



中国图形学报

2021
12
VOL.26

ISSN1006-8961
CN11-3758/TB





第26卷第12期（总第308期）
2021年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

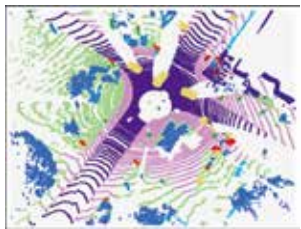
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



环视鱼眼图像处理深度学习
研究进展(第2778页)



结合空间深度卷积和残差
的大尺度点云场景分割
(第2848页)



曼哈顿世界环境下消失点非线
性优化估计算法(第2931页)

综述

深度对抗视觉生成综述

谭明奎, 许守恺, 张书海, 陈奇 2751

人机交互中的智能体情感计算研究

刘婷婷, 刘箴, 柴艳杰, 王瑾, 王媛怡 2767

环视鱼眼图像处理深度学习研究进展

包俊, 刘宏哲, 褚文博 2778

图像处理和编码

图像分解与色彩先验下的多曝光图像融合

李嫛源, 王琴, 朱智勤, 齐观秋 2800

双尺度分解和显著性分析相结合的红外与可见光图像融合

霍星, 邹韵, 陈影, 檀结庆 2813

轻量级注意力特征选择循环网络的超分辨率重建

徐雯捷, 宋慧慧, 袁晓彤, 刘青山 2826

区域级通道注意力融合高频损失的图像超分辨率重建

周波, 李成华, 陈伟 2836

图像分析和识别

结合空间深度卷积和残差的大尺度点云场景分割

刘盛, 黄圣跃, 程豪豪, 沈家瑜, 陈胜勇 2848

细粒度图像分类的互补注意力方法

赵勋, 王家宝, 李阳, 王亚鹏, 苗壮 2860

联合分块谱带优选和深度特征的高光谱人脸识别

谢志华, 李毅, 牛杰一 2870

时空双仿射微分不变量及骨架动作识别

李琪, 墨瀚林, 赵婧涵, 郝宏翔, 李华 2879

多尺度分割先验迭代重加权低秩恢复显著性检测

张荣国, 郑维佳, 赵建, 胡静, 刘小君 2892

融合环境特征与改进YOLOv4的安全帽佩戴检测

葛青青, 张智杰, 袁琰, 李秀梅, 孙军梅 2904

图像理解和计算机视觉

关节轴角先验对3维人体重建结果的影响

姚砾, 张幼安, 张梦雪, 万燕 2918

曼哈顿世界环境下消失点非线性优化估计算法

韩冰心, 卢惠民, 于清华, 张礼廉 2931

双目机器视觉及RetinaNet模型的路侧行人感知定位

连丽容, 罗文婷, 秦勇, 李林 2941

遥感图像处理

遥感图像中油罐目标精确定位与参数提取

江哈, 张月婷, 郭嘉逸, 赵鑫, 李芳芳, 黄丽佳, 胡玉新, 雷斌, 丁赤飏 2953

参数合成空间变换网络的遥感图像一致性配准

陈颖, 张祺, 李文举, 石艳娇, 陈磊 2964

改进Mask R-CNN模型的海洋锋检测

徐慧芳, 黄冬梅, 贺琪, 杜艳玲, 覃学标 2981

噪声鲁棒的轻量级深度遥感场景图像分类检索

王亚鹏, 李阳, 王家宝, 赵勋, 苗壮 2991

改进U-Net型网络的遥感图像道路提取

杨佳林, 郭学俊, 陈泽华 3005

弱监督深度语义分割网络的多源遥感影像水体检测

李鑫伟, 李彦胜, 张永军 3015

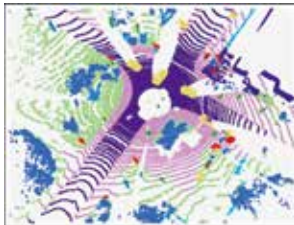
总目次 I

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Research progress of fisheye image processing based on deep learning(P2778)



A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation(P2848)



Vanishing point estimation based on non-linear optimization in manhattan world environments (P2931)

Review

A review on deep adversarial visual generation	
Tan Mingkui, Xu Shoukai, Zhang Shuhai, Chen Qi	2751
Agent affective computing in human-computer interaction	
Liu Tingting, Liu Zhen, Chai Yanjie, Wang Jin, Wang Yuanyi	2767
Research progress of fisheye image processing based on deep learning	
Bao Jun, Liu Hongzhe, Chu Wenbo	2778

Image Processing and Coding

Multi-exposure image fusion based on image decomposition and color prior	
Li Yuanyuan, Wang Qin, Zhu Zhiqin, Qi Guanqiu	2800
Dual-scale decomposition and saliency analysis based infrared and visible image fusion	
Huo Xing, Zou Yun, Chen Ying, Tan Jieqing	2813
Lightweight attention feature selection recursive network for super-resolution	
Xu Wenjie, Song Huihui, Yuan Xiaotong, Liu Qingshan	2826
Region-level channel attention for single image super-resolution combining high frequency loss	
Zhou Bo, Li Chenghua, Chen Wei	2836

Image Analysis and Recognition

A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation	
Liu Sheng, Huang Shengyue, Cheng Haohao, Shen Jiayu, Chen Shengyong	2848
Complemented attention method for fine-grained image classification	
Zhao Xun, Wang Jiabao, Li Yang, Wang Yapeng, Miao Zhuang	2860
Hyperspectral face recognition based on partition bands optimal selection and deep features	
Xie Zhihua, Li Yi, Niu Jieyi	2870
Spatio-temporal dual affine differential invariants for skeleton-based action recognition	
Li Qi, Mo Hanlin, Zhao Jinghan, Hao Hongxiang, Li Hua	2879
Iterative reweighted low-rank recovery salient object detection with multiscale segmentation prior	
Zhang Rongguo, Zheng Weijia, Zhao Jian, Hu Jing, Liu Xiaojun	2892
Safety helmet wearing detection method of fusing environmental features and improved YOLOv4	
Ge Qingqing, Zhang Zhijie, Yuan Long, Li Xiumei, Sun Junmei	2904

