DOI: 10.11992/tis.202001032

多模态情绪识别研究综述

潘家辉[1], 何志鹏[1], 李自娜[2], 梁艳[1], 邱丽娜[1]
(1. 华南师范大学 软件学院, 广东 佛山 528225; 2. 华南师范大学 计算机学院, 广东 广州 510641)

摘 要: 本文针对多模态情绪识别这一新兴领域进行综述。首先从情绪描述模型及情绪诱发方式两个方面对情绪识别的研究基础进行了综述。接着针对多模态情绪识别中的信息融合这一重难点问题, 从数据级融合、特征级融合、决策级融合、模型级融合 4 种融合层次下的主流高效信息融合策略进行了介绍。然后从多种行为表现模态混合、多神经生理模态混合、神经生理与行为表现模态混合这 3 个角度分别列举具有代表性的多模态混合实例, 全面合理地论证了多模态相较于单模态更具情绪区分能力和情绪表征能力, 同时对多模态情绪识别方法转为工程技术应用提出了一些思考。最后立足于情绪识别研究现状的分析和把握, 对改善和提升情绪识别模型性能的方式和策略进行了深入的探讨与展望。

关键词: 情绪识别; 情绪描述模型; 情绪诱发方式; 信息融合; 融合策略; 情绪表征; 模态混合
中图分类号: TP391.4 **文献标志码**: A **文章编号**: 1673-4785(2020)04-0633-13

中文引用格式: 潘家辉, 何志鹏, 李自娜, 等. 多模态情绪识别研究综述 [J]. 智能系统学报, 2020, 15(4): 633–645.
英文引用格式: PAN Jiahui, HE Zhipeng, LI Zina, et al. A review of multimodal emotion recognition[J]. CAAI transactions on intelligent systems, 2020, 15(4): 633–645.

A review of multimodal emotion recognition

PAN Jiahui[1], HE Zhipeng[1], LI Zina[2], LIANG Yan[1], QIU Lina[1]
(1. School of Software, South China Normal University, Foshan 528225, China; 2. School of Computer, South China Normal University, Guangzhou 510641, China)

Abstract: This paper reviews the emerging field of multimodal emotion recognition. Firstly, the research foundation of emotion recognition is summarized from two aspects: emotion description model and emotion-inducing mode. Then, aiming at the key and difficult problem of information fusion in multi-modal emotion recognition, some mainstream and high-efficiency information fusion strategies are introduced from four fusion levels: data-level fusion, feature-level fusion, decision-level fusion, and model-level fusion. By exemplifying representative multi-modal mixing examples from three perspectives: the mixing of multiple external presentation modalities, the mixing of multiple neurophysiological modalities, and the mixing of neurophysiology and external presentation modalities, it fully demonstrates that multi-modality is more capable of emotional discrimination and emotional representation than single-modality. At the same time, some thoughts on the conversion of multi-modal recognition methods to engineering technology applications are put forward. Finally, based on the analysis and grasp of the current situation of emotion recognition research, the ways and strategies for improving and enhancing the performance of the emotion recognition models are discussed and prospected.

Keywords: emotion recognition; emotion description model; emotion inducing mode; information fusion; fusion strategy; emotion representation; modality blend

收稿日期: 2020-01-30.
基金项目: 国家自然科学基金面上项目 (61876067); 广东省自然科学基金面上项目 (2019A1515011375); 广州市科技计划项目重点领域研发计划项目 (202007030005).
通信作者: 潘家辉, E-mail: panjh82@qq.com.

1 相关研究

1.1 背景与研究意义

情绪, 是一系列主观认知经验的高度概括, 由

多种感觉、思想和行为等产生的生理心理状态。人们在交流过程中无时无刻不传递着大量的情绪信息。从认知神经科学角度来看，情绪也属于经典认知的一种。情绪在人与人之间的沟通中意义重大，而在人机交互中，情绪识别是实现人性化必不可少的部分。

1995年，Picard等[1]提出了“情感计算”，情感计算要赋予计算机像人一样的观察理解和生成情感特征的能力，最终使得计算机像人一样进行自然亲近和生动交互。情感计算逐渐演变成高级人机交互的关键技术，而情感计算的子领域情绪识别更是人工智能领域中日益受到重点关注的研究方向。

情绪识别应用领域非常广阔，涉及日常生活的方方面面。在医学领域[2-3]，情绪识别能为精神疾病的诊断治疗提供依据。比如在意识障碍的诊断上，利用标准的行为量表不容易检测患者的意识状态，而计算机辅助评估意识障碍患者的情绪能帮助医生更好地做出诊断和治疗；在远程教育领域[4-5]，学生佩戴具有情绪识别功能的便携设备，以便教师可以监控学生在远程授课过程中的情绪状态，从而调整授课的进度和方式。在交通领域中[6-7]，对于那些需要高度集中注意力进行操作的工作人员，例如宇航员、长途旅行客车司机、飞行员等，他们的愤怒、焦虑、悲伤等负面情绪会严重影响他们的专注度，导致操作水平下降，造成交通事故的发生[8]。及时检测这类人员的情绪状态是避免事故发生的一种有效手段。

1.2 研究现状

近年来，随着人工智能和便携无创人体传感器技术的不断发展，多模态情绪识别已成为国内外情感计算领域的研究热点。目前，多模态情绪识别的研究主要集中于以下几个层面，基于多种外在行为表现模态、基于多种神经生理模态、基于神经生理状态和行为潜意识行为。我国在这些层面的情绪识别研究均取得了初步成果。

基于行为表现方面，中国地质大学吴敏教授情感计算团队在基于语音情感、人脸表情等基于外在行为的多模态情绪识别方面开展了研究[9-10]，并在此基础上对多维度情感意图理解、人机情感交互等领域也进行了较为深入的研究[11]。中国人民大学多媒体计算实验室融合了视听两种模态，采取条件注意融合策略进行连续维的情感预测研究[12]。同时着力于探索多种信道交互信息的有效替代、互补与干扰等融合机制，以实现人机交互自然性[13]。

兰州大学普适感知与智能系统团队主要研究基于人体生理信号和眼动、表情等行为表现模态的多模态知识建模及应用[14]。通过对生理信号等模态，组织与建模，研究适用于不同人群(抑郁症患者、心理高压患者)的模型，为准确、客观、实时地监控异常情感与心理状态变化。东南大学情感信息处理团队开展了基于脑电、表情、语音、肢体动作等模态的情绪分析研究[15-16]，在情感分析的基础探究和实践应用都取得了一定的成果。

基于神经生理信号的情绪识别的研究方面，上海交通大学吕宝粮教授团队建立了公开情绪数据集SEED[17]，并且在情绪最相关的信号频段和脑区、脑电的时间稳定性等基础研究方面做了不少相关的探索[18-19]。另外，中国科学院自动化研究所何晖光研究团队[20-21]和西南大学刘光远教授团队[22-23]都对生理信号的情绪识别进行了较为深入的研究。

2 情绪模型描述

情绪识别本质上是挖掘出有关情绪的特征数据与内在情绪状态的映射关系，如图1所示。情绪建模是指通过建立数学模型对情绪状态进行描述，从科学角度对情绪状态进行分类甚至量化。情绪模型的建立对于情绪测量有重要的意义，通过它可以对情绪状态做出较为准确的评估。2003年，Picard[24]就情绪建模描述进行了探讨。许多研究者提出了相应的情绪表征方法，依据表征方式的不同，可划分为离散情绪模型、维度情绪模型以及其他情绪模型，完整的描述如表1。

