他们将传统的面部表情的识别方法应用到微表情识别中,根据微表情本身的特性做了一定的适应.后期不少工作都致力于在特征的层面上改进微表情的识别性能<sup>[36-37,57]</sup>.这些工作都取得了不错的性能改进,其共性在于从时空纹理的角度挖掘面部表情的变化,具有很强的描述能力,但是计算得到的特征的可解释性欠佳.

近年来出现了不少从运动角度描述微表情的工作<sup>[47-48,67]</sup>,在保证识别性能的前提下,给出了良好的可解释性.然而,基于稠密光流场的特征耗时较长,对于微表情这样仅持续很短时间的面部运动显得代价过大,几乎无法应用到实时检测中.

除了上述从人工特征出发的工作,我国的王甦菁、贲晔分别进行了不少从理论算法上优化微表情特征表达的工作<sup>[44,51,53-54]</sup>.这些方法具有很强的理论性,同时不只适用于微表情的特征表达,也可用于普通面部运动甚至其他计算机视觉的识别分类问题.

上述两大类方法并不冲突,而是可以共同使用.例如文献<sup>[44]</sup>在 TICS 变换后的图像上提取 LBP-TOP 特征;文献<sup>[54]</sup>也可用于在已有特征下进一步计算表达.

此外,早期工作的假设较为简单,只在分割好的面部表情序列上进行检测和分类,而微表情分割本身是一个困难的问题.近年来有不少工作研究在长视频中检测微表情并进行进一步分割<sup>[64,66-67]</sup>,这对于微表情识别走向实用化是必不可少的.

表 3 比较了现有的微表情学习算法.

## 5 未来研究的可能方向

近年来微表情识别领域涌现了大量新的研究工作,这是对微表情的利用价值的肯定.可以预见,未来会有更多的工作尝试进一步提高微表情的识别性能,并逐渐将微表情识别应用到实际中.本文总结了现有技术的一些问题和未来可能的研究方向.

### 5.1 针对微表情的预处理技术

利用现有数据集的一个好处是可以直接在预处理好的图像上尝试新的算法,减轻了预处理流程的压力.然而预处理是微表情识别中非常重要的一项流程,其重要程度应当超过普通的表情识别或其他面部信息识别,其原因就是微表情的动作幅度很小,且持续时间太短.现有的数据集的录制过程中,被试位置相对稳定,在经过进一步的对齐处理,较容易得到高质量的图像.但在实际应用中是很难做到的,因此针对微表情的精细化预处理值得研究.

表 3 现有微表情识别方法  
Table 3 Existing approaches for microexpression recognition

方法	预处理方法	特征表达	学习算法	解决问题
文献 [23]	ASM、LWM、TIM	LBP-TOP	SVM、RF、MKL	检测/分类
文献 [27, 70]	ASM	Strain tensor	阈值	检测
文献 [28]	ASM、面部分块	时空梯度	近邻投票	检测/分类
文献 [36]	ASM、TIM	STCLQP	SVM	检测/分类
文献 [37]	N/A	LBP-SIP	SVM	检测/分类
文献 [38]	AAM	DTCM	SVM、RF	检测/分类
文献 [41]	ASM	Gabor	GentlSVM	检测/分类
文献 [42]	N/A	Gabor	LDA、PCA、SVM	分类
文献 [43–44]	ASM	TICS + LBP-TOP	SVM	检测/分类
文献 [45]	N/A	STLBP-IP	SVM	检测/分类
文献 [47]	光流场修正	FDM	SVM	检测/分类
文献 [48]	DRMF、光流场对齐	MDMO	SVM	检测/分类
文献 [51]	N/A	DTSA 变换	ELM	检测/分类
文献 [53]	ASM	STCCA	最近邻	检测/分类
文献 [54]	N/A	MMPTR 变换	最近邻	检测/分类
文献 [55]	ASM、面部分块	RPCA + LSDF	SVM	检测/分类
文献 [57]	面部检测、分割	CBP-TOP	ELM	分类
文献 [58]	N/A	Riesz 小波	MKL	分类
文献 [59]	ASM、LWM、TIM	EVM 处理后提前特征	SVM	分类
文献 [62]	N/A	特征点追踪	MKL	特定 AU 识别
文献 [64]	STASM、Procuste 分析	几何形变特征	随机过程计算概率	检测
文献 [66]	特征点定位	基于 LBP 的差异特征	阈值	检测
文献 [67]	DRMF	光流场	基于规则	阶段分割
文献 [68]	CLM	CLM、LBP	基于规则	顶点定位
文献 [71]	N/A	时间平滑的 Optical Strain	SVM	检测/分类

