

# 基于多视觉描述子及音频特征的 动态序列人脸表情识别

李宏菲<sup>1,2</sup>, 李 庆<sup>1,2</sup>, 周 莉<sup>1,2</sup>

(1. 中国科学院微电子研究所, 北京 100029; 2. 中国科学院大学, 北京 100049)

**摘 要:** 关于面部表情识别的应用也正在渗透至各个领域,如安全驾驶、商品销售、临床医学等等. 本文对面部表情识别相关技术进行研究,主要工作及贡献如下:研究非约束条件下人脸动态表情识别,提出了一种基于多视觉描述子及音频特征融合策略的动态表情识别算法. 借助多视觉描述子的空时局部特征描述实现动态表情特征的提取;而视频、音频特征的融合策略改善了表情识别性能. 基于协方差矩阵及时间轴分段的动态规整,有效地解决了具有不同时长的动态表情序列的样本描述. 为进一步改善表情识别模型的泛化性能,本文引入了基于多个个体识别模型加权投票的集成识别模型. 针对投票过程中的权值学习,提出了基于随机重采样的投票权重学习以及基于个体分类模型相对优势的投票权重学习方法. 集成决策进一步改善了表情识别性能. 基于 AFEW5.0 的动态表情库实验验证了算法的有效性.

**关键词:** 动态表情识别; 多视觉描述子; 集成分类器; 权重学习

**中图分类号:**           **文献标识码:** A           **文章编号:** 0372-2112 (2019)08-1643-11

**电子学报 URL:** <http://www.ejournal.org.cn>   **DOI:** 10.3969/j.issn.0372-2112.2019.08.006

## Dynamic Facial Expression Recognition Based on Multi-Visual and Audio Descriptors

LI Hong-fei<sup>1,2</sup>, LI Qing<sup>1,2</sup>, ZHOU Li<sup>1,2</sup>

(1. Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Communication in any form either verbal or non-verbal is vital to complete various daily routine tasks and plays a significant role in life. Facial expression is the most effective form of non-verbal communication and it provides a clue about emotional state, mindset and intention. Till now, facial expression recognition has been successfully applied to various fields such as safe driving, merchandise sales, clinical medicine, and so on. This thesis explores key techniques related to facial expression recognition. The main work and contributions are as follows. A dynamic facial expression recognition algorithm based on multi-visual descriptors and audio features is proposed under unrestricted conditions, in which dynamic facial feature extraction was conducted based on local spatial-temporal feature representation via multi-visual descriptors. Furthermore, the combination of video and audio features improves the recognition performance. Dynamic time warping based on timeline segmentation and covariance matrix proves to be effective in analyzing dynamic expression sequences of different time duration. To improve the generalization performance of facial expression recognition model, an integrated decision-making strategy based on weight voting by multiple individual recognition models is introduced. In order to effectively learning the weight for each individual recognition model, the method of voting weight learning by random re-sampling and the method of voting learning based on comparative advantages of individual recognition model are proposed. Finally the above ensemble model is applied and the recognition performance is further improved. Experiments on AFEW5.0 dataset validate the performance of the proposed dynamic facial expression algorithm.

**Key words:** active expression recognition; multi-visual descriptors; ensemble model; weight learning

## 1 引言

随着计算机技术和人工智能技术的发展,人脸表情识别更是作为人机交互的一个重要接口,吸引了大量的学者对其展开研究.许多著名的国外院校例如匹兹堡大学、加利福尼亚大学等研究机构已经开始致力于人脸表情识别的研究,并且取得了一定的进展.曼彻斯特大学 Cristinacce 等人将 AAM 模型与他们提出的特征响应成对增强算法相结合<sup>[1]</sup>,实现关于眼眶周围、鼻尖和两边嘴角等中心面部组件的特征点以及面部边缘区域的特征提取.20世纪90年代末,国内一些比较著名的院校陆续地注意到了这个领域并对其展开了研究.香港科技大学的 Liao 等人<sup>[2]</sup>,则是将基于 Gabor 滤波器的 Tsallis 能量以及梯度图和强度的局部二元模式特征的结合,实现关于人脸表情特征的描述,并基于线性判别分析法(LDA)训练脸部表情图像提取的特征.Zhu<sup>[3]</sup>等人开创了基于28个脸部特征点检测的表情特征提取方法.然而随着对表情识别研究的深入,人们发现表情有真有假,有些表情持续时间仅止于瞬间,快到让人难以察觉,它表达了人们试图压抑与隐藏的真实情感.国外研究人员将其定义为微表情.微表情在司法机关诊断案情,商业谈判等领域都有突出的成就.Ekman 团队是开展微表情研究的主要力量,他们进行了面部测绘行为的研究<sup>[4]</sup>,验证了微表情可提供更全面的隐蔽情感的泄露.在国内,一些研究者也将目光投放到了对微表情的研究上,认知心理学博士傅小兰教授主持了2011~2013年的国家自然科学基金项目“面向自动谎言识别的微表情表达研究”,成为国内微表情研究第一人.2010年,吴奇等<sup>[5]</sup>从心理学角度总结了微表情识别能力测验和早期训练工具.2011年吴冉等<sup>[6]</sup>从心理学角度研究微表情的启动效应.2013年,梁静等人<sup>[7]</sup>简略地总结了三种微表情数据库以及3种微表情自动识别方法.随着表情识别研究的不断深入,表情识别将会在我们的生活中发挥越来越大的作用.

本文结合 AFEW5.0 (Acted Facial Emotion in Wild),研究非约束条件下,动态序列的人脸表情识别.如何有效地捕捉和表征人脸不同类型表情的动态变化是动态表情识别研究的关键.首先获取人脸表情样本集 AFEW5.0,并对其进行预处理,对于动态序列的表情识别,表情样本的预处理包含两部分:(1)视频片段或动态表情序列中人脸检测与定位;(2)人脸区域的辐射校正、几何校正与裁剪.如图1所示为一个来自 AFEW5.0 库的7种动态表情类型的14个样本;之后,结合多种图像、视频特征检测算法,实现基于多视觉描述子的面部动态表情特征提取;同时,基于协方差矩阵以及时间分段方式实现动态表情视频特征的规范化;

由于会话场景中面部动态表情的情感信息表达也伴随着相同情感的语音语调信息的表达,因此研究中尝试基于视频、音频多模态特征实现动态表情特征描述.本文还尝试借助集成学习策略,进一步改善动态人脸表情的识别性能.最后,基于 AFEW5.0 验证算法的有效性.



图1 来自 AFEW5.0 库的7种动态表情类型的14个样本. A, Angry; B, Disgust; C, Sad; D, Fear; E, Neutral; F, Happy; G, Surprise

## 2 AFEW5.0 动态表情库及预处理

结合 AFEW5.0 表情数据集进行动态表情识别的动机源自于 EmotiW 挑战赛. AFEW 表情库提供的人脸表情视频片段是从众多电影片段剪裁得到,是最接近真实条件的、非控制前提下、具有自发意义的动态表情库,具有多样的背景、多变的光照条件、多姿势的人脸、脸部不同程度遮挡的特点,涉及不同种族、不同性别、不同人种、不同年龄阶段的人物表情.表情库中涉及7类基本表情:愤怒(Angry)、厌恶(Disgust)、恐惧(Fear)、快乐(Happy)、中性(Neutral)、悲伤(Sad)、及惊讶(Surprise).每个片段对应一个动态表情样本.