离散情绪模型指的是情绪由几种基本的离散的情绪构成，正如传统观念上的“喜怒哀乐”。Ekman等[25]认为情绪由悲伤、恐惧、厌恶、惊讶、高兴和愤怒组成，这6种情绪通过一定的组合模式可构成更为复杂的情绪类别。然而这种描述方式，无法科学地描述情绪的本质，也无法很好地从计算的角度来分析情绪状态。

维度情绪模型是将情绪状态映射为某一空间上的点，不同的情绪状态依据不同维度分布在空间中的不同位置，位置间的距离反映了不同的情绪状态间的差异。与离散情绪模型最大不同在于，维度情绪模型是连续的，具有表示情绪的范围广、能描述情绪的演变过程的优点[26]。

美国心理学家Johnston使用一维的坐标轴表征情绪，其正半轴为快乐，负半轴为悲伤[27]。而二维情绪模型则是从极性和强度两个维度区分情绪，极性指情绪具有正情绪和负情绪之分，强度

指情绪具有强烈程度和微弱程度的区别。目前使用最多的是 1980 年 Russell[28] 提出的效价–唤醒二维情绪模型。该模型将情绪划分为两个维度，分别为效价维度和唤醒维度。效价维度的负半轴表示消极情绪，正半轴表示积极情绪。唤醒维度的负半轴表示平缓的情绪，正半轴表示强烈的情绪。在三维情绪模型方面，当前认可度比较高的是 Mehrabian[29] 提出的高兴–唤醒–优势三维模型，该模型定义情绪具有愉悦度、唤醒度和优势度 3 个维度。

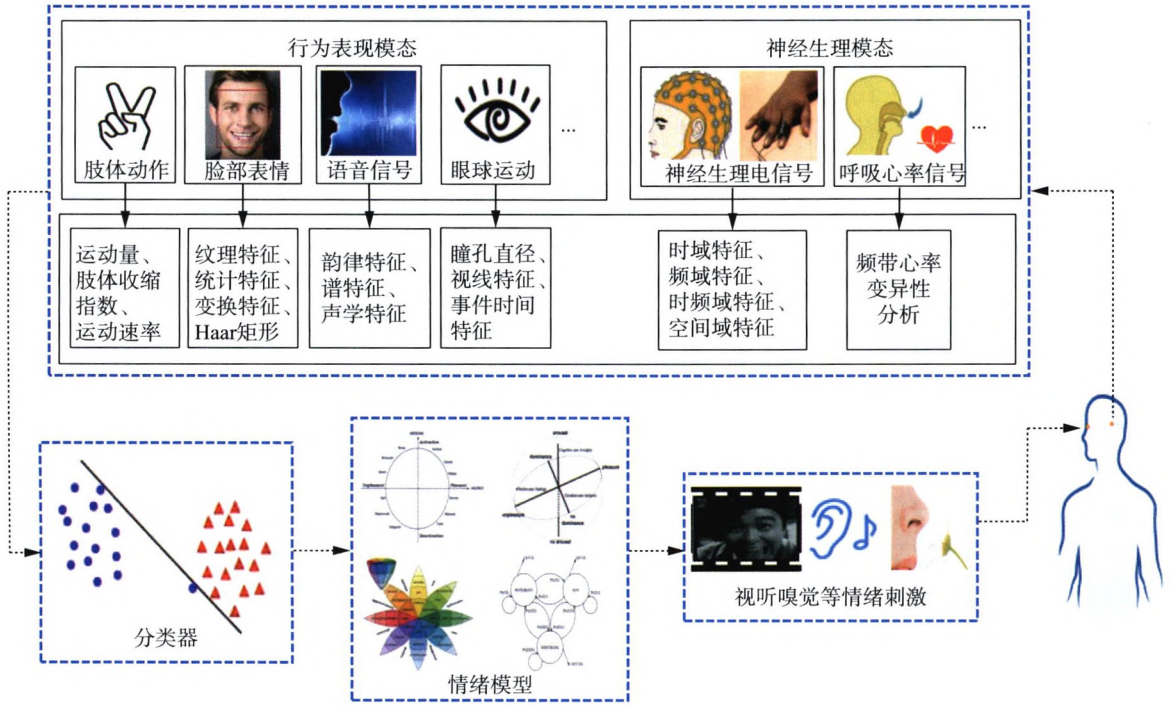

图 1　情绪识别研究流程

Fig. 1　Research process of emotion recognition

表 1　情绪描述模型

Table 1　Emotional description model

模型类别	情绪描述	提出者	基本内容
离散情绪	基本	Ekman[25]	悲伤、恐惧、高兴、厌恶、惊讶、愤怒
	复合	Plutchik[33]	复杂情绪由基本情绪的组合形成
维度情绪	一维	Johnston[27]	快乐–悲伤的正负两极
	二维	Russell[28]	效价–唤醒
	三维	Mehrabian[29]	高兴–唤醒–优势
	四维	Izard[34]	愉快度–紧张度–激动度–确信度
其他模型	Ortony-Clore-Collins	Ortony[30]	事件结果–仿生动作–对象感观的情感层次关系
	隐马尔可夫情感	Picard[31]	感兴趣–高兴–悲伤的环形情感状态转移模型
	分布式	Kesteren[32]	将特定的情感事件转换为相应的情绪状态

除此之外，许多心理学家和情绪研究学者根据各自不同的分析角度，提出了不同于上述两种模型的情绪描述模型。如 Ortony 等[30] 在 1988 年提出了 Ortony-Clore-Collins(OCC) 模型，Picard[31] 在 2000 年提出了隐马尔可夫模型 (hidden Markov model, HMM)。同年，Kesteren 等[32] 以情绪时间类型为分类标准，并针对外界刺激事件而建立了分布式情感模型。

3　情绪诱发方式

在进行情绪识别研究时，首要考虑的是如何才能有效地诱发被试者的情绪状态。目前的研究中，情绪的诱发方式主要有两种，分别是主体诱

发和事件诱发。

主体诱发指借助回忆相关情绪的事件来产生相关的情绪，这是一个由被试者主动诱发目标情绪的方式。Zhuang 等[35] 设计了 30 名被试者在观看每个电影片段 (共 18 个片段) 后，被试者被要求通过回忆每个电影中的特定场景来自我诱导 6 种离散情绪，包括喜悦、中立、悲伤、厌恶、愤怒和恐惧。Iacoviello 等[36] 也做了类似的自诱发情绪的研究，该研究是对厌恶情绪和中立情绪进行二分类实验，被试者通过回忆一种不愉快的气味进行自诱发情绪。主体诱发能充分反映被试者的主观情感体验，但难以控制情绪诱发的时间且易受外界干扰。

事件诱发是指使用不同情绪相关的刺激材料来诱发被试者的某一目标情绪状态，如图片、声音、视频、气味等。镜像神经元理论[37] 表明当一个人观察另一个人的活动时，其脑部也在做类似的活动。事件诱发正是根据该原理设计诱发情绪的方式。事件诱发能够更加高效地诱发被试者的目标情绪，因此其被广泛地用于设计诱发情绪方式。事件诱发相对常用的刺激方式包括图片刺激、听觉刺激以及多媒体材料。当前公认的图片情绪刺激材料有 Lang 等[38] 采集的国际情绪图片，听觉刺激材料库则有 Bradley 等[39] 采集的国际情绪数字化声音。

而目前公开的多模态情绪数据集大多采用多媒体材料来诱发情绪。其中采集了多种生理信号的最具代表性数据集的就是 DEAP 数据集[40]。DEAP 数据集诱发方式要求 32 位被试者在观看 40 段 1 min 的音乐视频片段后填写自我情绪评定量表 (self-assessment manikin, SAM)。收集了生理信号和行为表现的数据集有 MAHNOB-HCI 数据集[41]，是记录由情绪电影刺激产生神经生理信号和行为表现的多模态数据库。基于多种行为表现模态的数据集有 Martin 等[42] 制作的视听数据集 eNTERFACE'05，共有 42 位被试者，可以使用该数据库作为测试和评估视频、音频或者视听情感识别算法。