另外, 微表情预处理中应用到一些常见技术, 有比较多的候选项, 但是其中哪种最适合对微表情进行精细化的处理还少有研究. 例如人脸关键点定位几乎在所有工作中都会涉及, 候选技术有 ASM、AAM、DRMF 等; 又如帧数对齐, 有些工作使用 TIM 技术, 但是 Xu 等<sup>[47]</sup>报道了线性插值对微表情有更好的适应性, 因为微表情帧率较高, 相邻帧之间时间间隔很短, 基于流形的插帧方式反而可能造成较线性插值更高的误差.

确立一组对微表情有效的预处理流程, 可能在后续算法不变的情况下取得更好的效果.

## 5.2 长视频中的微表情检测

目前很多研究工作基于现有的数据集, 而数据

集中的图像序列已经预先进行了分割, 因此提出的算法只需要完成检测和分类两种模式识别的任务. 而实际任务中, 通常要分析长视频中对象的表情和微表情, 上述的技术模式很难处理这样的问题. 最简单的弥补方式是引入滑动窗口, 对窗口内的子序列进行检测和分类. 但这会极大地提升计算量, 且窗口的长度难以确定. 因此, 在长视频中高效地检测微表情的出现是一项非常关键的技术, 这将构成之后进一步检测和分类的输入. 这种任务和微表情序列的阶段划分有些类似, 而阶段划分需要寻找起始点、顶点和终止点.

目前已经有一些工作尝试解决这种问题, 但也存在局限. 例如 Shreve 等<sup>[27, 70]</sup>和 Moilanen 等<sup>[66]</sup>的工作都需要手工设定阈值作判断, 可能因个人差

异等因素,在真实场景中出现失效的问题. Wu 等<sup>[41]</sup>对每一帧进行检测,然后通过微表情的时间特性进行序列的区分,在单帧识别率较低时会造成整体检测的失效.

### 5.3 高效的微表情识别

对于一段微表情图像序列进行分类,识别其中的对应情绪,是微表情最有可能的实际应用之一.由于一段微表情序列的实际时间非常短,只有  $1/25\text{ s} \sim 1/3\text{ s}$ ,对处理速度提出了很高的要求.在单机实时处理场景下,如果不能以类似的时间对一段微表情图像序列进行分类,就可能积累更多的待处理任务,导致低效的系统.高效的方法总是受欢迎的,有两种应用场景需要极端高效的处理算法:1)在嵌入式设备或者移动终端上,只有非常高效的方法能胜任;2)将识别算法部署在服务器上,并以服务的方式向不同的终端提供,则高效的算法意味着更少的硬件投入.

目前还较少有工作探讨这方面的性能,而近期出现的一些关于光流场计算的方法虽然有着很好的识别率,但是不可避免地要在时间性能上有待提高.

### 5.4 微表情动作单元的识别

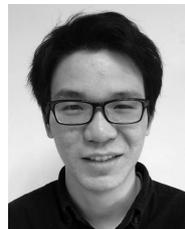
动作单元检测是常规表情识别中一项重要的子任务,但在微表情中对动作单元进行检测面临着更大的难度.然而这并不是没有价值的.正确的识别动作单元可以作为情绪识别结果的强有力证据,使情绪识别结果具有可解释性.另外,心理学对微表情的各种性质仍在研究之中,不少结论尚存在争议.动作单元级别的精细化识别,结合可视化的标定技术,可以作为心理学中微表情研究的重要依据,具有跨学科辅助研究的意义.