### 2.1 样本集的筛选、转换

并不是每个原始视频片段中一定包含人脸图像,这也是该数据集异于其他数据集的一点.基于 AFEW5.0 的上述特点,我们先对数据集进行了人工处理.首先,剔除了不含人脸区域的视频样本;其次,对于包含人脸

区域的视频样本,基于人脸检测及几何校准,得到了相应的基于静态图像序列描述的动态表情样本;此外,在 AFEW5.0 表情库的训练集、验证集中,每个动态表情样本均附有类别标签;对于测试集,并未提供相应的类别标签.但可以将测试集上的测试结果交由竞赛的主办方,主办方会返回关于测试集的总体识别率以及反映不同类别表情决策结果的混淆矩阵.

## 2.2 训练集的扩张

由于 AFEW5.0 动态表情样本数量有限,并且各类样本数目分布并不均衡,因此在实验中,需要进行训练样本集的扩张.主要采取了两种扩展策略:(1)将验证集部分样本补充到训练集内,避免训练集内各类样本的不均衡现象;(2)提取验证集各类样本的镜像数据.按照上面的扩张方法将镜像的验证集的一些样本补充到训练集中进行训练集的扩张.原始的验证集还继续作为验证集.

## 3 基于多视觉描述子的动态表情局部空时特征提取

基于局部空时特征(local spatial-temporal features),捕捉不同类型动态表情的面部不同程度局部形状、肌肉局部运动的空时变化特性,尝试结合几种典型的局部空间、空时特征检测算子进行动态表情特征描述.我们分别考察了基于 LBP-TOP、dense SIFT、dense HOG、dense HOF、dense MBHx、dense MBHy、Gabor 变换、分块二维 DCT 变换等多种不同形式的特征提取算法.

基于 LBP-TOP、dense SIFT、dense HOG、dense HOF、dense MBHx、dense MBHy 特征提取方式,是借助对表情样本图像进行规则分块、稠密采样而得到的以块直方图连接的形式来表达的局部空时域特征.基于 Gabor 变换、分块二维 DCT 变换方式可以实现关于表情样本的局部频域特征.本文仅对其中对表情视频片段识别作用性最大的关键视觉算子进行原理的介绍.

### 3.1 基于 LBP-TOP 算子的动态表情特征提取

LBP-TOP(Local Binary Patterns from Three Orthogonal Planes)<sup>[9]</sup>是一种基于局部空时域二值编码的特征提取算子.对于每一个表情矩形体,在如图 2 所示 LBP-TOP 三个正交平面采用 LBP 算子<sup>[10]</sup>进行基于局部二值编码的纹理特征,从而表述表情图像序列的局部空时信息.

LBP-TOP 算子本质上是对 VLBP<sup>[9]</sup>(Volume Local Binary Patterns)算子的一种改进.VLBP 算子也是一种基于空时域三维局部邻域提取表情纹理信息的算子.但是 VLBP 算子具有许多不足之处,为克服 VLBP 描述子的不足,引出 LBP-TOP 算子.基于 LBP-TOP 算子的局部时空纹理编码,不仅大大减少内存空间占用率、特征

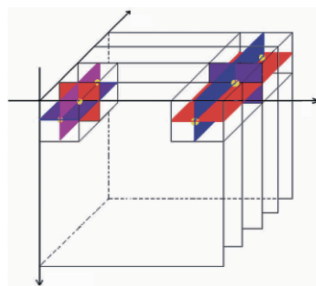


图2 LBP-TOP三个正交平面

编码速度快,而且对表情图像序列空间信息和时间信息表述更加完备、精确.

基于 LBP-TOP 算子的表情样本特征提取过程如下:首先对每个表情矩形体如图 3(a)所示,对 XT、YT、XY 三个正交平面进行 LBP 编码.其中,XY-LBP 编码主要描述了动态表情样本沿时间轴 T,在每个时刻人脸表情图像的空间信息;而 XT-LBP、YT-LBP 则主要描述了动态表情样本沿 Y 轴、X 轴的每个截面的 XT 或 YT 的空时信息.通过编码得到对应 XT、YT、XY 三个正交平面的三个 LBP 图像矩形体,按照一定的划分规则,对各图像矩形体进行 LBP 直方图描述如图 3(b)所示每个正交平面的 LBP 直方图;最后,将得到的多个 LBP 图像直方图顺次串接,即可得到一个如图 3(c)所示基于 LBP-TOP 编码的高维直方图特征向量.这就是基于 LBP-TOP 的动态表情样本特征表示.为编码直方图描述导致的特征向量维数过高,采用 uniform LBP 编码模式.对于 8 点邻域 LBP 编码,最多可得到 58 个不同编码模式.

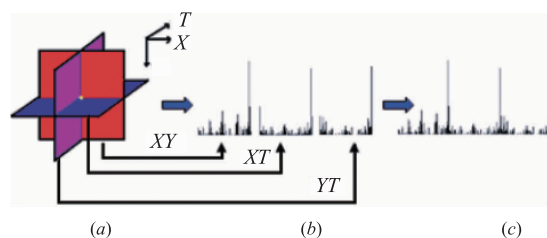


图3 (a)基于LBP-TOP纹理特征提取的三个正交平面;(b)每个正交平面的LBP直方图;(c)连接三平面LBP直方图,得到基于LBP-TOP编码的高维直方向量

因空间采样和时间采样的意义各不相同,因此,对于基于 LBP-TOP 算子的动态表情特征提取,时间轴和空间轴上设置编码邻域半径长度会有所区别.特别是高空间分辨率的短时表情图像序列.因此,可在 X、Y 轴和 T 时间轴上设置不同的邻域半径,使 LBP 算子中的圆形邻域采样转换成椭圆邻域采样.

设  $R_x$ 、 $R_y$ 、 $R_t$  分别表示 X 轴、Y 轴、T 轴方向编码邻域半径, $P_{xy}$ 、 $P_{xt}$ 、 $P_{yt}$  表示分别表示如图 4 所示 XY、XT、YT 三个正交平面中椭圆编码邻域编码像素数量.如图 5 所示,展示了表情矩形体某编码邻域坐标为  $(x_c$ ,

$y_c, t_c$ ) 的中心像素  $g_{t_c, c}$  在三个正交平面上编码邻域.