4 多模态融合策略

在单模态情绪识别领域，单模态信息易受各种噪声的影响，难以完整地反映情绪状态[43]。文献 [44] 使用统计方法深入讨论单模态和多模态情绪识别精度。他们比较了不同算法在不同数据集上的准确性，85% 多模态情绪识别系统比最佳单模态对应系统更准确，平均提高 9.83%(中位数为 6.60%)，综合分析出多个信号及其相互依赖可以构建出更准确地反映人类情感表达潜在本质的模型。与文献 [44] 不同，Poria 等[45] 在充分讨论了单模态识别方法的现状基础上，根据实验的准确性对基于同一数据集的多模态融合情绪识别研究与单模态情绪识别进行了横向比较，同样有力证明了高效的模态融合能极大地提高情绪识别系统的健壮性。利用不同类别的信号相互支持，对互补信息进行融合处理，能够有效地提高最终的识别效果[46]。根据目前已有的研究，模态融合的方式大致可分为 4 种，分别是数据级融合 (传感层融合)、特征级融合、决策级融合、模型层融合。

4.1 数据级融合

数据级融合[47-48]，又称传感器层融合。数据级融合是直接对各个传感器采集到的最原始的、没有经过特殊处理的数据进行组合，从而构造一组新的数据，如图 2。

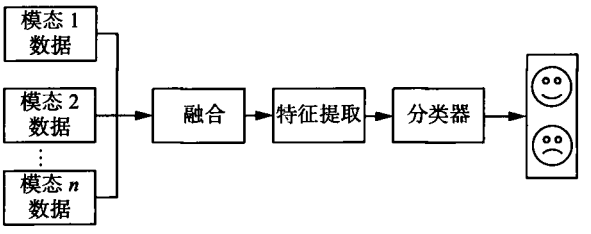

图 2 数据级融合

Fig. 2 Data level fusion

目前数据级融合处理的方法有数值处理、参数估计。具体为使用线性、非线性估计和统计运算方法对来自多个数据源的数据进行计算处理[49]。其优点是：可以很好地保留各个模态传感器上的数据信息，避免信息的丢失，保持信息的完整性。但其缺点也是明显的，因为数据是在原始状态下进行融合，因此处理过程极为繁琐复杂。

4.2 特征级融合

特征级融合[50-51] 是将多种模态数据经过提取、构建成相应的模态特征之后，再拼接成一个集成各个模态特征的特征集，如图 3。

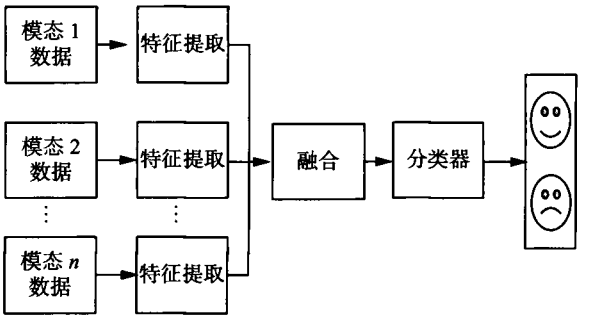

图 3 特征级融合

Fig. 3 Feature level fusion

在特征级层面，常用的融合策略是将经特征提取后全部模态特征数据级联为特征向量后再送入一个情感分类器。如 Emerich 等[52]将长度归一化的语音情感特征和面部表情特征级联起来，构造一个特征向量。实验结果表明语音信息系统提取的特征包含有价值的情感特征，这些特征是无法从视觉信息中提取出来的。当这两种模式融合时，情绪识别系统的性能和鲁棒性都得到了提高，但这种直接级联拼接的方式导致了新特征空间不完备，融合后维数过高，当特征维数达到一定规模后，模型的性能将会下降。为此，Yan 等[53]提出了一种基于稀疏核降秩回归 (sparse kernel reduced-rank regression, SKRRR) 特征级融合策略，SKRRR 方法是传统降秩回归 (RRR) 方法的非线性扩展，将预测量和响应特征向量分别通过两个非线性映射映射到两个高维特征空间中进行核化。openSMILE 特征提取器和 SIFT 描述子分别从语音模态和面部表情模态中提取有效特征，然后使用 SKRRR 融合方法融合两种模态的情感特征。而 Mansoorizadeh 等[54]提出了一种异步的特征级融合方法，在单个信号测量之外创建一个统一的混合特征空间，他们使用提出的方法从语音韵律和面部表情来识别基本的情绪状态。结果表明，与基于单模态人脸和基于语音的系统相比，基于特征级融合的系统性能明显提高。

当模态信息针对同一内容而又不互相包含时，特征级融合方法虽然能最大限度地保留原始信息，在理论上能达到最佳的识别效果[49]，但是其没有考虑到不同模态情绪特征之间的差异性。

4.3　决策级融合

决策级融合[55-56]是找出各个模态的可信度，再进行协调、联合决策，如图 4。决策级融合与特征级融合相比，更容易进行，但关键是要探究各个模态对情绪识别的重要度。

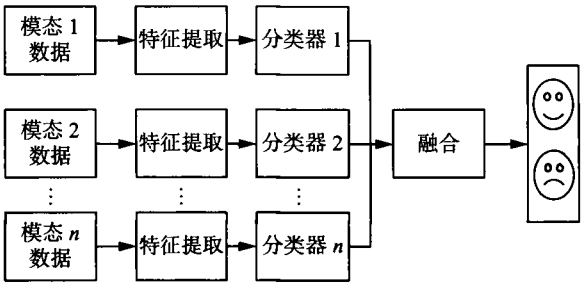

图 4　决策级融合

Fig. 4　Decision level fusion

决策级所采用的融合策略有基于统计学规则[57] (总和规则、乘积规则、最大/最小/中位数规则等)、枚举权重[58-59]、自适应增强[60-61]，贝叶斯推论及其推广理论 (Dempster-Shafer 理论[62]、动态贝叶斯网络[63])、模糊积分[64] 等。

Huang 等[65] 同时使用枚举权重及 adaboost 两种不同决策级融合策略来比较情绪识别效果，使用面部表情分类器和脑电图分类器作为增强分类器的子分类器，并分别应用于两个学习任务 (效价和唤醒)。结果表明这两种方法都能给出最后的效价和唤醒结果，在公开数据集 DEAP、MAHNOB-HCI 以及在线应用均取得不错的效果。

基于统计规则和概率理论均依赖于所有分类器相互独立的假设，这与实际情况不符。因此，预测结果在一定程度上是不准确的。Lu 等[66] 采用了一种称为模糊积分的融合策略。模糊积分是关于模糊测度的实函数的积分。实验发现眼球运动特征和脑电图对情绪识别具有互补作用，模糊积分融合策略的最佳准确率为 87.59%，相比于其他融合方式，模糊积分融合能显著提高情绪识别的准确性。通常情况下，多种模态间的信息并非完全独立，决策级融合会丢失不同模态之间的相关性，所以在实际应用环境下识别的结果未必会比单模态识别的效果好。

4.4　模型级融合

模型级模态融合[67-68]不依赖于以上 3 种融合层次的体系结构。决策级融合关键在于找出不同模态在决策阶段的可信程度，但模型级融合并不需要重点去探究各模态的重要程度，而是根据模态特性需要建立合适的模型，联合学习关联信息。特征级融合则主要先通过构建特征集合或混合特征空间，再送入到分类模型进行分类决策。模型级融合可以将不同模态特征分别输入到不同模型结构再进行进一步特征提取，如 Zheng 等[17] 采用将堆叠的受限玻尔兹曼机展开成深度置信网络，首先以手工提取出来的脑电和眼动特征分别作为两个玻尔兹曼机的输入并从神经网络中学习两种模式的共享表示，实验结果表明，基于深度神经网络的模型级融合能显著提高性能。总的来说，模型级融合相较于决策级融合和特征级融合最大的优势在于可以灵活地选择融合的位置。