Image Understanding and Computer Vision

The impact of joint axis angle prior on the results of 3D human body reconstruction	
Yao Li, Zhang Youan, Zhang Mengxue, Wan Yan	2918
Vanishing point estimation based on non-linear optimization in manhattan world environments	
Han Bingxin, Lu Huimin, Yu Qinghua, Zhang Lilian	2931
Roadside pedestrian detection and location based on binocular machine vision and RetinaNet	
Lian Lirong, Luo Wenting, Qin Yong, Li Lin	2941

Remote Sensing Image Processing

Accurate localization and parameter extraction of oil tank in remote sensing images	
Jiang Han, Zhang Yueting, Guo Jiayi, Zhao Xin, Li Fangfang, Huang Lijia, Hu Yuxin, Lei Bin, Ding Chibiao	2953
Consistent registration of remote sensing images in parametric synthesized spatial transformation network	
Chen Ying, Zhang Qi, Li Wenju, Shi Yanjiao, Chen Lei	2964
Ocean front detection method based on improved Mask R-CNN	
Xu Huifang, Huang Dongmei, He Qi, Du Yanling, Qin Xuebiao	2981
A robust lightweight deep neural network for remote sensing scene image classification and retrieval under label noise	
Wang Yapeng, Li Yang, Wang Jiabao, Zhao Xun, Miao Zhuang	2991
Road extraction method from remote sensing images based on improved U-Net network	
Yang Jialin, Guo Xuejun, Chen Zehua	3005
Weakly supervised deep semantic segmentation network for water body extraction based on multi-source remote sensing imagery	
Li Xinwei, Li Yansheng, Zhang Yongjun	3015

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2021)12-2767-11

论文引用格式: Liu T T, Liu Z, Chai Y J, Wang J and Wang Y Y. 2021. Agent affective computing in human-computer interaction. Journal of Image and Graphics, 26(12): 2767-2777 (刘婷婷, 刘箴, 柴艳杰, 王瑾, 王媛怡. 2021. 人机交互中的智能体情感计算研究. 中国图象图形学报, 26(12): 2767-2777) [DOI:10.11834/jig.200498]

人机交互中的智能体情感计算研究

刘婷婷¹, 刘箴^{2*}, 柴艳杰², 王瑾¹, 王媛怡¹

1. 宁波大学科学技术学院, 慈溪 315300; 2. 宁波大学信息科学与工程学院, 宁波 315211

摘要: 机器的情感是通过融入具有情感能力的智能体实现的, 虽然目前在人机交互领域已经有大量研究成果, 但有关智能体情感计算方面的研究尚处起步阶段, 深入开展这项研究对推动人机交互领域的发展具有重要的科学和应用价值。本文通过检索 Scopus 数据库选择有代表性的文献, 重点关注情感在智能体和用户之间的双向流动, 分别从智能体对用户的情绪感知和对用户情绪调节的角度开展分析总结。首先梳理了用户情绪的识别方法, 即通过用户的表情、语音、姿态、生理信号和文本信息等多通道信息分析用户的情绪状态, 归纳了情绪识别中的一些机器学习方法。其次从用户体验角度分析具有情绪表现力的智能体对用户的影响, 总结了智能体的情绪生成和表现技术, 指出智能体除了通过表情之外, 还可以通过注视、姿态、头部运动和手势等非言语动作来表现情绪。并且梳理了典型的智能体情绪架构, 举例说明了强化学习在智能体情绪设计中的作用。同时为了验证模型的准确性, 比较了已有的情感评估手段和评价指标。最后指出智能体情感计算急需解决的问题。通过对现有研究的总结, 智能体情感计算研究是一个很有前景的研究方向, 希望本文能够为深入开展相关研究提供借鉴。

关键词: 人机交互; 智能体; 情感计算; 情绪; 感知; 表现

Agent affective computing in human-computer interaction

Liu Tingting¹, Liu Zhen^{2*}, Chai Yanjie², Wang Jin¹, Wang Yuanyi¹

1. College of Science and Technology, Ningbo University, Cixi 315300, China;

2. Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: Human computer interaction technology has been promoting to realize intelligent human-computer interaction. The user's emotional experience in the human-computer interaction system has been facilitated based on the realization of emotional interaction. Emotional interaction has been intended to use widely via Gartner's analysis next decade. The agent can be real or virtual to detect the user's emotion and adjust the user's emotion. It can greatly enhance the user's experience in human-computer interaction on the aspects of psychological rehabilitation, E-education, digital entertainment, smart home, virtual tourism, E-commerce and etc. The research of agent's affective computing has involved in computer graphics, virtual reality, human-computer-based interaction, machine learning, psychology, social science. Based on Scopus database, 2 080 journal papers have been optioned via using virtual human (agent, multimodal) plus emotional interaction as the key words each. The perspective of agent's perceptions and influence of users' emotions have been analyzed and summarized. The importance of multi-channel in emotion perception and the typical machine learning algorithms in emotion recognition have been summarized from the perspective of agent's perception of users' emotions. The external and

收稿日期: 2020-08-25; 修回日期: 2021-02-03; 预印本日期: 2021-02-10

* 通信作者: 刘箴 liuzhen@nbu.edu.cn

基金项目: 国家自然科学基金项目(61761166005); 宁波市科技计划项目(2020Z082)