## References

- 1 Shan C F, Gong S G, McOwan P W. Facial expression recognition based on local binary patterns: a comprehensive study. *Image and Vision Computing*, 2009, **27**(6): 803–816
- 2 Rahulamathavan Y, Phan R C W, Chambers J A, Parish D J. Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. *IEEE Transactions on Affective Computing*, 2013, **4**(1): 83–92
- 3 Wang S F, Liu Z L, Wang Z Y, Wu G B, Shen P J, He S, Wang X F. Analyses of a multimodal spontaneous facial expression database. *IEEE Transactions on Affective Computing*, 2013, **4**(1): 34–46
- 4 Sun Xiao, Pan Ting, Ren Fu-Ji. Facial expression recognition using ROI-KNN deep convolutional neural networks. *Acta Automatica Sinica*, 2016, **42**(6): 883–891 (孙晓, 潘汀, 任福继. 基于 ROI-KNN 卷积神经网络的面部表情识别. *自动化学报*, 2016, **42**(6): 883–891)
- 5 Liu Shuai-Shi, Tian Yan-Tao, Wang Xin-Zhu. Illumination-robust facial expression recognition based on symmetric bilinear model. *Acta Automatica Sinica*, 2012, **38**(12): 1933–1940 (刘帅师, 田彦涛, 王新竹. 基于对称双线性模型的光照鲁棒性人脸表情识别. *自动化学报*, 2012, **38**(12): 1933–1940)
- 6 Liu Shuai-Shi, Tian Yan-Tao, Wan Chuan. Facial expression recognition method based on Gabor multi-orientation features fusion and block histogram. *Acta Automatica Sinica*, 2011, **37**(12): 1455–1463 (刘帅师, 田彦涛, 万川. 基于 Gabor 多方向特征融合与分块直方图的人脸表情识别方法. *自动化学报*, 2011, **37**(12): 1455–1463)
- 7 Taheri S, Patel V M, Chellappa R. Component-based recognition of faces and facial expressions. *IEEE Transactions on Affective Computing*, 2013, **4**(4): 360–371
- 8 El Mostafa M K A, Levine M D. Fully automated recognition of spontaneous facial expressions in videos using random forest classifiers. *IEEE Transactions on Affective Computing*, 2014, **5**(2): 141–154
- 9 Ekman P. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 2003, **1000**: 205–221
- 10 Haggard E A, Isaacs K S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. *Methods of Research in Psychotherapy*. US: Springer, 1966. 154–165
- 11 Ekman P, Friesen W. Nonverbal Leakage and Clues to Deception. Technical Report, DTIC Document, 1969.
- 12 Gottman J M, Levenson R W. A two-factor model for predicting when a couple will divorce: exploratory analyses using 14-year longitudinal data. *Family Process*, 2002, **41**(1): 83–96
- 13 Salter F, Grammer K, Rikowski A. Sex differences in negotiating with powerful males. *Human Nature*, 2005, **16**(3): 306–321
- 14 Whitehill J, Serpell Z, Lin Y C, Foster A, Movellan J R. The faces of engagement: automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 2014, **5**(1): 86–98
- 15 Pool L D, Qualter P. Improving emotional intelligence and emotional self-efficacy through a teaching intervention for university students. *Learning and Individual Differences*, 2012, **22**(3): 306–312
- 16 Porter S, ten Brinke L. Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 2008, **19**(5): 508–514
- 17 Warren G, Schertler E, Bull P. Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior*, 2009, **33**(1): 59–69
- 18 Yan W J, Wu Q, Liang J, Chen Y H, Fu X L. How fast are the leaked facial expressions: the duration of micro-expressions. *Journal of Nonverbal Behavior*, 2013, **37**(4): 217–230
- 19 Ekman P. *MicroExpression Training Tool (METT)*. University of California, San Francisco, 2002.
- 20 Frank M G, Herbasz M, Sinuk K, Keller A, Nolan C. I see how you feel: training laypeople and professionals to recognize fleeting emotions. In: Proceedings of the 2009 Annual Meeting of the International Communication Association. New York, 2009. [http://www.allacademic.com/meta/p15018\\_index.htm](http://www.allacademic.com/meta/p15018_index.htm)