目前的模型级融合主要采取的策略是通过构建深度网络模型，建立多层结构，逐层学习可以学习到更加复杂的变换，从而可以拟合更加复杂的特征，增加非线性表达能力。Zhang 等[67] 提出一种充分利用深度神经网络强大的特征学习能力的混合深度学习模型，将视听数据经卷积神经网络 (convolutional neural networks, CNN) 和 3DCNN (three dimensional convolutional neural networks) 产

生视听片段特征，然后将视听片段特征融合到深度置信网络中，联合学习了一种视听特征表示，在情感识别任务上比先手工特征再深度学习融合方法表现得更好。

5 多模态混合形式

情绪是一个人在某种环境状态下的行为和心理综合表现。从人体的表现角度来讲情绪表现主要分为神经生理层面及外在行为层面。神经生理层面情绪识别利用人的脑电信号 (electroencephalogram, EEG)、血容量搏动 (blood volume pulsation, BVP)、瞳孔变化、肌电 (electromyogram, EMG)、心电 (electrocardiogram, ECG)、皮肤电 (galvanie skin response, GSR)、皮温 (skin temperature, SKT)、呼吸 (respire, RSP) 等内在的神经生理信号变化进行识别，而行为表现层面情绪识别，则可以通过面部表情、姿态手势、语音等外在表现行为进行识别。多模态数据的混合可以为决策提供更多信息，从而提高了决策总体的准确率[69]。表 2 针对不同的模态类型对情绪识别的效果进行粗略的比较。表 3 总结了近年基于公开数据集的多模态情绪识别相关研究。

表 2 不同模态类型下的情绪识别效果对比

Table 2 Comparisons of emotion recognition effects under different model types

模态属性	模态类型	采集难度	信噪比	识别率
神经生理	脑中枢变化	高	低	高
	外周生理变化	高	高	中
行为表现	脸部表情	低	高	中
	语音语调	低	中	中
	肢体动作	低	高	低

表 3 不同混合模式的多模态情绪识别方案及效果

Table 3 Schemes and effects of multimodal emotion recognition with different mixed modes

模态混合模式	公开数据集	作者及文献	模态与特征提取	模态融合层次 (融合策略)	分类算法	准确率/% (分类标准/#)
多种行为表现模态混合	eNTERFACE'05	Nguyen等[70]	语音+人脸: 串联三维卷积神经网络	特征级融合 (双线性池理论)	DBN	90.85(Avg./6)
		Dobrišek等[71]	语音: open SMILE; 人脸: 子空间的构建与匹配	决策级融合 (加权乘积规则)	支持向量机	77.50(Avg./6)
		Zhang等[72]	语音: CNN 人脸: 3D-CNN	模型级融合 (DBN融合网络)	支持向量机	85.97(Avg./6)
多种神经生理模态混合	DEAP	Koelstra等[40]	EEG+PPS: Welch's method	决策级融合 (多数类别投票)	朴素贝叶斯	58.6 (Valence/2) 64.4 (Arousal/2)
		Tang等[73]	EEG:频域微分熵; PPS: 时域统计特征	模型级融合 (BLSTM编码融合)	支持向量机	83.8 (Valence/2) 83.2 (Arousal/2)
		Yin等[74]	EEG+PPS: 堆叠自编码器	模型级融合 (多层网络融合)	Ensemble- SAE	83.0 (Valence/2) 84.1 (Arousal/2)
行为表现与神经生理模态混合	MAHNOB - HCI	Soleymani等[75]	EEG: 频域功率谱密度; 眼动: 瞳孔功率谱眨眼、注视等时域统计特征	决策级融合 (置信度求和融合)	支持向量机	76.4 (Valence/3) 68.5 (Arousal/3)
		Huang等[65]	EEG: 小波变换提取功率谱; 人脸: CNN深度特征提取	决策级融合 (自适应增强)	SVM(脑电) CNN(表情)	75.2 (Valence/2) 74.1 (Arousal/2)
		Koelstra等[58]	EEG: 支持向量机递归特征消除; 人脸: 动作单元映射	特征级融合 (特征拼接)	朴素贝叶斯	73.0 (Valence/2) 68.5 (Arousal/2)

注: 1) # 分类类别数; 2) Avg. 平均准确率; 3) Valence, Arousal 维度情绪模型的效价和唤醒维度; 4) PPS, peripheral physiological signals (外周生理信号); 5) DBN, deep belief networks (深度置信网络); 6) BLSTM, bi-directional long short-term memory (向长短时记忆); 7) Ensemble-SAE, ensemble-stacked auto encoder (堆叠自编码器)。

5.1 多种行为表现模态混合

情绪表达的途径多种多样，情绪的载体也极其丰富。情绪过程与其他心理过程不一样，表现在情绪活动过程中伴随着一定的行为表现。James[76]认为情绪是人们对于自己身体所发生变化的一种感觉，先有身体的变化然后才有情绪的感知，任

何情绪的产生都伴随着一定的身体变化，包括面部表情、肌肉紧张、内脏活动等，从而可以通过行为表现感知推测情绪变化。在众多行为表现模态信息中，语音语调[77-78]、脸部表情[79-80]、肢体动作[81-82]这几类模态在情绪识别中取得较好的效果。

语音信息具有复杂多样的声学属性。人在不同情绪状态下的语调、语速等声学属性将显现其不同的情绪特性。从语音语调等声学属性中提取出能体现情绪差异的参数特征即可进行情绪识别。面部表情是一种最为直观的情绪表达方式，对人脸信息进行分析就可以解读人的情绪。表情是所有面部肌肉变化所组成的模式。面部表情模式能精细地表达不同性质的情绪和情感。此外，人在特定的情绪状态下往往会伴随着相应的肢体动作，比如人在失落时捶胸顿足。

因此也有不少研究者在情绪识别框架中整合来自各种行为表现特征。Kaliouby 和 Robinson[83]提出了一个基于视觉的计算模型，从头部运动和面部表情来推断行为相应的精神状态。Castellano 等[84]融合了面部表情、语音和手势这 3 种模态，对 8 种情绪状态进行识别，准确率达到 78.3%。与 Castellano 不同的是，Scherer 等[85]在面部表情、语音、手势 3 种模态的基础上加上身体姿态，对 14 种不同情绪进行识别，分类准确率达到 79%。

5.2　多种神经生理模态混合

美国心理学之父 James[76]1884 年首次定义了情绪，他认为情绪是人的身体上发生某种变化的一种综合效应，一切情绪都引起了身体上的一定变化。人的情绪总是伴随着心理活动的产生，而任何心理活动又与我们的神经生理状态息息相关。

因为情绪的变化会引起相应人体神经中枢的活动，进而反应为脑区的活跃，因此可以通过采集脑电图、核磁共振图像等脑成像技术来监测大脑结构和功能活动状态的变化。在这些大脑信号中，脑电被证实能够更好地提供反映情绪状态变化的信息[86-87]。Davidson 等[88]提出脑额叶皮层活跃活动能表征人体积极消极情绪变化。不少研究[89-90]也印证了 EEG 能够很好地反映情绪的波动变化。

为了利用脑电图信号识别情绪，需要执行以下步骤：记录用户大脑中观察到的电压变化；去除记录信号中的噪音和伪影；分析结果数据，提取相关特征；利用计算出的特征对分类器进行训练，得到原始脑信号与情绪状态的映射关系[91]。