Supported by: National Natural Science Foundation of China (61761166005); Ningbo Science and Technology Plan Project (2020Z082)

internal factors affecting users' emotions have been analyzed from the perspective of agent's influence on users' emotions. The emotional architecture, emotional generation and expression algorithms have been implemented. Customized evaluation methods have been applied to improve the accuracy of the affective computing algorithm. The importance of emotional agent in human-computer interaction has been analyzed. Four key steps of agent affective computing have been summarized based on current studies: 1) An agent expressed its emotion to the user. 2) The user gave their feedback to the agent (they may or may not express their satisfaction or dissatisfaction via some channels like facial expressions). 3) The real-time agent perceived the user's emotional state and intention and adjustable emotional performance to respond to user's feedback. 4) A standard (e.g., the completion of emotion regulation task, the end of plot) has been reached, the agent stopped interacting with the user, otherwise, returns to step 1). The current studies have shown that user's expressed emotions via facial expressions, voices, postures, physiological signals and texts on the aspect of user emotion recognition. The multi-channel method has been more reliable than the single channel method. Machine learning can be used to extract emotional features. Typical machine learning algorithms and their applicable scenarios have been sorted out based on CNN (convolutional neural network) nowadays. Some solutions have been facilitated to resolve insufficient data and over fitting issues. Spatial distance, the number of agents, the appearance of the agent, brightness and shadow have been set as external factors. Agent's autonomous emotion expression has been targeted as the internal factor. An agent should have an emotional architecture and use facial expression, eye gaze, posture, head movement gesture and other channels to express its emotion. The accuracy of the emotional classification model and users' feelings has been assessed based on an affective computing model. The statistical sampling analysis has been listed in the table. The existing emotional agents such as low intelligence, lack of common-sense database, lack of interactivity have been as the constrained factors. The research of agent affective computing in the field of human-computer interaction has been developed further. An affective computing for human-computer-based interaction and an agent could be a channel of emotional interaction. Knowledge-based database and appropriated machine learning algorithms have been adopted to build an agent with the ability of emotion perception and emotion regulation. Qualified physiological detection equipment and sustainable enrichment of emotional information assessment methods have developed the affective computing further.

Key words: human-computer interaction; agent; affective computing; emotion; perception; performance

0 引言

随着人工智能技术的发展,人机交互广泛渗透到社会生活中(Sharp 等,2003)。人机交互涉及很多学科,不仅包含计算机技术,也包含心理学、艺术设计等人文学科。目前,人机交互发展的方向是与人工智能技术相融合,目标是建立自然和谐的人机交互环境,更好地满足人类的交互需求,实现人与交互技术的协同共进(范向民 等,2019;范俊君 等,2018;陶建华 等,2019)。随着体验经济的兴起,人机交互中的用户情绪体验越来越受重视。利用机器部分代替人实现情感交流成为迫切的社会需求。为了使机器具有情感交互能力,需要创建具有情绪表现力的智能体,这也是人工智能与人机交互相融合的重要交叉点(Sharp 等,2003)。

智能体是人工智能领域的一个核心概念,智能体有多种表现形式,可以是虚拟的也可以是实体的

(Pferifer 和 Bongard,2009)。智能体的结构越来越丰富,不仅具有感知能力,也具有情绪和行为表现能力。

智能体情绪感知和情绪表达的研究属于情感计算的研究范围。本文主要围绕智能体情感计算的设计需求,从交互的角度总结了情绪在智能体设计中的作用及相关实现方法,梳理了目前智能体情感计算方面的工作,为进一步开展情感交互的研究提供参考。

1 人机交互中的情感交流过程

在人与机器的情感交互中,智能体(可以是虚拟人也可以是机器人等)在其中扮演了重要角色。对于具体的一次交互任务来说,情感体现为情绪。智能体最重要的特征是具有自主感知和情绪表现力。它能够感知用户的情绪和状态,通过内部的学习调整,做出合适的情绪反应,建立与用户的情感

交流,甚至影响用户的情绪。任何一种机器系统,若具有情感交互能力,就是指其包含的智能体具有情绪感知和情绪表现能力。

一般来说,用户和智能体的情感交流是双向的,具体过程如下:1)情感从用户流向智能体。用户在人机交互中,表达自己的情感,通过智能体的感知模块,情感从用户流向智能体。在感知到用户情绪之后,智能体可以通过学习模块,调整情绪的表达,以满足用户个性化情感交流的需求。换言之,情绪从用户流向智能体是实现用户个性化情感交流的前提。2)情感从智能体流向用户。当情感从智能体流向用户时,智能体会对用户的情感产生影响。Roxas 等人(2018)研究发现,当智能体做出愤怒、悲伤和快乐等舞蹈动作后,用户情绪会受到相应影响。了解智能体如何对用户产生影响,可以合理构建智能体的情绪模型。在算法的辅助下,情绪可以定量表现。这有助于建立宜人型的人机交互界面,更好地满足用户的情感体验。

智能体 and 用户情感交互的算法步骤可以概括如下:

- 1) 用户和机器交互,智能体表现出一定情绪。
- 2) 用户对智能体的情绪做出一定反馈(可以有反馈也可以没有反馈;反馈可能满意也可能不满意;反馈通道可以是一个也可以是多个)。
- 3) 智能体实时感知用户的情绪状态和意图,调整情绪表现,以响应用户反馈。
- 4) 达到一定标准(如完成情绪调节任务、剧情结束),智能体停止与用户交互。否则返回步骤1)。

2 用户情绪感知

要实现情感从用户流向智能体,最重要的是智能体能够对用户表现出来的情绪进行准确的感知,即能够识别用户的情绪类型和强度。情绪的表现形式多样,目前对于单个模态的情绪检测及分类已有许多研究。Quan 等人(2017)使用组合倒谱距离方法对语音中的情感进行识别。饶元等人(2018)对基于语义分析的情感计算技术进行了综述。但是,仅凭一个通道获取的情绪信息并不完整,可能存在缺失或误判的情况。即使是作为主要通道的面部表情,也会存在这样的问题。Samara 等人(2019)使用一种分层的机器学习方法(hierarchical parallelised

binary support vector machines, HPBSVM) 对用户 in 主动交互(更改系统设置、网购、操作 Excel、玩 Pac-man)和被动交互(观看视频)过程中的情绪进行分析,发现虽然与被动交互相比,在主动交互任务期间,用户的面部表情发生的变化更多,但面部表情仍不能准确揭示用户与计算机交互时的实际感受,最重要的是用户不会以与人类交互时相同的方式来与机器交换情绪。因此,通常需要结合多个通道信息来感知用户情绪。