- 21 Wu Qi, Shen Xun-Bing, Fu Xiao-Lan. Micro-expression and its applications. *Advances in Psychological Science*, 2010, **18**(9): 1359–1368  
(吴奇, 申寻兵, 傅小兰. 微表情研究及其应用. 心理科学进展, 2010, **18**(9): 1359–1368)
- 22 Ben Xian-Ye, Yang Ming-Qiang, Zhang Peng, Li Juan. Survey on automatic micro expression recognition methods. *Journal of Computer-Aided Design and Computer Graphics*, 2014, **26**(9): 1385–1395  
(贲峴烨, 杨明强, 张鹏, 李娟. 微表情自动识别综述. 计算机辅助设计与图形学学报, 2014, **26**(9): 1385–1395)
- 23 Pfister T, Li X B, Zhao G Y, Pietikäinen M. Recognising spontaneous facial micro-expressions. In: Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 1449–1456
- 24 Li X B, Pfister T, Huang X H, Zhao G Y, Pietikäinen M. A spontaneous micro-expression database: inducement, collection and baseline. In: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, China: IEEE, 2013. 1–6
- 25 Yan W J, Wu Q, Liu Y J, Wang S J, Fu X L. CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, China: IEEE, 2013. 1–7
- 26 Yan W J, Li X B, Wang S J, Zhao G Y, Liu Y J, Chen Y H, Fu X L. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS One*, 2014, **9**(1): e86041
- 27 Shreve M, Godavarthy S, Goldgof D, Sarkar S. Macro and micro-expression spotting in long videos using spatiotemporal strain. In: Proceedings of the 2011 IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Santa Barbara, CA, USA: IEEE, 2011. 51–56
- 28 Polikovskiy S, Kameda Y, Ohta Y. Facial micro-expression detection in hi-speed video based on facial action coding system (FACS). *IEICE Transactions on Information and Systems*, 2013, **E96-D**(1): 81–92
- 29 Ekman P, Friesen W V. *Facial Action Coding System*. Palo Alto: Consulting Psychologists Press, 1977.
- 30 Cootes T F, Taylor C J, Cooper D H, Graham J. Active shape models-their training and application. *Computer Vision and Image Understanding*, 1995, **61**(1): 38–59
- 31 Goshtasby A. Image registration by local approximation methods. *Image and Vision Computing*, 1998, **6**(4): 255–261
- 32 Zhou Z H, Zhao G Y, Pietikäinen M. Towards a practical lipreading system. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA: IEEE, 2011. 137–144
- 33 Ojala T, Pietikäinen M, Maenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7): 971–987
- 34 Zhao G Y, Pietikäinen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(6): 915–928
- 35 Huang X H, Zhao G Y, Hong X P, Pietikäinen M, Zheng W M. Texture description with completed local quantized patterns. *Image Analysis*. Berlin Heidelberg: Springer, 2013. 1–10
- 36 Huang X H, Zhao G Y, Hong X P, Zheng W M, Pietikäinen M. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing*, 2016, **175**: 564–578
- 37 Wang Y D, See J, Phan P C W, Oh Y H. LBP with six intersection points: reducing redundant information in LBP-TOP for micro-expression recognition. In: Proceedings of the 12th Conference on Computer Vision, Singapore. Singapore: Springer, 2014. 21–23
- 38 Lu Z Y, Luo Z Q, Zheng H C, Chen J K, Li W H. A delaunay-based temporal coding model for micro-expression recognition. *Computer Vision-ACCV Workshops*. Switzerland: Springer International Publishing, 2014.
- 39 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
- 40 Barber B C, Dobkin D P, Huhdanpää H. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 1996, **22**(4): 469–483
- 41 Wu W, Shen X B, Fu X L. The machine knows what you are hiding: an automatic micro-expression recognition system. In: Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction. Memphis, TN, USA: Springer-Verlag, 2011. 152–162
- 42 Zhang P, Ben X Y, Yan R, Wu C, Guo C. Micro-expression recognition system. *Optik — International Journal for Light and Electron Optics*, 2016, **127**(3): 1395–1400
- 43 Wang S J, Yan W J, Li X B, Zhao G Y, Fu X L. Micro-expression recognition using dynamic textures on tensor independent color space. In: Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE, 2014. 4678–4683
- 44 Wang S J, Yan W J, Li X B, Zhao G Y, Zhou C G, Fu X L, Yang M H, Tao J H. Micro-expression recognition using color spaces. *IEEE Transactions on Image Processing*, 2015, **24**(12): 6034–6047
- 45 Huang X H, Wang S J, Zhao G Y, Pietikäinen M. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. In: Proceedings of the 2015 IEEE International Conference on Computer Vision Workshops. Santiago, Chile: IEEE, 2015. 1–9
- 46 Houam L, Hafiane A, Boukrouche A, Lespessailles E, Jennane R. One dimensional local binary pattern for bone texture characterization. *Pattern Analysis and Applications*, 2014, **17**(1): 179–193
- 47 Xu F, Zhang J P, Wang J Z. Microexpression identification and categorization using a facial dynamics map. *IEEE Transactions on Affective Computing*, **PP**(99): 1–1, DOI: 10.1109/TAFFC.2016.2518162
- 48 Liu Y J, Zhang J K, Yan W J, Wang S J, Zhao G Y, Fu X L. A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 2016, **7**(4): 299–310
- 49 Asthana A, Zafeiriou S, Cheng S Y, Pantic M. Robust discriminative response map fitting with constrained local models. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013. 3444–3451