生理唤醒是情绪与情感产生的生理反应，可以用于衡量生理激活水平。在不同情绪状态下，人体的生理反应模式是不一样的，如高兴时心跳节律平稳；恐惧时，血压升高、心跳加速。Picard 等[1]首次应用计算机模式识别的思路来解决情感生理识别问题，并使用图片诱发材料，采集一个被试者的 4 种生理信号 (肌电、脉搏、皮肤电导、呼吸)，提取其中的生理信号特征对 8 种情感状态进行分类，总体识别率为 81%。

神经生理模态直接受神经系统和内分泌系统支配，不受主观影响，难以伪装，因此与脸部表情等行为模态相比更具真实性、客观性。Chanel 等[92]结合脑电信号与皮肤电传导、血压、心率、呼吸等外围生理信号对积极情绪、消极情绪和中立情绪进行情绪识别，该实验达到 80% 的准确率，可以发现将外围生理信号特征与脑电特征融合，情绪识别效果明显。显然，脑电信号有着与情绪相关的大量信息，而外围神经系统生理信号也有与情绪相关的信息，两者的结合可以增强情绪识别的效果。

Chen 等[93]通过脑电图增强周围生理信号的情感识别，这是利用脑电信号特征作为特殊信息从周围信号中识别情绪的新方法，首先提取周围生理特征和脑电图特征，利用典型相关分析方法，结合脑电特征构建了一个新的边缘生理特征空间。最后，训练支持向量机模型，只利用边缘生理特征从构建的边缘生理特征空间来识别情绪。该方法在 DEAP 数据集和 MAHNOB-HCI 数据集上的实验结果表明，提出的以脑电图特征为优先信息的方法优于仅从周围生理信号识别情绪的方法。

5.3　神经生理模态与行为表现模态混合

上述研究分别以外在行为和神经生理信号作为输入信号来进行情绪识别。而基本情绪理论[94]已经证明，当情绪被激发时，会激活人类的多种生理和外在行为反应系统。比如高兴时会出现特定的面部表情、语音语调、手势姿态及相应神经生理状态的变化。

随着神经生理及行为表现信号采集设备的发展与推广，基于神经生理模态与行为表现模态的混合模式受到越来越多情绪识别研究者的关注。Huang 等[95]基于效价-唤醒情绪模型，将面部表情和脑电图相结合，采用了一种多模式情绪识别框架。对于面部表情检测，他们遵循用于多任务卷积神经网络架构的迁移学习方法来检测效价和唤醒状态。对于脑电图检测，通过支持向量机分类

器检测到两个学习目标。最终的混合模态识别准确率均达到82%左右。

就应用型情感识别系统而言，较具代表性的是2007年Kapoor等[96]提出的学习伴侣系统，该系统不仅使用面部表情,还结合了皮肤电导、姿势等信息。该系统通过识别用户是否处于失落状态而自动调整呈现的学习内容，给用户提供了极大的协助[23]。Liu等[97]提出了一种将眼动和脑电图相结合的多模态情绪识别框架，利用提取出的特征对3类(正面、中性和负面)情绪进行识别。他们所提出的双模态深度自编码器能够充分利用脑神经信号(脑电)与外围生理信息(眼球运动)的互补性来提高模型的情绪识别准确率，最终的准确率达到91.01%。

6 工程应用的讨论与建议

要实现具有一定情绪识别功能的人机交互应用，建立自然的人机交互过程，需要满足数据高质量获取、识别过程中准确性和鲁棒性、在线识别的时效性这3个方面的要求。因此在实际工程应用方面，我们针对数据的获取与处理、系统应用模型泛化设计、实时在线系统设计等方面的挑战，提出了操作性建议以及在实践应用需要考虑的问题。

6.1 多模态情绪数据的获取与处理

情绪识别的实际应用过程中，如何快速准确地采集高质量的多模态信号是一个关键问题。2017德国柏林工业大学的脑机接口研究团队发布了一款可同时采集脑电信号、近红外光谱，以及其他常规生理参数(如心电、肌电和加速度等)的无线模块化硬件架构[98]。类似的高精准、便携式、可扩展性的多生理参数采集硬件架构，是多模态情绪识别研究走向工程应用是先决条件。

使用多模态进行情绪识别的意义在于通过不同模态之间的情绪信息互补，融合各模态潜在的共享信息，进而提升情绪识别任务的有效性。但混合了各种异质高维的特征，带来的最直接问题就是维数灾难，从而显著降低了模型的性能。同时，数据中包含的负相关特征，也将会极大影响模型识别的准确率，因此需要针对不同的数据源采用合适的降维方法，保留数量适当且分类效果好的特征。

6.2 系统应用模型泛化设计

情绪刺激反应普遍都存在着个体差异性和非平稳特性，即不同个体在面对同一刺激产生的情绪反应不尽相同，从这种差异中寻找稳定的情绪反应与模态特征之间的对应关系，构建普适的稳健的情感模型是当前工程应用研究中急需解决且具有挑战性的问题。文献[99]提出了一种归一化数据转换方法，将模态信号中依赖于个体的分量去除，构建不依赖于个体的共用特征空间，从而消除模态数据特征个体差异性所带来的影响，实现了跨被试者、可迁移的非个体依赖的情感生理状态识别，提出了更接近实际应用的情绪识别方法。

目前大多的情绪识别模型训练需要花费大量前期时间进行系统标定，这极大地限制了工程应用的场景，同时无标签多模态数据的获取相对比较容易，因此基于无标签样本的多模态学习对于工程应用具有重要意义。Du等[21]提出了一个多视角深层生成多视图情绪识别模型，将无标签的半监督分类问题转化为一个专门的缺失数据输入任务，其中丢失的视图被视为一个潜在变量，并在推理过程中被整合出来。

6.3 多模态在线系统设计

当我们需要借助多种模态对情绪状态进行综合分析时,首先工程应用过程中多种模态数据采集仪器设备分别记录的每种模态信号在时间上必须准确对应或同步。最直接方式是保证每种模态采集的频率一致，但在工程应用中要保证异质多源模态数据采集频率一致是不现实的。清华大学高小榕团队[100]使用伪随机序列编码信号同时标记视频和脑电信号的数据同步方法，完成了眼动仪与脑电同步采集平台的搭建。

与多模态情绪识别离线算法分析不同，多模态情绪识别在实际的工程应用中我们更强调系统需要实时针对当前新的样本，不断学习新的特征并保存大部分已经学习到的知识，适当调整模型结构，从而不断提升模型的泛化能力。大连理工大学赵亮[101]提出一种多模态数据增量共聚类融合算法，设计了3种增量聚类策略，即簇创建、簇合并和聚类划分，对多模态数据进行增量聚类融合，同时设计一种自适应的模态权重机制，在共聚类融合过程中对模态权重进行动态调整以应对多模态数据处理的实时性问题。

7 未来展望

本文对目前多模态情绪识别研究现状进行了综述，包括情绪模型的描述、情绪诱发的范式设计、按多模态的融合层次介绍了各层次下的融合策略、不同模态类型混合的情绪识别方案，论证了多模态融合技术可以极大提高情绪识别效果，同时对多模态情绪识别方法转为工程技术应用提

出了一些思考，但无论是算法研究还是工程应用都存在着一些可以努力尝试和改进的方向。下面将提出值得进一步探究和尝试的方面，旨在对相关研究者起到一定的启发作用。

1) 量化情绪描述模型

现在进行的情感识别工作多集中在简单的情感分类，对于情绪强弱程度方面，目前并没有太多的研究者去探究。精准地识别情绪强弱能够更加明确情绪状态，具有非常重要的现实意义。