2.1 多通道情绪感知

情绪的通道有很多种,人脸表情、语音信号、肢体行为、生理信号和文本信息是目前研究较多的几种通道(毛峡和薛雨丽, 2011; Jiang 等, 2020)。D' Mello和 Graesser(2010)针对对话、身体语言和面部表情3个通道进行了详细分析。对一个32 min的带对话智能体的辅导课程,让参与者进行了两种形式的情感判断。即自然情感判断(spontaneous affect judgments)(在课程的任何时候停顿)和固定时间判断(fixed judgments)(课程每20 s停顿)。结果显示,在自然情感判断中,面部表情为主要判断依据,在固定时间判断中,上下文对话更为有利。对于固定时间判断,三通道的准确率略高于单通道,并能减少对话通道中不同情绪的精度差异。但是面部表情和对话通道不仅能较好地地区分不同情绪,且在统计学上与三通道相当,可用这两个通道作为三通道模型的替代。同时,在自然情感判断中也发现结合身体语言的双通道不具备竞争力。因此,相对于其他两个通道来说,身体语言通道不是那么重要,在双通道表示中可以舍弃。在固定时间判断中对某些情感有超加效应(superadditive effects),而在自然情感判断中没有发现此类效应。同样是对语音、姿势和面部特征三个通道作分析,Kessous 等人(2010)却对双通道的组合有不同见解。他们发现,姿势和语音的组合效果最好,比表情和语音的组合识别率高10%。该差异的产生可能是两种方法提取姿势和语音时所用特征不同所致。D' Mello 和 Graesser(2010)对身体语言采集的数据比较粗糙,仅使用 tekscan 座椅压力分布量测系统检测测试者在椅背和坐凳上的压力。而 Kessous 等人(2010)使用身体的运动量和收缩指数、手掌质心的速度、加速度和流动性五个指数来检测姿势和手势特征。此外,在 Kessous 等人(2010)的工作中,语音不包含上下文

信息,仅对强度、音调、梅尔频率倒谱系数(Mel frequency cepstrum coefficient, MFCC)、bark 谱带、浊音段特征和暂停长度进行分析。因此,身体和手势姿势能比面部表情更好地补充语音通道中缺失的信息,从而分辨出更多的情绪信息。与以上常规通道不同,任福继等人(2018)使用了血容量脉冲(blood volume pulse, BVP)通道作为面部表情通道的补充通道,相比于其他生理信号, BVP 可以直接由摄像头获取,无需接触人体,比较便捷,而且其中含有心率值和呼吸频率值,能反映人的紧张、害怕和放松的状态。

总体而言,多通道通常能比单通道提供更多的信息,更容易对情绪进行准确判断。但是,目前的多通道通常是表情、语音和肢体的互相结合,很少有研究将生理信号与以上几种通道相结合。此外,多模态情绪信息的融合和数据处理也是一个需要进一步研究的课题。

2.2 情绪感知中的机器学习算法

无论使用二通道还是三通道,特征提取都是非常重要的工作。由于机器学习算法在特征提取和分类方面具有明显优势,这两部分工作多采用机器学习算法来实现。Kanjo 等人(2019)使用卷积神经网络和长短期记忆递归神经网络(convolutional neural network and long short-term memory, CNN-LSTM)的混合方法对采自智能手机和可穿戴设备的人体、环境和位置信息进行自动的特征提取。Ma 等人(2019)针对视、听两个通道,使用 2D CNN 作为音频情感特征提取器,3D CNN 作为图像中的面部表情提取器来自动提取两个通道中的情绪特征。之后,使用深度信念网络(deep belief networks, DBN)用于提取到的多模态情绪特征的高度非线性融合,再用支持向量机(support vector machine, SVM)进行情绪分类。陈师哲等人(2018)使用支持向量机和随机森林(random forest)作为情感分类器,并通过网格搜索(grid search)选择验证集上性能最好的超参数作为最终模型。同时,为了实现多文化场景下的情感识别,设置了情感相关隐层特征和文化相关隐层特征。Filko 和 Martinović (2013)使用主成分分析和 15 个神经网络分析关键面部区进行情绪识别。Zhang 等人(2015)为了解决 2D 模型鲁棒性不足的问题,实现运动和光照变化情况下的情绪识别,提出一个 3D 面部动作单元(action unit, AU)强度估计和

情绪识别系统,使用前馈神经网络和支持向量回归估算 16 个 AU 的强度,并提出一套 6 个自适应的分类器对 6 种基本情绪进行分类,每个分类器都使用基于距离的聚类 and 能进行不确定度量的互补神经网络(complementary neural network, CMTNN)。但很多已有的表情识别研究针对的是实验室环境下的单一表情,没有关注真实情况下的混合表情。为此,邓伟洪课题组建立了带有复合表情标签的数据集 RAF-DB(real-world affective face database)(单标签)和 RAF-ML(multi-label facial expression database)(多标签),并提出了适应这两个数据库的 DLP-CNN(deep locality-preserving convolutional neural network)算法和 Deep Bi-Manifold CNN 算法以更好地提取特征(Li 和 Deng, 2019a, b)。

除了直接对情绪进行分类之外,也有一些研究通过对动作进行分类识别而间接地对情绪进行分类(Ajili 等, 2019)。不同于情绪识别,韩加旭等人(2020)使用融合排序与回归的卷积神经网络(joint rank and regression convolution neural network, JRR-CNN)对表情的强度进行了估计。

早期对情绪分类使用 SVM 较多,随着数据量增大,以 CNN 为代表的深度学习算法开始逐渐增多。Ko(2018)关于面部情绪识别综述中有大量 CNN 或 CNN 与 LSTM(RNN)相结合的工作。Shu 等人(2018)在对生理信号的情绪识别研究工作中列举了不少关于 CNN、DBN、PNN(product-based neural network)和 LSTM 的工作。但是深度学习虽然不依赖于模型,却对数据及其标注的要求比较高,不适用于小数据集分类。而且在一些非受控的外部条件(如光照、头部姿态)下会出现过拟合现象。为了解决这些问题,出现了很多预训练、微调技术以及多样化的网络输入设置(Li 和 Deng, 2020)。情绪分类中常用的机器学习算法如表 1 所示。