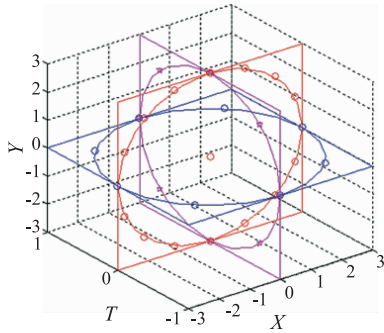


图4 三个正交平面中不同的半径和不同的邻域点数量

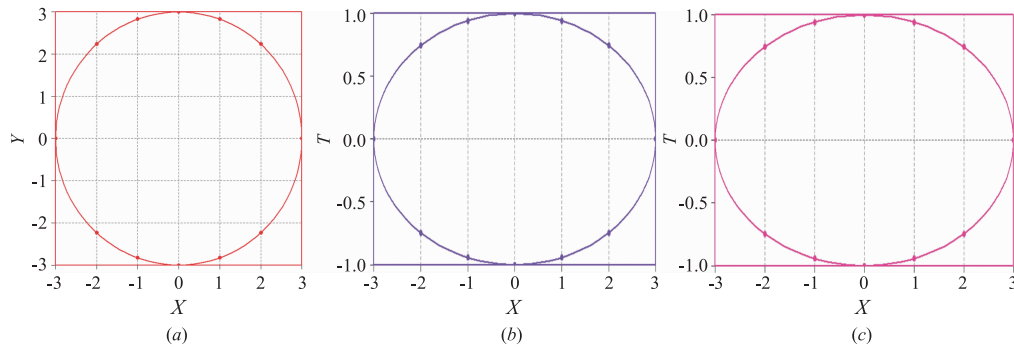


图5 表情形体中坐标为  $(x_c, y_c, t_c)$  的像素  $g_{t_c, c}$  在三个正交平面的编码邻域: (a)XY平面; (b)XT平面; (c)YT平面

其中:  $H_{i,j}$  表示 XY、XT、YT 三个正交平面中第  $j$  个正交平面直方图的第  $i$  个 bin 的取值;  $n_j$  表示第  $j$  个平面生成的二值编码模式的数量;  $j = 0, 1, 2$  分别表示 XY 平面、XT 平面、YT 平面;  $f_j(x, y, t)$  表示中心像素  $(x, y, t)$  在第  $j$  个平面的 LBP 编码;  $I(\bullet)$  为指示函数, 并且有:

$$I(A) = \begin{cases} 1, & \text{若 } A \text{ 为真} \\ 0, & \text{其它} \end{cases} \quad (2)$$

由于 XY、XT、YT 三个正交平面中, 每个平面空时特征的直方图长度不一, 因此需要进行动态表情纹理特征直方图的归一化处理, 即:

$$N_{i,j} = \frac{H_{i,j}}{\sum_{k=0}^{n_j-1} H_{k,j}}, i=0, 1, \dots, n_j-1, j=0, 1, 2 \quad (3)$$

按照一定的方式将上述 LBP 图像形体进行划分, 分别对每个划分子块进行直方图描述, 最终将各子块直方图按照一定方式串接, 即可得到关于动态表情图像序列的空时特征描述.

## 3.2 基于 HOG 与 HOF 的动态表情特征提取

### 3.2.1 基于 HOG 的图像特征提取

方向梯度直方图 (Histogram of Oriented Gradient, HOG)<sup>[11]</sup> 是基于空间梯度统计信息对图像区域局部表现、形状的一种描述, 是关于目标区域的一种视觉纹理特征描述子. HOG 特征的一个成功应用就是行人检测. 基于 HOG 的样本图像特征提取过程如下:

设  $\{(x_c, y_c, t_c) | x_c \in \{0, 1, \dots, X-1\}, y_c \in \{0, 1, \dots, Y-1\}, t_c \in \{0, 1, \dots, T-1\}\}$  为  $X \times Y \times T$  的 3D 表情形体, 它表达了一个完整的动态表情过程. 其中, 那些处于表情立方体的边缘区域相对于形体靠近中心的区域, 承载的信息较少, 因此实际处理时, 仅考虑对靠近形体中心区域的像素邻域进行 LBP-TOP 编码, 所以, 可对 XY、XT、YT 三个正交平面的动态表情纹理特征直方图进行如下定义:

$$H_{i,j} = \sum_{x,y,t} I(f_j(x, y, t) = i), i=0, 1, \dots, n_j-1, j=0, 1, 2 \quad (1)$$

Step1. 获取灰度图像  $I(x, y)$ , 借助 Gamma 校正进行图像局部区域对比度调整.

Step2. 获取图像中各个像素的梯度信息. 梯度幅值及梯度相位计算方式如下:

$$G(x, y) = \sqrt{[I(x+1, y) - I(x-1, y)]^2 + [I(x, y+1) - I(x, y-1)]^2} \quad (4)$$

$$\theta(x, y) = \arctan \frac{I(x, y+1) - I(x, y-1)}{I(x+1, y) - I(x-1, y)} \quad (5)$$

Step3. 将样本图像划分成多个细胞 (cells).

Step4. 生成关于每个细胞的梯度方向直方图.

Step5. 按照一定规则生成一系列图像块 (block); 对其进行直方图描述.

Step6. 将图像中各块直方图顺次串接, 得到 HOG 算子对样本图像的特征描述.

### 3.2.2 基于 HOF 的动态图像序列特征提取

与 HOG 统计样本图像局部区域的梯度信息不同, HOF 算子 (Histogram of Optical Flow, 光流方向直方图)<sup>[12]</sup> 算子捕获的是动态图像序列 (视频对象) 的局部运动信息. HOF 算子采用经典的 Horn-Schunck<sup>[13]</sup> 方法提取样本图像序列中每个像素处的光流信息, 因而得到每个时刻的光流场; 用光流场替换 HOG 算子计算所使用的梯度场, 即可实现基于 HOF 的动态图像序列直方图描述.

### 3.3 基于 MBH 算子的动态表情特征提取

运动边界直方图 (Motion Boundary Histograms, MBH) 最初是由 Dalal et al.<sup>[12]</sup> 提出的一种用于人体检测的算子, 后来被 Wang et al.<sup>[14]</sup> 用于基于运动边界动作识别. 不同于 HOF 算子捕捉两个不同帧之间的绝对局部运动, MBH 算子通过描述光流梯度的变化来表征两个像素间的相对运动, 从而抑制局部不变的运动信息, 保留光流梯度变化的信息 (如运动边界).

MBH 算子将光流域分解为水平方向和垂直方向的光流分量; 分别将其视为独立的图像, 按照 HOG 特征提取方式, 对它们进行直方图描述, 分别得到了基于 MBHx、MBHy 的动态序列特征提取. 此 MBH 是 MBHx 和 MBHy 的统称.

与 HOG 算子相似, MBH 算子也是基于空时图像块提取直方图特征. 本文采用文献<sup>[15]</sup> 快速算法提取 dense HOG、dense HOF、dense MBH 等直方图特征.

本次研究提出了一种如图 6 所示多模态表情识别框架. 其“多模态”不仅体现在引入了多种视觉算子提取人脸表情特征, 同时, 还体现在基于音频特征的描述.