- 50 Chaudhry R, Ravichandran A, Hager G, Vidal R. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida: IEEE, 2009. 1932–1939
- 51 Wang S J, Chen H L, Yan W J, Chen Y H, Fu X L. Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. *Neural Processing Letters*, 2014, **39**(1): 25–43
- 52 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications. *Neurocomputing*, 2006, **70**(1–3): 489–501
- 53 Wang S J, Yan W J, Sun T K, Zhao G Y, Fu X L. Sparse tensor canonical correlation analysis for micro-expression recognition. *Neurocomputing*, 2016, **214**: 218–232
- 54 Ben X Y, Zhang P, Yan R, Yang M Q, Ge G D. Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation. *Neural Computing and Applications*, 2015, **127**(3): 1–18
- 55 Wang S J, Yan W J, Zhao G Y, Fu X L, Zhou C G. Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features. *Computer Vision — ECCV 2014 Workshops*. Switzerland: Springer International Publishing, 2014.
- 56 Fu X F, Wei W. Centralized binary patterns embedded with image euclidean distance for facial expression recognition. In: Proceedings of the 4th International Conference on Natural Computation. Jinan, China: IEEE, 2008. 115–119
- 57 Guo Y C, Xue C H, Wang Y Z, Yu M. Micro-expression recognition based on CBP-TOP feature with ELM. *Optik — International Journal for Light and Electron Optics*, 2015, **126**(23): 4446–4451
- 58 Oh Y H, Le Ngo A C, See J, Liong S T, Phan R C W, Ling H C. Monogenic riesz wavelet representation for micro-expression recognition. In: Proceedings of the 2015 IEEE International Conference on Digital Signal Processing. Singapore: IEEE, 2015. 1237–1241
- 59 Li X B, Hong X P, Moilanen A, Huang X H, Pfister T, Zhao G Y, Pietikäinen M. Reading hidden emotions: spontaneous micro-expression spotting and recognition. arXiv Preprint arXiv: 1511.00423 [Online], available: <https://arxiv.org/abs/1511.00423>, February 20, 2017
- 60 Wu H Y, Rubinstein M, Shih E, Guttag J, Durand F, Freeman W T. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, 2012, **31**(4): 65
- 61 Chavali G K, Bhavaraju S K N V, Adusumilli T, Puripanda V. Micro-expression Extraction for Lie Detection Using Eulerian Video (Motion and Color) Magnification [Master dissertation], Blekinge Institute of Technology, Swedish, 2014.
- 62 Yao S Q, He N, Zhang H Q, Yoshie O. Micro-expression recognition by feature points tracking. In: Proceedings of the 10th International Conference on Communications. Bucharest, Romania: IEEE, 2014. 1–4
- 63 Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(7): 1409–1422