3 智能体对用户的情绪影响

研究智能体如何对用户情绪产生影响有助于更好地设计满足用户体验需求的智能体。关于智能体对用户情绪影响的研究主要从两方面展开,一是影响智能体情绪传递的外部条件,二是智能体本身的情绪表现。

表1 情绪分类中常用的机器学习算法

Table 1 Classical machine learning algorithms in emotion classification

算法	适用场景
前馈神经网络(Zhang等,2015)	诊断面部动作单元(AU)的强度
CMTNN(Zhang等,2015)	适合于二元分类问题,而且能够提供模糊的分类结果估计
聚类(Zhang等,2015)	情绪分类
support vector regressor(Zhang等,2015)	诊断面部动作单元(AU)的强度
DAG-SVM(Quan等,2017)	情绪分类
随机森林(陈师哲等,2018)	情绪分类
SVM(Kanjo等,2019;陈师哲等,2018)	情绪分类
HPBSVM(Samara等,2019)	情绪分类
CNN-LSTM(Kanjo等,2019)	采自智能手机和可穿戴设备的人体、环境和位置信息
2D CNN(Ma等,2019)	音频情感特征提取
3D CNN(Ma等,2019)	图像中的面部表情特征提取
深度信念网络(Ma等,2019)	特征融合
DLP-CNN(Li和Deng,2019a)	复合表情识别
Deep Bi-Manifold CNN(Li和Deng,2019b)	复合表情识别

3.1 影响情绪传递的外部条件

自主情绪表达是智能体向用户传递情绪的一个主要途径。除此之外,外部条件设置也会对用户的情绪产生影响。在已有的研究中,可查到的外部影响因素主要有空间距离、智能体个数、智能体外观真实感和光照。

1)空间距离和智能体个数的研究。Bönsch等人(2018)从个人空间的角度出发,对智能体情绪可能造成的用户舒适度进行研究。结果显示,智能体情绪和智能体个数都会对用户舒适度产生影响,与高兴情绪相比,愤怒智能体需要的舒适空间更大;与单个智能体相比,群组智能体需要的舒适空间更大。比如同样是高兴的情绪表现,群组智能体可能会给用户带来侵略感。

2)智能体真实感对用户情绪的研究。Volonte等人(2016)使用具有真实、卡通和素描外观的智能体(见图1)进行实验,以发现智能体外观的真实感对用户情绪的影响。通过对参与者的情绪状态在虚拟病人病情恶化(一种情绪紧张的互动)不同时间步骤中的分析,发现真实外观对负面情绪的影响远低于其他外观,认为这是由恐怖谷效应引起的,即过于真实的外表可能会抑制参与者在关键情绪上的反

应。但是真实的外观有助于激发适当的社会情感结构,如害羞和羞耻感。此外,性别会影响用户的情绪,男性具有比女性更高的皮电EDA(electrodermal activity)觉醒水平。这与智能体外观的真实感无关。

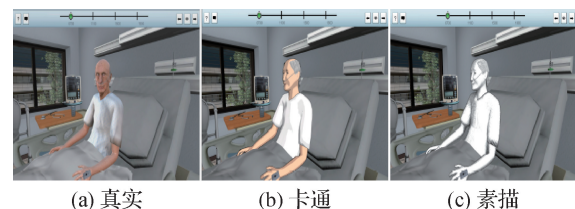


图1 虚拟患者的外观

Fig. 1 The appearances of the virtual patient
((a) real; (b) cartoon; (c) sketch)

Zibrek等人(2019)的实验与Volonte等人(2016)的类似,并有类似的发现。即在中性场景中,具有真实外观的智能体能使用户体会到更多的消极情绪,而在消极场景中,具有真实外观的智能体对用户情绪的影响反而不如风格化的智能体。其后,Volonte等人(2019)使用同样的一组智能体检测智能体外表及其有效的行为对用户情绪和注意力的影响程度。检测结果显示,在互动过程中,具有素描外观的智能体得到了用户更多的关注。此外,视觉

刺激优先于情绪,但他们之间存在互动关系。在病人病情恶化的第1和2阶段,智能体的负面情绪较低,用户的视觉注意力会影响用户情绪;在病情恶化的第4阶段,用户的情绪不受其视觉注意力的影响。以上研究的共同之处在于,在智能体负面情绪较高时,真实的外观对用户情绪的影响没有其他外观大,说明了不是在所有情况下智能体都需要具有很高的真实感。因此。要针对不同场景和目的进行智能体外观设计。

3) 光照对用户情绪的影响。Wisessing 等人(2020)针对光照对用户情绪的影响进行了一系列详细实验,使用主光(key light)、辅光(fill light)和轮廓光(rim light)三点照明,讨论亮度和主辅光比率(key-to-fill ratio, KTFR)对情绪(愤怒、悲伤、害怕、高兴)识别率、强度和吸引力的影响。为了去除外观影响,分别对比了具有不同真实感(真实、中性、卡通)的男女智能体,如图2所示,其中第1—4行分别展示了100%亮度下,1:1 KTFR的悲伤男性、16:1 KTFR的悲伤男性、1:1 KTFR的高兴女性和16:1 KTFR的高兴女性。