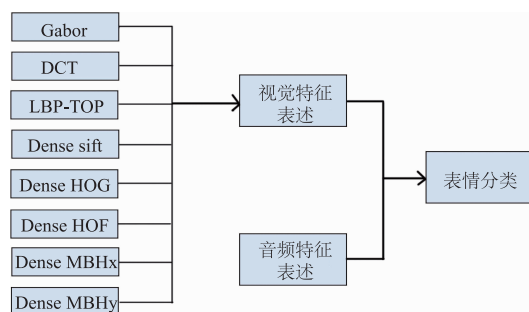


图6 多模态表情识别框架

## 4 音频特征的获取

音频特征的提取是借助开源的情感及情绪识别库 openEAR<sup>[16]</sup> 以及后台工具包 openSMILE<sup>[17]</sup> 从伴随视频的音频文件中提取, 共得到了 1583 维音频特征. 该特征集包含: 714 个与能量和频谱相关的低级描述子以及相应的 714 个 delta 系数、76 个与发音有关的低级描述子及其相应的 76 个 delta 系数、以及 3 个关于发音或不发音的持续性特征表述. 如表 1 所示音频特征低级描述子.

## 5 表情视频片的特征表述及分类

### 5.1 多视觉描述子的时间规整

提取静态表情图像序列的特征表述过程如下: 将静态表情序列 (表情视频片段) 看成一个图像立方体, 并且对图像立方体进行无重叠地分块, 生成若干个小

的图像立方体, 提取每个小图像立方体的特征描述, 连接所有小图像立方体的特征描述, 生成关于静态表情序列的特征描述. 其次, 将已经提取的静态表情序列的特征表述在时间轴上平均分为三段, 计算每个时间段表情图像序列的平均特征表述, 将三个时间段的平均特征向量依次串联起来. 最后, 生成上述各个视觉算子关于当前表情图像序列的特征表述.

表 1 音频特征低级描述子

Low Level Descriptors (LLD)	
Energy/Spectral LLD 714 × 2 = 1428	PCM Loudness
	MFCC[0 - 14]
	log Mel Frequency Band[0 - 7]
	Line Spectral Pairs (LSP) frequency[0 - 7]
	F0
	F0 Envelope
Voicing related LLD 76 × 2 + 3 = 155	Voicing Prob
	Jitter Local
	Jitter consecutive frame pairs
	Shimmer Local

### 5.2 特征融合策略

本研究中, 基于表情视频片段提取的特征可以分为两类: 视觉特征和音频特征. 对于音频特征既可以单独用于表情的分类, 同时也可以与视觉特征相结合用于表情的分类. 同样对于视觉特征既可以单独用于人脸表情的分类, 也可以相互结合生成一个高维的特征表述用于最后的表情视频片段分类. 经过特征融合之后, 特征的维数过高可能会带来数据冗余、随机噪声, 影响表情识别效果. 考虑到降维可能会提高人脸表情的识别率. PCA<sup>[18]</sup> 不但可以实现特征的降维, 同时还能消除数据中的噪声. 因此, 尝试采用 PCA 对原始的特征描述进行降维, 可能会提高人脸表情的识别率.

### 5.3 表情视频片段分类

在完成了关于动态表情样本的特征提取之后, 即可结合具体的分类器进行人脸表情识别模型的学习. 我们主要采用偏最小二乘 (Partial Least Squares, PLS) 以及 C-SVM 等分类器进行动态表情的分类. 基于 C-SVM, 主要结合 linear SVM、polynomial SVM、RBF SVM、histogram intersection SVM (IKSVM), 进行表情识别模型的性能分析.

## 6 多识别模型的集成

基于不同特征提取方式与分类模型相结合, 可以学习得到多个用于表情识别的模型. 本节我们考虑多个识别模型的结合. 按照文献[20 - 21] 关于集成学习 (Ensemble Learning) 的理解与认识, 多识别模型集成的优势表现在如下几个方面: 从统计的角度来看, 将多个

识别模型相结合,可以减弱基于单个模型用于表情识别可能因误选而导致的识别模型泛化能力不佳的风险;从计算的角度看,多个识别模型的结合,会减弱基于单个识别模型的学习陷入局部极值的风险;从表示方面来说,一些学习任务的真实假设可能不在当前学习模型所考虑的假设空间,使用多个识别模型相结合的方式,会使假设空间扩大,得到更好的识别结果.

### 6.1 基于加权投票机制的多识别模型集成决策

由文献[22]可知,集成模型的构建通常有两种途径:方式一,基于独立训练(Independent Training)个体模型的集成;方式二,基于协作式训练(Coordinated Training)的个体模型的集成.

方式一中,每个个体模型的训练都无需借助于参与集成的其它个体模型而独立完成的,因此个体模型之间不存在强依赖关系,是一种并行化集成,这种集成方式中,各个个体模型彼此相互独立.其中一个重要的途径就是基于加权投票机制的模型集成.

方式二中,参与集成的每个个体模型的学习都是在前一个个体模型学习的基础上完成的,而当前个体模型训练的目的也是为了克服前一个个体模型的不足.因此,各个个体模型的学习是串行进行的,个体模型之间存在明显依赖关系.

本文中由于各个个体表情识别模型的训练是彼此独立的,因此我们将考虑基于多识别模型决策结果加权投票的集成方式.我们首先对这种加权投票的集成方式进行描述.

设表情类别数目为  $M$ ,  $H = \{h_1, h_2, \dots, h_L\}$  为多个基于不同特征描述、采用 SVM 或 PLS 分类器得到的  $L$  个个体识别模型的集合,其中第  $l$  个个体识别模型为  $h_l$ ;分别基于相同验证集对上述各识别模型进行评价,得到各识别模型的总体识别率  $\mathbf{Acc} = \{Acc_1, Acc_2, \dots, Acc_L\}$ ,其中  $Acc_l$  为第  $l$  个识别模型的识别率;将上述  $L$  个识别模型对集成模型决策结果的贡献权重,组成权重向量  $\boldsymbol{\omega} = [\omega_1 \ \omega_2 \ \dots \ \omega_L]^T$ ,其中第  $l$  个识别模型的贡献权重为  $\omega_l \in [0, 1]$ ,  $l = 1, 2, \dots, L$ , 并且有:

$$\sum_{l=1}^L \omega_l = 1.$$

给定观测样本  $x$ ,基于上述多个识别模型的决策结果可得到各类别的加权投票值:

$$g_j(x) = \sum_{l=1}^L \omega_l I(h_l(x), j), j = 1, 2, \dots, M \quad (6)$$

上式中  $I(\bullet, \bullet)$  为指示函数,即:

$$I(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \quad (7)$$

最终关于样本  $x$  的类别标号  $y$  的决策结果为:

$$\hat{y}(x) = \underset{j \in \{1, 2, \dots, M\}}{\operatorname{argmax}} g_j(x) \quad (8)$$

上述基于投票机制的集成模型的性能,依赖于各个个体识别模型的投票权重.因此,如何结合各个个体识别模型的识别性能,进行权重的学习是这种加权投票机制的关键.

本文基于个体识别模型总体识别率  $\mathbf{Acc} = \{Acc_1, Acc_2, \dots, Acc_L\}$ ,给出了两种用于投票权重  $\boldsymbol{\omega} = [\omega_1 \ \omega_2 \ \dots \ \omega_L]^T$  的学习方式,分别是:基于随机重采样的投票权重学习,以及基于个体分类模型相对优势的投票权重学习.下面分别予以详述.