2) 尝试多模态的混合融合策略

混合融合属于深层融合的方式，能够很好地挖掘不同模态之间的非线性关联。我们可以同时考虑结合深度学习，针对不同模态，在不同的融合层次进行相关融合，进而实现情绪识别。

3) 针对不同实际应用场景，选择合适模态

利用情绪信息之间的互补性无疑是可以提高识别准确率的，但我们也有必要根据不同的应用环境选择不同的模态进行混合识别。例如针对健康人群，采集脸部表情这一对于健康人最能反映情绪的外部行为表现的基础上，结合能充分揭露用户的内在认知状态和情感状态的脑电信号，增强了现实应用的可用性和可靠性。针对意识障碍患者等运动障碍患者，基于脑电信号的方法被认为是更可靠的解析情绪的方法，因为它们是对中枢神经系统和自主神经系统的客观测量。

4) 借助深度学习算法提取情感特征

目前情绪识别研究大多是从数字信号分析的角度出发，探究情绪与信号特征的关系。深度学习具有提取鲁棒性特征的潜力，大量基于深度学习的情绪识别算法被提出。利用海量数据学习到的特征比人工构造的特征更能反映数据的内在本质，能在一定程度上克服特征冗余问题，从而极大提高情绪识别效果。

5) 设计更加有效的情绪诱发方式

主体诱发难以控制情绪诱发的时间且易受外界干扰，使用情绪诱发材料的事件诱发情绪和现实生活中个体感受到情绪的情境相差甚远，不同通道的情绪刺激对被试者的认知和行为产生的影响也不尽相同。在未来，一方面可以通过设计特殊情境来操控诱发被试者情绪体验的范式，例如开发情绪诱发游戏、基于虚拟现实技术的沉浸式情境诱发情绪等。另一方面，由于基于高度控制素材诱发情绪的范式推广的局限性，可通过让被试者处于一种自然情境下并使用相对开放素材诱发情绪，探究真实场景的模态信息与情绪状态之间的关系。

参考文献：

[1] PICARD R W, HEALEY J. Affective wearables[J]. Personal technologies, 1997, 1(4): 231–240.

[2] PAN J, XIE Q, HUANG H, et al. Emotion-related consciousness detection in patients with disorders of consciousness through an EEG-Based BCI system[J]. Frontiers in human neuroence, 2018, 12: 198–209.

[3] HUANG H, XIE Q, PAN J, et al. An EEG-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness[J]. IEEE transactions on affective computing, 2019, 99: 1–10.

[4] WANG S, PHILLIPS P, DONG Z, et al. Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm[J]. Neurocomputing, 2018, 272: 668–676.

[5] WANG W, WU J. Notice of retraction emotion recognition based on CSO&SVM in e-learning[C]//Poceedings of the 2011 Seventh International Conference on Natural Computation. Shanghai, China, 2011: 566–570.

[6] LIU W, QIAN J, YAO Z, et al. Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection[J]. Future internet, 2019, 11(5): 115.

[7] BORIL H, OMID SADJADI S, KLEINSCHMIDT T, et al. Analysis and detection of cognitive load and frustration in drivers' speech[J]. Proceedings of interspeech, 2010: 502–505.

[8] 陆怡菲. 基于脑电信号和眼动信号融合的多模态情绪识别研究 [D]. 上海: 上海交通大学, 2017.
LU Yifei. Research on multi-modal emotion recognition based on eeg and eye movement signal fusion[D]. Shanghai: Shanghai Jiaotong University, 2017.

[9] LIU Z, WU M, TAN G, et al. Speech emotion recognition based on feature selection and extreme learning machine decision tree[J]. Neurocomputing, 2018, 10: 271–280.

[10] LIU Z, WU M, CAO W, et al. A facial expression emotion recognition based human-robot interaction system[J]. Journal of automation: english version, 2017, 4(4): 668–676.

[11] LIU Z, PAN F, WU M, et al. A multimodal emotional communication based humans-robots interaction system[C]//Poceedings of the Control Conference. Chengdu, China, 2016: 6363–6368.

[12] CHEN S, JIN Q. Multi-modal conditional attention fusion for dimensional emotion prediction[C]//Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, the Netherlands, 2016: 571–575.

[13] CHEN S, LI X, JIN Q, et al. Video emotion recognition in the wild based on fusion of multimodal features[C]// Proceedings of the 18th ACM International Conference

on Mmultimodal Interaction. Tokyo, Japan, 2016: 494–500.

[14] ZHANG X, SHEN J, DIN Z U, et al. Multimodal depression detection: fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble[J]. IEEE journal of biomedical and health informatics, 2019, 23(6): 2265–2275.

[15] ZONG Y, ZHENG W, HUANG X, et al. Emotion recognition in the wild via sparse transductive transfer linear discriminant analysis[J]. Journal on multimodal user interfaces, 2016, 10(2): 163–172.

[16] ZHANG T, ZHENG W, CUI Z, et al. Spatial–temporal recurrent neural network for emotion recognition[J]. IEEE transactions on systems, man, and cybernetics, 2019, 49(3): 839–847.

[17] ZHENG W, LIU W, LU Y, et al. Emotionmeter: A multimodal framework for recognizing human emotions[J]. IEEE transactions on cybernetics, 2018, 49(3): 1110–1122.

[18] ZHENG W, ZHU J, LU B. Identifying stable patterns over time for emotion recognition from EEG[J]. IEEE transactions on affective computing, 2019, 10(3): 417–429.

[19] YAN X, ZHENG W, LIU W, et al. Investigating Gender differences of brain areas in emotion recognition using LSTM neural network[C]//Poceedings of the International Conference on Neural Information Processing. Guangzhou, China, 2017: 820–829.

[20] LI J, QIU S, SHEN Y, et al. Multisource transfer learning for cross-subject EEG emotion recognition[J]. IEEE transactions on systems, man, and cybernetics, 2019, 50(7): 1–13.

[21] DU Changde, DU Changying, LI J, et al. Semi-supervised bayesian deep multi-modal emotion recognition[J]. arXiv preprint arXiv: 170407548, 2017.

[22] 程静. 基本情感生理信号的非线性特征提取研究 [D]. 重庆: 西南大学, 2015.
CHENG Jing. Research on nonlinear feature extraction of basic emotional physiological signals[D]. Chongqing: Southwest University, 2015.

[23] 温万惠. 基于生理信号的情感识别方法研究 [D]. 重庆: 西南大学, 2010.
WEN Wanhui. Research on emotion recognition method based on physiological signals[D]. Chongqing, Southwest university, 2010.

[24] PICARD R W. Affective computing: challenges[J]. International journal of human-computer studies, 2003, 59(1-2): 55–64.

[25] EKMAN P E, DAVIDSON R J. The nature of emotion: fundamental questions[M]. Oxford: Oxford university press, 1994.

[26] 高庆吉, 赵志华, 徐达, 等. 语音情感识别研究综述 [J]. 智能系统学报, 2020, 15(1): 1–13.
GAO Qingji, ZHAO Zhihua, XU Da, et al. Review on speech emotion recognition research[J]. CAAI transactions on intelligent systems, 2020, 15(1): 1–13.

[27] JOHNSTON V S. Why we feel: The science of human emotions[M]. New York: Perseus publishing, 1999.

[28] RUSSELL J A. A circumplex model of affect[J]. Journal of personality and social psychology, 1980, 39(6): 1161.

[29] MEHRABIAN A. Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies[M]. Cambridge: Oelgeschlager Gunn & Hain Cambridge, MA, 1980.

[30] ORTONY A, CLORE G L, COLLINS A. The cognitive structure of emotion[J]. Contemporary sociology, 1988, 18(6): 2147–2153.