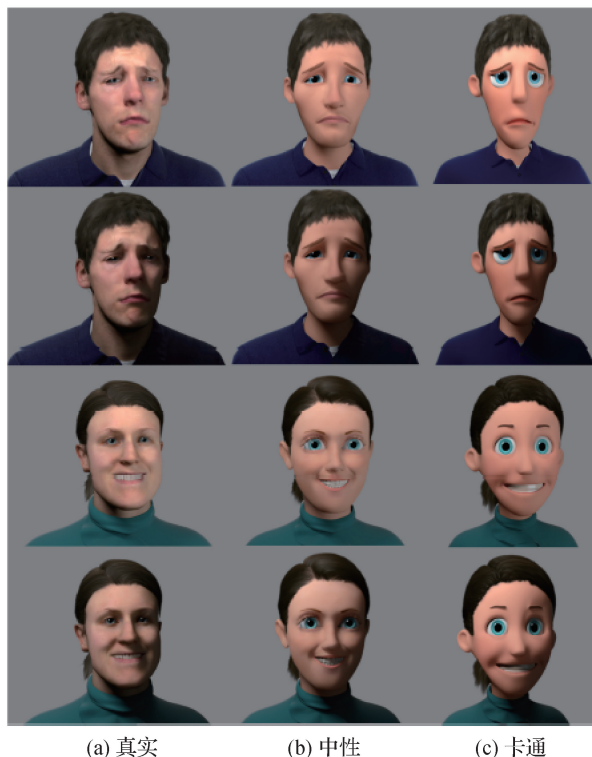


图2 具有不同真实感的智能体

Fig. 2 Agent with different realism
((a) realistic; (b) middle; (c) cartoon)

结果发现,无论智能体外观如何,亮度都能增强高兴和悲伤的吸引力,但不能影响愤怒和恐惧情绪。对于卡通形象来说,情绪表现不受 KTFR 的影响,而对于真实形象来说,快乐和愤怒情绪对用户的吸引力会受 KTFR 的影响。需要注意的是,暗度虽然可能增加悲伤的表达,但也会降低识别率,所以在设计时应避免将光线调得过暗。

3.2 智能体的情绪生成和表现

智能体的自主情绪表达是情感计算的一个主要研究方向,主要分为宏观的情绪框架设计和微观的情绪表现两部分。

3.2.1 情绪框架

在情绪生成过程中,可以从宏观层面对整个过程进行描述,也可以从微观层面对某一具体部分进行算法设计实现。情绪框架是一种宏观描述手段,包含了智能体情绪生成过程中可能涉及的因素和相互之间的作用。情绪框架设计的合理性直接决定着智能体的情绪生成能力。

EMA(emotion and adaptation)是 Gratch 和 Marsella(2004)提出的著名的情绪框架,允许智能体综合评价外部环境与自身情况,并累计叠加评价结果,生成情绪,进而产生相对应的控制信号以推翻或维持导致个体评估的因果解释特征,能在交互式环境中对人类情绪行为进行抽象并建模,且易于扩展,是使用较多的一个基础的情绪框架。

WASABI(Wasabi affect simulation for agents with believable interactivity)情绪架构(Becker-Asano 和 Wachsmuth,2010)是另一个理论架构,主要贡献是将本能产生的原生情绪和需要较高认知推理能力的次生情绪统一起来进行表现,将结合过程放在 PAD (pleasure-arousal-dominance)空间上完成,选取希望(hope)、恐惧成真(fears-confirmed)和宽慰(relief)3种次生情绪进行演示。在感受到外界刺激后,智能体会对刺激进行无意识评估和有意识评估。无意识部分在投射到 PAD 空间后,以一定强度引发原生情绪。而有意识评估会产生次生情绪。一部分认知合理的次生情绪会像原生情绪一样,直接引发情绪冲动,进而采用原生情绪的生成机制,进入到原生情绪的生成;而另一部分次生情绪则会在 PAD 空间中进行过滤,与原生情绪一起形成智能体能感知到的情绪。作为第1个将原生情绪和次生情绪进行过程描述的结构,WASABI 的工作非常有开创性。目

前虽然有一些研究使用相应的参数表示认知对情绪的影响,但完整使用该结构的工作并不多见。

现有的情绪理论不止一个,涉及用于描述情绪的成分也较多。为了找出一种通用模型,Rosales 等人(2019)对已有的工作进行综合研究,将这些工作中提到的所有参数归纳为情绪生成中必须包含的3个基本要素:情感、感觉和情绪行为。并提出了一个包含3个基本要素的通用情绪框架。该框架分为3个阶段:阶段1)情感对刺激进行评估,产生感觉及情绪行为。这一阶段的情绪行为与基本情绪(即反应性情绪)有关;阶段2)这一阶段有认知过程的参与,可产生次生情绪;阶段3)此阶段是认知过程的结果,能控制情感和情绪行为。其中情感控制将重新解释所感知的刺激的含义,情绪行为控制则会抑制相关的行为,试图不表达感觉。

除特征提取和分类问题外,机器学习算法在决策类问题上也具有优势。若情绪框架涉及选择决策,可以考虑使用相关机器学习算法。Moussa 和 Magnenat-Thalmann(2013)在情绪框架中使用Q学习(Q-learning)确定智能体将要执行的动作,如图3所示。



图3 具有情绪学习能力的智能体

Fig.3 The agent with emotional learning ability

该模型设置依恋管理模块,根据用户之前引起的情绪计算智能体对该用户应该具有的情感依恋。学习模块使用Q学习方法,通过执行动作后的奖励,不断学习在交互的每个访问状态下动作的Q值,计算动作效用。奖励值 R_s^{user} 由交互状态 s 时具有的正面情绪 PE_i^{user} 和负面情绪 NE_i^{user} 决定,具体为

$$R_s^{\text{user}} = \frac{\max_i PE_i^{\text{user}} - \max_i NE_i^{\text{user}}}{2} + 0.5 \quad (1)$$

对式(1)进行归一化,得到1~0之间的奖励。通过奖励,可以计算Q值,具体为

$$Q_{s,a}^{\text{user}} = (1 - \alpha)Q_{s,a}^{\text{user}} + \alpha(R_s^{\text{user}} + \gamma \max_{a'} Q_{s',a'}^{\text{user}}) \quad (2)$$