#### 6.1.1 基于随机重采样的投票权重学习

给定足够大的采样组数  $N$  (例:  $N \geq 5000$ ),从集合  $H = \{h_1, h_2, \dots, h_L\}$  中,对  $L$  个个体识别模型进行  $N$  组随机重采样;每一组重采样都需要按照一定的选择概率、有放回地抽取  $L$  个可能重复出现的个体识别模型.

每个个体识别模型的选择概率,取决于其总体识别率,具体估计方式为:

$$P_l = \frac{(Acc_l)^q}{\sum_{k=1}^L (Acc_k)^q}, l = 1, \dots, L \quad (9)$$

其中:  $P_l$  是第  $l$  个识别模型的选择概率;  $q \geq 1$  是控制参数,当各个个体识别模型的总体识别率性能差异微弱时,通过增大  $q$  值,增强识别性能好的个体被选择的机会,默认值为 1.

采用“轮盘赌法”生成每一组重采样的  $L$  个个体识别模型;其中,候选个体识别模型  $H = \{h_1, h_2, \dots, h_L\}$  中  $h_l$  模型的概率选择区间记为  $[sumP_{l-1}, sumP_l)$ ,  $l = 1, \dots, L$ . 具体计算方式为:

$$sumP_l = \begin{cases} 0, & l = 0 \\ \sum_{k=1}^l P_k, & l = 1, 2, \dots, L \end{cases} \quad (10)$$

以第  $n$  组重采样得到的  $L$  个个体识别模型  $\hat{H}_n = \{h_{(n)}^{(1)}, h_{(n)}^{(2)}, \dots, h_{(n)}^{(L)}\}$  为例.其中  $h_{(n)}^{(j)} \in H$  为第  $n$  组重采样过程中第  $j$  次有放回抽取得到的个体识别模型,设该个体识别模型对应的总体识别率为  $Acc_{(n)}^{(j)} \in Acc$ . 随机产生  $L$  个  $0 \sim 1$  之间均匀分布的随机数  $num_j, j = 1, \dots, L$ ;若  $num_j \in [sumP_{l-1}, sumP_l)$ ,则以候选个体识别模型  $h_l$  作为  $h_{(n)}^{(j)}$ ,相应的总体识别率为  $Acc_{(n)}^{(j)} = Acc_l$ .

综上,基于上述重采样,得到  $N$  组重采样结果;统计每个个体识别模型作为识别率最高的模型在  $N$  组重采样中出现的次数;记  $N_l$  为第  $l$  个识别模型出现次数,  $l = 1, 2, \dots, L$ . 基于上述重采样方式,可得识别模型  $h_l$  对最终决策结果的贡献权重为:

$$\omega_l = \frac{N_l}{N} = \frac{N_l}{\sum_{k=1}^L N_k} \quad (11)$$

#### 6.1.2 基于个体分类模型相对优势的投票权重学习

受文献[9]识别过程中关于特征重要性的分析,我

们给出一种基于个体分类模型相对优势的投票权重学习方式. 基本思路是:(1)对于每一种个体识别模型,以总体识别率表达每个个体进行表情识别的绝对优势,识别率越高,绝对优势越强;(2)对于一组个体识别模型,以这组模型中绝对优势最弱的个体识别模型为基准,评判各个个体识别模型的相对优势,并将各个个体识别模型的相对优势进行得分等级的量化成 1~10,其中满分 10 分对应绝对优势最大的个体识别模型,而最低分 1 分对应绝对优势最小的个体识别模型,其它个体识别模型得分按照比例量化,最终基于各个个体的相对优势等级评分;(3)将其归一化,得到各识别模型  $h_l, l=1, 2, \dots, L$ , 对最终决策结果的贡献权重. 下面我们给出详细的实现步骤.

Step1. 基于相同验证集, 获取  $H = \{h_1, h_2, \dots, h_L\}$  中每一个体识别模型  $h_l$  的总体识别率  $Acc_l$ , 以 0~1 之间的小数表示, 从而有  $Acc = \{Acc_1, Acc_2, \dots, Acc_L\}$ .

Step2. 计算所有个体识别模型的相对优势, 使其量化等级得分为  $[1, 10]$ :

$$r_l = \frac{Acc_l - \min_{k \in \{1, 2, \dots, L\}} \{Acc_k\}}{\max_{k \in \{1, 2, \dots, L\}} \{Acc_k\} - \min_{k \in \{1, 2, \dots, L\}} \{Acc_k\}} \times 9 + 1, l = 1, \dots, L \quad (12)$$

Step3. 基于相对优势, 估计各个个体识别模型的归一化权重, 得:

$$\omega_l = \frac{r_l}{\sum_{k=1}^L r_k}, l = 1, \dots, L \quad (13)$$

基于 6.1.1 或 6.1.2 所描述的两种方式完成了投票权重的估计之后, 结合 PLS 或者 C-SVM 即可进行未知样本的决策.

## 7 实验与分析

### 7.1 参数设置

考虑到光照强度的不一致会对人脸表情的识别造成一定的影响, 因此, 将竞赛提供的已经经过人脸探测和校准的静态表情图像序列采用 DoG (Difference of Gauss, 高斯差分) 滤波器进行光照归一化, 并将样本集中的  $128 \times 128$  的人脸表情图像采用双三次插值法将图像重新设置为  $64 \times 64$  大小. 由于最初的 AFEW5.0 样本集中有个别表情图像序列并未检测到人脸, 因此, 去除样本集中并未检测到人脸的表情图像序列, 得到含有 700 个表情图像序列的训练集, 366 个表情图像序列的验证集, 539 个表情图像序列的测试集. 经过预处理阶段之后, 尽管大多数表情视频片段含有足够多的人脸表情样本图像, 但是有的视频片段中检测到人脸的表情帧数量过少. 因此, 对于检测到人脸的表情帧数过少的表情视频片段, 采用统一地复制表情帧两次的方

法来扩充表情视频片段中可用的表情帧的数量.

### 7.2 基于多视觉描述子的动态表情特征描述

我们在表情样本集 AFEW5.0 上提取了多种视觉特征. 由于上述视觉算子提取表情视频特征的处理方式类似. 下面仅展示了其中两种视觉算子提取特征的具体过程.

#### 7.2.1 基于 MBH、HOF 特征的动态表情描述

HOF 特征的提取过程的关键是基于相邻时刻两幅人脸表情图像进行光流场的计算. 给定一个动态表情样本的“表情矩形体”, 借助该矩形体内相邻两幅表情图像光流场的计算, 可得到一个关于人脸表情变化的“光流矩形体”.

以光流替换梯度, 以光流方向替换梯度方向, 以光流幅值替换梯度幅值, 采用类似 HOG 特征计算方式, 若采用与 HOG 特征描述中相同的参数设置, 即:“光流矩形体”中每一幅  $64 \text{ 像素} \times 64 \text{ 像素}$  的光流图像, 每个细胞 (cell) 的大小为  $8 \text{ 像素} \times 8 \text{ 像素}$ , 每个细胞的光流方向直方图量化为 9 个方向; 各细胞以  $2 \times 2$  布局构成一个图像块 (block), 因此每个图像块含 4 个细胞; 以重叠方式提取每个图像块, 水平或垂直方向, 相邻图像块的重叠率 50%.