- 64 Xia Z Q, Feng X Y, Peng J Y, Peng X L, Zhao G Y. Spontaneous micro-expression spotting via geometric deformation modeling. *Computer Vision and Image Understanding*, 2016, **147**: 87–94
- 65 Milborrow S, Nicolls F. Active shape models with SIFT descriptors and MARS. In: Proceedings of the 2014 International Conference on Computer Vision Theory and Applications. Lisbon, Portugal: IEEE, 2014. 380–387
- 66 Moilanen A, Zhao G Y, Pietikäinen M. Spotting rapid facial movements from videos using appearance-based feature difference analysis. In: Proceedings of the 2nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE, 2014. 1722–1727
- 67 Patel D, Zhao G Y, Pietikäinen M. Spatiotemporal integration of optical flow vectors for micro-expression detection. *Advanced Concepts for Intelligent Vision Systems*. Switzerland: Springer International Publishing, 2015. 369–380
- 68 Yan W J, Wang S J, Chen Y H, Zhao G Y, Fu X L. Quantifying micro-expressions with constraint local model and local binary pattern. *Computer Vision — ECCV 2014 Workshops*. Switzerland: Springer International Publishing, 2014.
- 69 Cristinacce D, Cootes T F. Feature detection and tracking with constrained local models. In: Proceedings of the 2006 BMVC. Edinburgh: BMVA, 2006. 929–938
- 70 Shreve M, Godavarthy S, Manohar V, Goldgof D, Sarkar S. Towards macro- and micro-expression spotting in video using strain patterns. In: Proceedings of the 2009 IEEE Workshop on Applications of Computer Vision. Snowbird, UT, USA: IEEE, 2009. 1–6
- 71 Liong S T, Phan R C W, See J, Oh Y H, Wong K. Optical strain based recognition of subtle emotions. In: Proceedings of the 2014 International Symposium on Intelligent Signal Processing and Communication Systems. Kuching, Sarawak, Malaysia: IEEE, 2014. 180–184
- 72 House C, Meyer R. Preprocessing and descriptor features for facial micro-expression recognition [Online], available: [https://web.stanford.edu/class/ee368/Project\\_Spring\\_1415/Reports/House\\_Meyer.pdf](https://web.stanford.edu/class/ee368/Project_Spring_1415/Reports/House_Meyer.pdf), February 20, 2017



徐峰 复旦大学计算机科学技术学院  
硕士研究生. 主要研究方向为计算机视觉,  
人脸表情识别.

E-mail: feng\_xu@fudan.edu.cn

(XU Feng Master student at the  
School of Computer Science, Fudan  
University. His research interest covers  
computer vision and facial expres-

sion recognition.)



张军平 复旦大学计算机科学技术学院  
教授. 主要研究方向为机器学习, 智能交  
通, 生物认证与图像识别. 本文通信作者.

E-mail: jpzhang@fudan.edu.cn

(ZHANG Jun-Ping Professor at  
the School of Computer Science, Fudan  
University. His research interest covers  
machine learning, intelligent trans-  
portation systems, biometric authentication,  
and image processing. Corresponding author  
of this paper.)