式中, $Q_{s,a}^{\text{user}}$ 是原有的Q值, $Q_{s',a'}^{\text{user}}$ 是智能体从新的交互状态 s' 可以期望得到的新的Q值。 $\alpha \in [0,1]$ 是学习率,决定多大程度上新学习的信息将覆盖旧信息, $\gamma \in [0,1]$ 是折现系数,确定智能体应在多大程度上考虑将来的奖励高于当前奖励。通过机器学习,可以使智能体学会什么动作是可接受的,什么动作是不可接受的,产生能够适应于特定用户或用户群体的动作。

3.2.2 情绪表现

在宏观框架上需要对情绪进行具体的微观表现。情绪的表现形式有许多种,最主要的是面部表情。Wojdel 和 Rothkrantz(2005)在 Ekman(1993)的面部肌肉活动编码系统(facial action coding system, FACS)的基础上,提出基于参数的面部动画。每个参数对应一个运动单元(AU),将FACS中定义的同现原则作为一种特殊的模糊逻辑来实现面部动画,以避免不自然的面部表情。Tinwell 等人(2011)对面部表情的缺失是否会增加角色的恐怖感进行研究,探讨恐怖感的增加是否与情绪相关。实验测试了6种情绪,结果显示当上面部表情缺失时,生气和高兴两种情绪不会增加角色的恐怖感,而害怕、悲伤、厌恶和惊讶这4种情绪则会增加角色的恐怖感。Castillo 等人(2014)研究了语义驱动的面部动画,关注面部表情的语义空间,提取视频中的表情,试图寻找语义空间和面部表情的映射。Amini 等人(2015)重点开发了一款免费动画生成软件 HapFACS,用于生成3维会话智能体,以FACS系统为标准,可以有效生成逼真的面部动画。

除了表情之外,眼睛注视、姿态、头部动画和手势等其他模态也可以进行情绪的展现,但相关研究比较零散。Queiroz 等人(2008)关注交互中智能体的注视行为,建模使智能体具有表现力的眼动行为,模型能够生成与不同情感状态相对应的眼动行为,进而在交互中表达情感。Lee 和 Marsella(2010)对人们说话时的点头运动展开研究,采用机器学习方法,使用手势语料库使智能体产生点头行为,在点头时间方面比基于规则方法产生的点头更自然。Shiratori 等人(2006)通过舞蹈动作使角色体现沮丧或高兴等情感。Aristidou 等人(2015)研究人类如何

通过身体动作表达情感,采用罗素的情感模型,根据动作分析 LMA (laban movement analysis) 识别动作表达的情感状态,提出的特征可用于分析不同情感的相似性、合成和分类。

近年来,人们在触觉的情绪表现形式上也进行了一些探索。Tewell 等人(2017)使用由 3 个热刺激器组成的热刺激阵列提高文本信息的情绪唤醒能力,如图 4 所示。



图 4 用户手上的热刺激阵列
Fig. 4 The thermal array display on user's arm

4 评价方法

在人机交互中,对所提模型和算法进行评估是必不可少的。但人机交互类研究的评估与图形、图像类研究有很大不同。这类研究不一定要速度高和精度高,更多的是用户感受。如何合理地对用户

主观感受进行评估和描述是一个非常值得讨论的交叉学科问题。在人机交互中,情感计算的评估主要分为两类,一类是对情绪分类模型的准确度进行评估,另一类是评价用户与智能体交互时的感受。第 1 类评估相对容易,因为情绪数据集集中的数据本身已有情绪标签,仅需测试验证结果 (Quan 等, 2017; Samara 等, 2019) 即可,如果是自发收集的数据,则引入一些交叉验证方法 (D' Mello 和 Graesser, 2010; 任福继 等, 2018)。第 2 类评估比较主观,需要使用统计分析方法并招募评估者,一般的分析方法和评估者人数如表 2 所示。此外,可以使用李克特量表 (Bönsch 等, 2018; Volonte 等, 2016; Zibrek 等, 2019) 和名义 (定量) 尺度 (nominal scale) (Volonte 等, 2016, 2019) 描述程度,使用生理数据反映评估者的情绪状况 (Volonte 等, 2016), 使用正负性情绪量表 (positive and negative affect schedule) (Volonte 等, 2016)、情绪量表 (differential emotions scale-IV, DES IV) (Volonte 等, 2016, 2019) 和 SAM (self-assessment manikin) 量表 (Samara 等, 2019) 使评估者进行自我情绪报告,使用社会存在感量表 (social presence scale) (Bönsch 等, 2018; Zibrek 等, 2019) 测量游戏中玩家与其他玩家/游戏角色交互过程的心理和行为参与程度。

表 2 常见的评估人数和统计分析方法
Table 2 Common number of participants and statistical analysis methods

统计分析方法	评估人数
SPSS22. 0、P-value、成对比较、Bonferroni 校正	27 人 (Bönsch 等, 2018)
均值、标准差、方差分析 ANOVA、Greenhouse-Geisser 校正、图基事后检验	62 人, 分 3 组实验 (真实: 20 人, 卡通: 22 人, 素描: 20 人) (Volonte 等, 2016)
基于 Wald 统计量的回归模型、方差分析 ANOVA、Levene 检验、图基事后检验	622 人, 每个场景约 70 人/附加实验: 175 人 (Zibrek 等, 2019)
交叉滞后面板模型 (cross-lagged panel model, CLPM)、路径分析	45 人 (Volonte 等, 2019)
平均值、标准差、双尾 t 检验	23 人, 分两个实验, 一个 11 人, 一个 12 人 (Becker-Asano 和 Wachsmuth, 2010)

5 讨 论

虽然人机交互中的智能体情感计算研究已经取得不少进展,但仍存在很多亟待研究的问题。

1) 多通道的情绪感知问题。基于单通道的用

户情绪感知研究已经日趋成熟,但是基于多通道交互的用户情绪感知研究尚不够深入。由于用户可能会隐藏自己的真实情绪,不同用户的情绪可能存在不同的表达方式,仅凭单通道的表情信息,有时难以识别喜极而泣这类表情背后的真实情绪。因此,通过多通道人机交互来感知用户的真实情绪是一个值