每一幅  $64 \text{ 像素} \times 64 \text{ 像素}$  的光流场图像, 会生成  $7 \times 7 = 49$  个图像块; 对每个图像块提取  $4 \times 9 = 36$  维的光流方向直方图, 并进行规范化; 将 49 个图像块的直方图特征表述, 串联生成当前表情样本图像  $49 \times 36 = 1764$  维的光流方向直方图表述.

为得到一个动态表情序列的 HOF 直方图描述, 首先在时间轴上, 将该“光流矩形体”等分成三段; 计算每个时间段表情图像光流方向直方图的均值向量; 三段均值向量顺次串接, 得到关于该动态表情序列的一个  $1764 \times 3 = 5292$  维的 HOF 直方图向量描述. MBH 特征与 HOF 特征计算方式类似, 对于每个表情图像序列最终得到  $1764 \times 3 = 5292$  维的 MBH 直方图特征表述.

#### 7.2.2 直方图向量的规范化

在提取 HOF、MBHx 和 MBHy 特征时, 为了增强特征分量之间的差异, 分别运用了 L1-范数和 L2-范数对特征向量进行规范化. L1-范数规范化即是向量中每个元素除以向量的 L1-范数, L2-范数规范化即是向量中每个元素除以向量的 L2-范数. 向量  $x = \{x_1, x_2, \dots, x_n\}$  的 L1-范数规范化过程如下:

$$\text{norm}(x) = |x_1| + |x_2| + \dots + |x_n| \quad (14)$$

$$x_i^s = \frac{x_i}{\text{norm}(x)}, i = 1, 2, \dots, n \quad (15)$$

向量  $x = \{x_1, x_2, \dots, x_n\}$  的 L2-范数规范化过程如下:

$$\text{norm}(x) = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (16)$$

$$x_i^g = \frac{x_i}{\text{norm}(x)}, i = 1, 2, \dots, n \quad (17)$$

$x_i^s, x_i^g$  分别是 L1 - 范数和 L2 - 范数规范化后的结果.

实验结果表明,相比于 L1 - 范数归一化的 HOF、MBHx、MBHy 特征,采用 L2 - 范数归一化的 HOF、MBHx、MBHy 特征能提高人脸表情识别率.

此外,对于 HOF、MBHx、MBHy 特征,因其本质是捕捉光流及其梯度变化的信息,基于光照校正后的表情图像提取特征削弱了梯度变化的信息.因此,我们只对表情本图像提取 HOF、MBH 特征之前,不进行辐射校正.

### 7.3 实验结果比较

表 2 在验证集和测试集上,基于 IKSVM、linear SVM 以及 PLS 分类器的表情识别结果

序号	表情识别模型		识别率 (%)	
	特征	分类模型	验证集	测试集
1	Baseline (LBP-TOP) 注: EmotiW2015 竞赛组织者提供	IKSVM	36.08	39.33
2	音频特征	线性 SVM (c = 256)	31.14	--
3	视频特征 (SIFT)	IKSVM (c = 0.5)	36.34	--
4	视频特征 (L2-Normalized MBHy)	IKSVM (c = 0.5)	36.34	--
5	视频特征 (HOG)	PLS (ncomp = 4)	39.34	--
6	视频 (LBPTOP + HOG)	PLS (ncomp = 4)	42.90	40.81
7	视频特征 (HOG)	IKSVM (c = 0.5)	41.53	--
8	音频 + 视频 (分块 2D DCT)	线性 SVM (c = 256)	40.43	
9	音频 + 视频 (Gabor 特征)	线性 SVM (c = 256)	43.44	
10	视频特征 (LBPTOP + SIFT + HOG + L2-normalized MBHy)	IKSVM (c = 0.5)	49.19	--
11	音频 + 视频 (LBPTOP + HOG + MBHy)	线性 SVM (c = 256)	48.63	--
12	音频 + 视频 (HOG + L2-normalized HOF)	线性 SVM	48.63	--
13	音频 + 视频 (HOG + SIFT)	线性 SVM	48.36	--
14	音频 + 视频 (HOG + HOF)	线性 SVM	47.81	--
15	模型 1 -- 音频 + 视频 (LBPTOP + HOG)	线性 SVM	48.08	46.94
16	模型 2 -- 音频 + 视频 (LBPTOP + HOG + HOF)	线性 SVM	50	46.38
17	模型 3 -- 音频 + 视频 (LBPTOP + L2-normalized MBHy)	线性 SVM	50	42.86
18	模型 4 -- 音频 + 视频 (LBPTOP + HOG + L2-normalized MBHy)	线性 SVM	49.45	46.94
19	音频 + 视频 (HOG)	线性 SVM	44.53	43.60
20	基于随机重采样进行投票权重学习的多识别模型集成 (q = 1) 参与集成的个体模型来自该表的 (6, 10, 15 ~ 19 行)	50.82 (186/366)	--	
21	基于随机重采样进行投票权重学习的多识别模型集成 (q = 2.5) 参与集成的个体模型来自该表的 (6, 10, 15 ~ 19 行)	51.09 (187/366)	--	
22	基于个体分类模型相对优势进行投票权重学习的多识别模型集成 (6, 10, 15 ~ 19 行)	51.09 (187/366)	--	

在完成了基于几种典型特征组合及分类器的识别模型学习之后,基于 AFEW5.0 验证集的评价结果,我们从中选取出识别性能较高的特征组合,考察基于 PCA 的特征降维,对识别性能的影响.

我们基于几种视频特征组合、以及视频特征与音频特征组合方式,结合不同类型的分类模型进行动态表情的分类.如表 2 所示在验证集和测试集上,基于 IKSVM、linear SVM 以及 PLS 分类器的表情识别结果,序号 3 ~ 19 对应的每一行即可视为一个基于训练样本集结合不同特征组合方式及分类模型得到的一个个体识别模型.对于分类模型,实验结果表明:基于最近邻分类器的识别最差,基于线性 SVM 和 PLS 分类器的表情视频片段识别效果最好.因此,最终采用 PLS 分类器和线性 SVM 用于 AFEW5.0 测试集的分类.