[31] PICARD R W. Affective computing[M]. Cambridge: MIT press, 2000.

[32] VAN KESTEREN A, OPDEN AKKER R, POEL M, et al. Simulation of emotions of agents in virtual environments using neural networks[J]. Learning to behave: internalising knowledge, 2000: 137–147.

[33] PLUTCHIK R. Emotions and life: Perspectives from psychology, biology, and evolution[M]. Washington: American Psychological Association, 2003.

[34] IZARD. Human emotions[M]. Berlin: Springer Science & Business Media, 2013.

[35] ZHUANG N, ZENG Y, YANG K, et al. Investigating patterns for self-induced emotion recognition from EEG signals[J]. Sensors, 2018, 18(3): 841.

[36] IACOVIELLO D, PETRACCA A, SPEZIALETTI M, et al. A real-time classification algorithm for EEG-based BCI driven by self-induced emotions[J]. Computer methods and programs in biomedicine, 2015, 122(3): 293–303.

[37] RIZZOLATTI G, CRAIGHERO L. The mirror-neuron system[J]. Annu rev neurosci, 2004, 27: 169–192.

[38] LANG P J, BRADLEY M M, CUTHBERT B N. International affective picture system (IAPS): Technical manual and affective ratings[J]. NIMH center for the study of emotion and attention, 1997, 1: 39–58.

[39] BRADLEY M, LANG P J. The International affective digitized sounds (IADS)[M]. Rockville: NIMH center, 1999.

[40] KOELSTRA S, MUHL C, SOLEYMANI M, et al. Deap: A database for emotion analysis; using physiological signals[J]. IEEE transactions on affective computing, 2011, 3(1): 18–31.

[41] SOLEYMANI M, LICHTENAUER J, PUN T, et al. A

multimodal database for affect recognition and implicit tagging[J]. IEEE transactions on affective computing, 2012, 3(1): 42–55.

[42] MARTIN O, KOTSIA I, MACQ B. The eNTERFACE'05 audio-visual emotion database[C]//Poceedings of the international conference on data engineering workshops IEEE computer society. Atlanta, USA, 2006: 8.

[43] 何俊, 刘跃, 何忠文. 多模态情感识别研究进展 [J]. 计算机应用研究, 2018, 35(11): 3201–3205.
HE Jun, LIU Yue, HE Zhongwen. Research progress of multimodal emotion recognition[J]. Computer application research, 2018, 35(11): 3201–3205.

[44] D'MELLO S K, KORY J. A review and meta-analysis of multimodal affect detection systems[J]. ACM computing surveys (CSUR), 2015, 47(3): 43.

[45] PORIA S, CAMBRIA E, BAJPAI R, et al. A review of affective computing: from unimodal analysis to multimodal fusion[J]. Information fusion, 2017, 37: 98–125.

[46] 黄泳锐, 杨健豪, 廖鹏凯, 等. 结合人脸图像和脑电的情绪识别技术 [J]. 计算机系统应用, 2018, 27(2): 9–15.
HUANG Yongrui, YANG Jianhao, LIAO Pengkai, et al. Emotion recognition technology combining face image and EEG[J]. Computer system application, 2018, 27(2): 9–15.

[47] 孙皓莹, 蒋静坪. 基于参数估计的多传感器数据融合 [J]. 传感器技术, 1995, 6: 32–36.
SUN Haoying, JIANG Jingping. Multi-sensor data fusion based on parameter estimation[J]. Sensor technology, 1995, 6: 32–36.

[48] MINOTTO V P, JUNG C R, LEE B. Multimodal multichannel on-line speaker diarization using sensor fusion through SVM[J]. IEEE transactions on multimedia, 2015, 17(10): 1694–1705.

[49] 张保梅. 数据级与特征级上的数据融合方法研究 [D]. 兰州: 兰州理工大学, 2005.
ZHANG Baomei. Research on data fusion methods at data level and feature level[D]. Lanzhou: Lanzhou University of Technology, 2005.

[50] PORIA S, CHATURVEDI I, CAMBRIA E, et al. Convolutional MKL based multimodal emotion recognition and sentiment analysis[C]//Poceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM). Barcelona, Spain, 2016: 439–448.

[51] HAGHIGHAT M, ABDELMOTTALEB M, ALHALABI W. Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition[J]. IEEE transactions on information forensics and security, 2016, 11(9): 1984–1996.

[52] EMERICH S, LUPU E, APATEAN A. Bimodal approach in emotion recognition using speech and facial expressions[C]//Poceedings of the 2009 International Symposium on Signals, Circuits and Systems. Iasi, Romania, 2009: 1–4.

[53] YAN J, ZHENG W, XU Q, et al. Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech[J]. IEEE transactions on multimedia, 2016, 18(7): 1319–1329.

[54] MANSOORIZADEH M, CHARKARI N M. Multimodal information fusion application to human emotion recognition from face and speech[J]. Multimedia tools and applications, 2010, 49(2): 277–297.

[55] ZHALEHPOUR S, ONDER O, AKHTAR Z, et al. BAUM-1: A spontaneous audio-visual face database of affective and mental states[J]. IEEE transactions on affective computing, 2016, 8(3): 300–313.

[56] WU P, LIU H, LI X, et al. A novel lip descriptor for audio-visual keyword spotting based on adaptive decision fusion[J]. IEEE transactions on multimedia, 2016, 18(3): 326–338.

[57] GUNES H, PICCARDI M. Bi-modal emotion recognition from expressive face and body gestures[J]. Journal of network and computer applications, 2007, 30(4): 1334–1345.

[58] KOELSTRA S, PATRAS I. Fusion of facial expressions and EEG for implicit affective tagging[J]. Image and vision computing, 2013, 31(2): 164–174.

[59] SOLEYMANI M, ASGHARIESFEDEN S, PANTIC M, et al. Continuous emotion detection using EEG signals and facial expressions[C]//Poceedings of the 2014 IEEE International Conference on Multimedia and Expo. Chengdu, China, 2014: 1–6.

[60] PONTI JR M P. Combining classifiers: from the creation of ensembles to the decision fusion[C]//Poceedings of the 2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials. Alagoas, Brazil, 2011: 1–10.

[61] FREUND Y, SCHAPIRE R E. Experiments with a new boosting algorithm[C]//Poceedings of the 1996 International Conference on Machine Learning. Bari, Italy, 1996: 148–156.

[62] CHANG Z, LIAO X, LIU Y, et al. Research of decision fusion for multi-source remote-sensing satellite information based on SVMs and DS evidence theory[C]//Poceedings of the Fourth International Workshop on Advanced Computational Intelligence. Wuhan, China, 2011: 416–420.

[63] NEFIAN A V, LIANG L, PI X, et al. Dynamic bayesian networks for audio-visual speech recognition[J]. EURASIP journal on advances in signal processing, 2002, 2002(11): 783042.

[64] MUROFUSHI T, SUGENO M. An interpretation of

fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure[J]. Fuzzy sets and systems, 1989, 29(2): 201–227.

[65] HUANG Y, YANG J, LIU S, et al. Combining facial expressions and electroencephalography to enhance emotion recognition[J]. Future internet, 2019, 11(5): 105.

[66] LU Y, ZHENG W, LI B, et al. Combining eye movements and EEG to enhance emotion recognition[C]//Poceedings of the Twenty-fourth International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina, 2015: 1170–1176.

[67] ZHANG S, ZHANG S, HUANG T, et al. Learning affective features with a hybrid deep model for audio-visual emotion recognition[J]. IEEE transactions on circuits & systems for video technology, 2017, 28(10): 1–1.