PCA (Principal Component Analysis) 的主要思想即是从原始的空间中顺序地找一组相互正交的坐标轴,新的坐标轴的选择与数据本身是密切相关的.其主要是通过计算样本数据的协方差矩阵,得到协方差矩阵

的特征值及其特征值所对应的特征向量,选择特征值最大(即方差最大)的  $k$  个特征所对应的特征向量组成矩阵.这样就可以将样本数据转换到新的特征空间中,实现数据特征的降维.求解 PCA 的过程如下:

输入:样本数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}, \mathbf{x}_i \in R^d, i = 1, 2, \dots, N, \mathbf{X} \in R^{d \times N}$

输出:样本数据集  $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}, \mathbf{y}_j \in R^k, j = 1, 2, \dots, N, \mathbf{Y} \in R^{k \times N}$

Step1. 去均值,即分别求每列的均值,之后所有列的样本数据都减去其对应的均值

Step2. 计算协方差矩阵  $\frac{1}{N}\mathbf{X}\mathbf{X}^T$

Step3. 用特征值分解方法求协方差矩阵  $\frac{1}{N}\mathbf{X}\mathbf{X}^T$  的特征值与特征向量

Step4. 对特征值从大到小排序,选择其中最大的  $k$  个特征值并将其所对应的  $k$  个特征向量分别作为行向量组成特征向量矩阵  $\mathbf{P}$

Step5. 即  $\mathbf{Y} = \mathbf{P}\mathbf{X}$

使用 PCA 方法降维,将音频特征降到了 45 维,其它特征降到了 637 维.对降维前后识别性能进行比较,实验结果表明,特征经过降维之后,表情视频片段的识别率降低.由于识别结果的提交次数受限,最多能提交 10 次.这里只给出了四次最好的提交结果对应的表情识别模型在验证集及测试集上评价得到的混淆矩阵.为便于查阅,我们分别将其记为模型 1 ~ 模型 4,分别对应表 2 的 15 至 18 行.表 2 关于个体识别模型的实验表明,相对于基线识别模型在测试集及验证集的识别率 39.33%、36.08%,本文对应模型 4 的识别模型在测试集、验证集上的识别率达到了 46.94%、49.45%.模型识别性能均有明显提高.如图 7 ~ 图 10 所示,分别展示了模型 1 和模型 4 相对应的混淆矩阵结果.

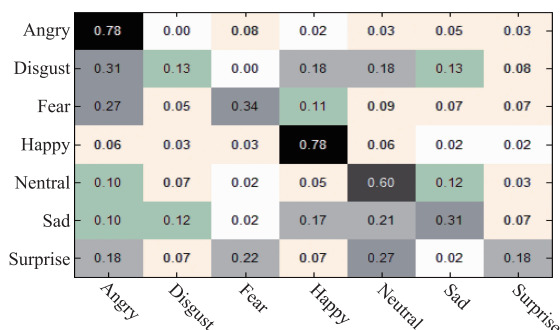


图7 模型1--基于“视频(LBPTOP+HOG)+音频”特征(9707维)及线性SVM(c=256)的验证集决策结果.其中:原始训练集700样本、验证集366样本

#### 7.4 基于加权投票的多个个体识别模型集成决策

由于各个体识别模型的学习都是独立于其它识别

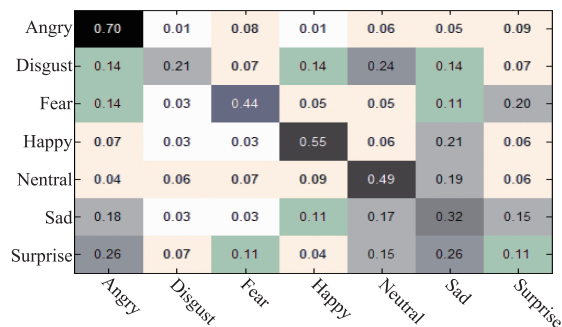


图8 模型1--基于“视频(LBPTOP+HOG)+音频”特征(9707维)及线性SVM(c=256)的测试集决策结果.其中:原始训练集700样本、测试集539样本

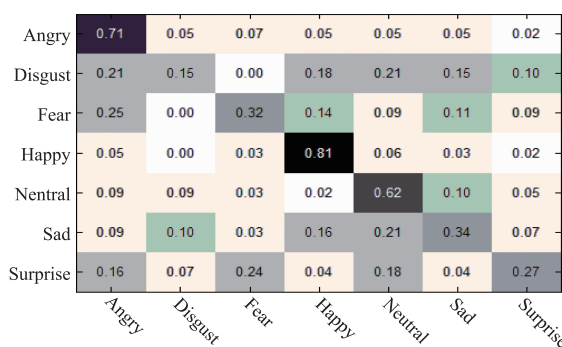


图9 模型4--基于“视频(LBPTOP+HOG+L2-normalized MBHY)+音频”特征(14999维)及线性SVM(c=256)的验证集决策结果.其中:原始训练集700样本、验证集366样本

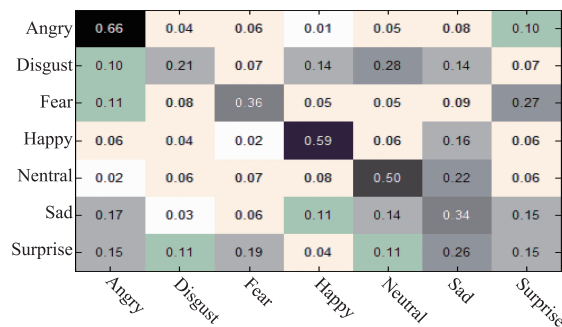


图10 模型4--基于“视频(LBPTOP+HOG+L2-normalized MBHY)+音频”特征(14999维)及线性SVM(c=256)的测试集决策结果.其中:原始训练集700样本、测试集539样本

模型完成的.进一步为改善动态表情的识别性能,我们结合表 2 中序号为 6、10、15 ~ 19 共计 7 个个体识别模型,结合本文提出的两种权重学习方法进行了基于加权投票方式的多个个体识别模型的集成决策.其在验证集上的识别性能可以提高至 51.09%.

## 8 结语

本文结合 AFEW5.0(Acted Facial Emotion in Wild)动态表情库,提出了一种基于多视觉描述子及音频特征融合策略的动态表情识别算法.有效地解决了具有

不同时长的动态表情序列的样本描述并且进一步改善了表情识别模型的泛化性能。但是本文尚有许多不足之处,未来还可以在如下方面进行改善:在进行表情视频预处理实验时,可以看到视频样本中存在大量重复的图像帧,因此在表情视频样本的预处理上可以考虑基于一定的约束条件选择视频样本中具有明显特征变化的数帧图像来进行特征提取;基于多视觉算子对样本图像的特征表述上,受到 CNN 模型的启发,取代直接串联的形式,可以考虑不同的视觉算子以多层堆叠的形式来表述视频样本特征;基于时间轴的规划上可以根据一个表情的起始、高潮、结束三个状态所持续时间的长短来进行更加合理的时间段规划;基于视觉、音频多模态特征结合以及多视觉算子集成过程中,可以考虑引入权重机制。