[68] METALLINOU A, WOLLMER M, KATSAMANIS A, et al. Context-sensitive learning for enhanced audiovisual emotion classification[J]. IEEE transactions on affective computing, 2012, 3(2): 184–198.

[69] MCGURK H, MACDONALD J. Hearing lips and seeing voices[J]. Nature, 1976, 264(5588): 746.

[70] NGUYEN D, NGUYEN K, SRIDHARAN S, et al. Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition[J]. Computer vision and image understanding, 2018, 174: 33–42.

[71] DOBRIŠEK S, GAJŠEK R, MIHELIČ F, et al. Towards efficient multi-modal emotion recognition[J]. International journal of advanced robotic systems, 2013, 10(1): 53.

[72] ZHANG S, ZHANG S, HUANG T, et al. Learning affective features with a hybrid deep model for audio–visual emotion recognition[J]. IEEE transactions on circuits and systems for video technology, 2017, 28(10): 3030–3043.

[73] TANG H, LIU W, ZHENG W, et al. Multimodal emotion recognition using deep neural networks[C]//Poceedings of the International Conference on Neural Information Processing. Guangzhou, China, 2017: 811–819.

[74] YIN Z, ZHAO M, WANG Y, et al. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model[J]. Computer methods and programs in biomedicine, 2017, 140: 93–110.

[75] SOLEYMANI M, PANTIC M, PUN T. Multimodal emotion recognition in response to videos[J]. IEEE transactions on affective computing, 2011, 3(2): 211–223.

[76] JAMES W. What is an Emotion?[J]. Mind, 1884, 9(34): 188–205.

[77] COWIE R, DOUGLASCOWIE E. Automatic statistical analysis of the signal and prosodic signs of emotion in speech[C]//Poceedings of the Fourth International Conference on Spoken Language Processing ICSLP'96. Philadelphia, USA, 1996: 1989–1992.

[78] SCHERER, KLAUS. Adding the affective dimension: a new look in speech analysis and synthesis[C]//Poceedings of the ICSLP. Dayton, USA, 1996.

[79] PANTIC M, ROTHKRANTZ L J. Automatic analysis of facial expressions: the state of the art[J]. IEEE transactions on pattern analysis & machine intelligence, 2000, 12: 1424–1445.

[80] IOANNOU S V, RAOUZAIOU A T, TZOUVARAS V A, et al. Emotion recognition through facial expression analysis based on a neurofuzzy network[J]. Neural networks, 2005, 18(4): 423–435.

[81] CASTELLANO G, VILLALBA S D, CAMURRI A. Recognising human emotions from body movement and gesture dynamics[C]//Poceedings of the International Conference on Affective Computing and Intelligent Interaction. Lisbon, Portugal, 2007: 71–82.

[82] CAMURRI A, LAGERLöF I, VOLPE G. Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques[J]. International journal of human-computer studies, 2003, 59(1-2): 213–225.

[83] KALIOUBY, ROBINSON P. Generalization of a vision-based computational model of mind-reading[C]//Poceedings of the International Conference on Affective Computing and Intelligent Interaction. Beijing, China, 2005: 582–589.

[84] CASTELLANO G, KESSOUS L, CARIDAKIS G. Emotion recognition through multiple modalities: face, body gesture, speech[M]. Berlin: Springer-verlag. 2008: 92–103.

[85] SCHERER K R, ELLGRING H. Multimodal expression of emotion: affect programs or componential appraisal patterns?[J]. Emotion, 2007, 7(1): 158–171.

[86] PETRANTONAKIS P C, HADJILEONTIADIS L J. A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition[J]. IEEE transactions on information technology in biomedicine, 2011, 15(5): 737–746.

[87] LIN Y, WANG C, JUNG T, et al. EEG-based emotion recognition in music listening[J]. IEEE transactions on biomedical engineering, 2010, 57(7): 1798–1806.

[88] DAVIDSON R J, FOX N A. Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants[J]. Science, 1982, 218(4578): 1235–1237.

[89] TURETSKY B I, KOHLER C G, INDERSMITTEN T, et al. Facial emotion recognition in schizophrenia: when and why does it go awry?[J]. Schizophrenia research, 2007, 94(1-3): 253–263.

[90] HAJCAK G, MACNAMARA A, OLVET D M. Event-related potentials, emotion, and emotion regulation: an integrative review[J]. Developmental neuropsychology, 2010, 35(2): 129–155.

[91] ALARCAO S M, FONSECA M J. Emotions recognition using EEG signals: A survey[J]. IEEE transactions on affective computing, 2019, 10(3): 374–393.

[92] CHANEL G, KIERKELS J J M, SOLEYMANI M, et al. Short-term emotion assessment in a recall paradigm[J]. International journal of human-computer studies, 2009, 67(8): 607–627.

[93] CHEN S, ZHEN G, WANG S. Emotion recognition from peripheral physiological signals enhanced by EEG[C]//Poceedings of the IEEE International Conference on Acoustics. Shanghai, China, 2016.

[94] EKMAN P. An argument for basic emotions[J]. Cognition & emotion, 1992, 6(3-4): 169–200.

[95] HUANG Y, YANG J, LIAO P, et al. Fusion of facial expressions and EEG for multimodal emotion recognition[J]. Computational intelligence and neuroscience, 2017: 2107451.

[96] KAPOOR A, BURLESON W, PICARD R W. Automatic prediction of frustration[J]. International journal of human-computer studies, 2007, 65(8): 724–736.

[97] LIU W, ZHENG W, LU B. Emotion recognition using multimodal deep learning[M]. Berlin: Springer International Publishing, 2016: 521–529.

[98] LüHMANN A V, WABNITZ H, SANDER T, et al. M3BA: A mobile, modular, multimodal biosignal acquisition architecture for miniaturized EEG-NIRS based hybrid BCI and monitoring[J]. IEEE transactions on biomedical engineering, 2016, 64(6): 1199–1210.

[99] AREVALILLO-HERRáEZ M, COBOS M, ROGER S, et al. Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from EEG signals[J]. Sensors, 2019, 19(13): 2999.

[100] 郭琛, 高小榕. 用于眼动检测和脑电采集的数据同步方法 [C]// 第九届全国信息获取与处理学术会议论文集Ⅱ. 丹东, 中国. 2011.
GUO Chen, GAO Xiaorong. A data synchronization method for eye movement detection and eeg acquisition[C]// Proceedings of the 9th National Conference on Information Acquisition and Processing Ⅱ. Dandong, China. 2011.

[101] 赵亮. 多模态数据融合算法研究 [D]. 大连: 大连理工大学, 2018.
ZHAO Liang. Multi-modal data fusion algorithm research[D]. Dalian: Dalian University of Technology, 2018.

[102] ZHENG W, ZHU J, PENG Y, et al. EEG-based emotion classification using deep belief networks[C]//Poceedings of the 2014 IEEE International Conference on Multimedia and Expo (ICME). Chengdu, China, 2014: 1–6.

[103] MOWER E, MATARIC M J, NARAYANAN S. A framework for automatic human emotion classification using emotion profiles[J]. IEEE transactions on audio, speech, and language processing, 2010, 19(5): 1057–1070.

作者简介：

潘家辉，副教授，博士，广东医学会数字医学分会常务委员，主要研究方向为机器学习、脑机接口、模式识别与智能系统。广州市珠江科技新星，华南师范大学教学名师，曾两次获得广东省科学技术奖一等奖、中华医学科技奖三等奖等。主持国家自然科学基金项目 2 项、广东省自然科学基金项目 2 项、广州市重点研发领域项目 1 项、广州市科技创新人才项目 1 项。发表学术论文 80 余篇。

何志鹏，硕士研究生，主要研究方向为情感计算、混合脑机接口。

李自娜，硕士研究生，主要研究方向为机器学习、情感识别。