#### 参考文献

- [1] Liao C T, Duan C H, Lai S H. Accurate facial feature localization on expressional face images based on a graphical model approach [A]. *Advances in Multimedia Information Processing-PCM 2010-11th Pacific Rim Conference, Part II* [C]. Berlin: Springer, 2010. 672 - 681.
- [2] Liao S, Fan W, Chunga C S, et al. Facial expression recognition using advanced local binary patterns, Tsallis entropies and global appearance features [A]. *The 2006 IEEE International Conference on Image Processing* [C]. Atlanta: IEEE, 2006. 665 - 668.
- [3] Tong Y, Wang Y, Zhu ZW, et al. Facial feature tracking using a multi-state hierarchical shape model under varying face pose and facial expression [A]. *The 18th International Conference on Pattern Recognition* [C]. HongKong: IEEE, 2006. 283 - 286.
- [4] Ekman P, Rosenberg E L. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* [M]. New York: Oxford University Press, 1997.
- [5] 吴奇, 申寻兵, 傅小兰. 微表情研究及其应用 [J]. *心理科学进展*, 2010, 18(9): 1359 - 1368.  
Wu Qi, Shen Xuebing, Fu Xiaolan. Micro-expression and its applications [J]. *Advances in Psychological Science*, 2010, 18(9): 1359 - 1368. (in Chinese)
- [6] 吴冉, 任衍具. 微表情的启动效应研究 [J]. *应用心理学*, 2011, 17(3): 241 - 248.  
Wu Ran, Ren Yanju. Study on Priming effect of micro-expression [J]. *Chinese Journal of Applied Psychology*, 2011, 17(3): 241 - 248. (in Chinese)
- [7] 梁静, 颜文婧, 吴奇, 等. 微表情研究的进展与展望 [J]. *中国科学基金*, 2013, 27(2): 75 - 78, 82.  
Liang Jing, Yan Wenjing, Wu Qi, et al. Recent advances and future trends in micro-expression research [J]. *Bulletin of National Science Foundation of China*, 2013, 27(2): 75 - 78, 82. (in Chinese)
- [8] Winnemöller H, Kyprianidis J E, Olsen S C. XDoG: An extended difference-of-Gaussians compendium including advanced image stylization [J]. *Computers & Graphics*, 2012, 36(6): 740 - 753.
- [9] Guoying Z, Matti P. Dynamic texture recognition using local binary patterns with an application to facial expressions [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2007, 29(6): 915 - 928.
- [10] Jain S, Bagga S, Hablani R, et al. Facial Expression Recognition Using Local Binary Patterns with Different Distance Measures [M]. *Intelligent Computing, Networking, and Informatics*. Springer India, 2014. 853 - 862.
- [11] Watanabe T, Ito S, Yokoi K. Co-occurrence histograms of oriented gradients for pedestrian detection [A]. *Advances in Image and Video Technology, Third Pacific Rim Symposium. Proceedings* [C]. Berlin: Springer, 2009. 37 - 47. (4)
- [12] Dalal N, Triggs B, Schmid C. Human detection using oriented histograms of flow and appearance [A]. *Computer Vision-ECCV 2006, 9th European Conference on Computer Vision, Proceedings, Part II* [C]. Berlin: Springer, 2006. 428 - 441. (5)
- [13] Horn B K P, Schunck B G. Determining optical flow [J]. *Artificial Intelligence*, 1980, 17(81): 185 - 203.
- [14] Xu H, Tian Q, Wang Z, et al. A survey on aggregating methods for action recognition with dense trajectories [J]. *Multimedia Tools & Applications*, 2015, 55(03): 1 - 17.
- [15] Uijlings J, Duta I C, Sangineto E, et al. Video classification with densely extracted HOG/HOF/MBH features: an evaluation of the accuracy/computational efficiency trade-off [J]. *International Journal of Multimedia Information Retrieval*, 2014, 4(1): 33 - 44.
- [16] Eyben F, Wöllmer M, Schuller B. OpenEAR — Introducing the munich open-source emotion and affect recognition toolkit [A]. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* [C]. Amsterdam: IEEE, 2009. 1 - 6.
- [17] Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor [J]. *Acm Mm*, 2010: 1459 - 1462.
- [18] 胡振兴. 一种基于 PCA 类内平均脸法和支持向量机模型的人脸识别算法 [J]. *软件导刊*, 2012, 11(6): 33 - 34.  
Hu Zhenxing. Face recognition algorithm based on PCA class mean face method and support vector machine model

- [J]. Software Guide, 2012, 11(6): 33 - 34. (in Chinese)
- [19] Rosipal R., Krámer N. Overview and recent advances in partial least squares [A]. In Subspace, Latent Structure and Feature Selection [C]. Berlin: Springer, 2006. 34 - 51.
- [20] Dietterich, T. G. Ensemble methods in machine learning [A]. Proceedings of the 1st International Workshop in Multiple Classifier Systems [C]. Berlin: Springer, 2000. 1 - 15
- [21] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2015.  
Zhou Zhihua. Machine Learning [M]. Beijing: Tsinghua University Press, 2015. (in Chinese)
- [22] Baroque B, Corchado E. Fusion Methods for Unsupervised Learning Ensembles [J]. Studies in Computational Intelligence, 2011, 322.
- [23] 刘志华, 李改燕, 刘晓爽. 基于最小二乘法的蒙特卡洛移动节点定位算法 [J]. 传感技术学报, 2012, 25(4): 541 - 544.  
Liu Zhihua, Li Gaiyan, Liu Xiaochen. Monte Carlo mobile node localization algorithm based on least square method [J]. Journal of sensing technology, 2012, 25(4): 541 - 544. (in Chinese)
- [24] Maji S, Berg A C, Malik J. Classification using intersection kernel support vector machines is efficient [A]. 2008 IEEE Conference on Computer Vision and Pattern Recognition [C]. Anchorage: IEEE, 2008. 1 - 8.
- [25] Buenaposada J M, Muñoz E, Baumela L. Recognising facial expressions in video sequences [J]. Pattern Analysis & Applications, 2008, 11(1): 101 - 116.
- [26] Huang X, Zhao G, Pietikinen M, et al. Dynamic facial expression recognition using boosted component-based spatiotemporal features and multi-classifier fusion [A]. Advanced Concepts for Intelligent Vision Systems-12th International Conference, Proceedings, Part II [C]. Berlin: Springer, 2010. 312 - 322.
- [27] Dubuisson S, Davoine F, Masson M. A solution for facial expression representation and recognition [J]. Signal Processing Image Communication, 2002, 17(9): 657 - 673.
- [28] Liu C, Yuen J, Torralba A, et al. SIFT Flow: Dense correspondence across different scenes [A]. Computer Vision-ECCV 2008, 10th European Conference on Computer Vision, Proceedings, Part III [C]. Berlin: Springer, 2008. 28 - 42.
- [29] 贲晔, 杨明强, 张鹏, 李娟. 微表情自动识别综述 [J]. 计算机辅助设计与图形学 2014, 26(9): 1385 - 1395.  
Ben Xianye, Yang Mingqiang, Zhang Peng, Li Juan. Survey on automatic micro expression recognition methods [J]. Journal of Computer-Aided Design & Computer Graphics. 2014, 26(9): 1385 - 1395. (in Chinese)

#### 作者简介



李宏菲 女, 1990 年生于河北省三河市. 中国科学院微电子研究所博士研究生. 研究方向: 基于视觉的汽车主动安全与智能驾驶.



李庆 (通信作者) 男, 1972 年出生于吉林省四平市. 中国科学院微电子研究所博士生导师. 研究方向: 智能网联汽车.  
E-mail: liqing@ime.ac.cn