得深入研究的方向。已经有工作开展了这方面的探索(张凯乐等, 2020), 结合多个通道的信息进一步感知用户的真实情绪。但是, 这方面的研究还需要进一步推进, 包括增强智能体的感知能力以及解决数据的融合和语义理解问题。

2) 智能体的自主学习问题。现有智能体的情绪表现力针对性不强, 不能在与用户的交互过程中进行情绪表达的自动调节。为了提升智能体情绪表达的适应性, 需要融合机器学习的方法。目前机器学习在情感计算方面的应用主要集中在情绪分类, 很少应用于智能体的情绪表现。此外, 具有学习能力的智能体应有能力通过交互捕获用户的背景信息, 并基于用户当前的情绪状态, 调整情绪表现相关参数, 做出合适的情绪反应。为此, 需要构建智能体的常识知识库和相关的推理计算方法, 以理解用户的情绪语义信息, 对用户的复杂情绪做出合理的反应, 完成内容丰富的情感交互任务。

3) 多学科交叉问题。人机交互中的智能体情感计算是一个需要多学科交叉融合的研究领域, 仅靠单一学科很难实现从研究到应用的转化。为此, 需要融合心理学、计算机、通讯、电子、医学、艺术和社会学等多学科知识, 针对不同领域需求, 提出合理可行的情感计算算法。目前, 有关情感计算的成果主要来自于计算机、通讯、电子、心理学和医学等领域, 缺少哲学、社会科学等人文领域研究人员的参与。未来需要不同专业、不同团队的研究人员开展深度合作, 共同推进研究。

6 结 语

人机交互中的智能体情感计算研究的目的是使智能体具有个性化的情感交互能力, 体现了以人为本的人机交互设计理念。智能体不仅要能准确感知用户情绪, 还要能展示合适的表情对用户进行反馈或对用户情绪进行调节, 从而达到情感关怀等功能需求。随着中国城镇化和人口老龄化趋势的发展, 采用智能体进行情感交互的市场需求越来越大, 智能体情感计算研究具有巨大的发展潜力。

本文通过检索文献数据库, 从用户情绪感知和智能体情绪表现方面对现有工作进行总结。在用户情绪感知方面, 智能体通过视觉和听觉等通道感知用户情绪状态的研究已经日趋成熟, 以深度学习为

代表的机器学习方法也在不断发展, 但关于多通道信息融合的工作尚不多见。在智能体情绪表现方面, 有关智能体造型和环境设计对用户的心理影响研究开始得到重视, 在智能体的情绪模型中集成机器学习也成为实现定量情绪表现的方法之一, 如何增强智能体的情绪表现力是下一步值得关注的方向。

智能体情感计算研究需要融合多学科领域, 学科交叉创新是推动该领域发展的关键。希望本文的工作对后续研究工作能起到抛砖引玉的作用。

参考文献 (References)

- Ajili I, Malle M and Didier J Y. 2019. Human motions and emotions recognition inspired by LMA qualities. *The Visual Computer*, 35(10): 1411-1426 [DOI: 10.1007/s00371-018-01619-w]
- Amini R, Lisetti C and Ruiz G. 2015. HapFACS 3.0: FACS-based facial expression generator for 3D speaking virtual characters. *IEEE Transactions on Affective Computing*, 6(4): 348-360 [DOI: 10.1109/TAFFC.2015.2432794]
- Aristidou A, Charalambous P and Chrysanthou Y. 2015. Emotion analysis and classification: understanding the performers' emotions using the LMA entities. *Computer Graphics Forum*, 34(6): 262-276 [DOI: 10.1111/cgf.12598]
- Becker-Asano C and Wachsmuth I. 2010. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, 20(1): #32 [DOI: 10.1007/s10458-009-9094-9]
- Bönsch A, Radke S, Overath H, Asché L M, Wendt J, Vierjahn T, Habel U and Kühlen T W. 2018. Social VR: how personal space is affected by virtual agents' emotions//*Proceedings of 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Tübingen/Reutlingen, Germany: IEEE: 199-206 [DOI: 10.1109/VR.2018.8446480]
- Castillo S, Wallraven C and Cunningham D W. 2014. The semantic space for facial communication. *Computer Animation and Virtual Worlds*, 25(3/4): 223-231 [DOI: 10.1002/cav.1593]
- Chen S Z, Wang S and Jin Q. 2018. Multimodal emotion recognition in multi-cultural conditions. *Journal of Software*, 29(4): 1060-1070 (陈师哲, 王帅, 金琴. 2018. 多文化场景下的多模态情感识别. *软件学报*, 29(4): 1060-1070) [DOI: 10.13328/j.cnki.jos.005412]
- D' Mello S K and Graesser A. 2010. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20(2): 147-187 [DOI: 10.1007/s11257-010-9074-4]
- Ekman P. 1993. Facial expression and emotion. *American Psychologist*,

- 48(4): 384-392 [DOI:10.1037/0003-066x.48.4.384]
- Fan J J, Tian F, Du Y, Liu Z J and Dai G Z. 2018. Thoughts on human-computer interaction in the age of artificial intelligence. *Scientia Sinica Informationis*, 48(4): 361-375 (范俊君, 田丰, 杜一, 刘正捷, 戴国忠. 2018. 智能时代人机交互的一些思考. *中国科学: 信息科学*, 48(4): 361-375) [DOI: 10.1360/N112017-00221]
- Fan X M, Fan J J, Tian F and Dai G Z. 2019. Human-computer interaction and artificial intelligence: from competition to integration. *Scientia Sinica Informationis*, 49(3): 361-368 (范向民, 范俊君, 田丰, 戴国忠. 2019. 人机交互与人工智能: 从交替浮沉到协同共进. *中国科学: 信息科学*, 49(3): 361-368) [DOI: 10.1360/N112018-00181]
- Filko D and Martinović G. 2013. Emotion recognition system by a neural network based facial expression analysis. *Automatika*, 54(2): 263-272 [DOI: 10.7305/automatika.54-2.73]
- Gratch J and Marsella S. 2004. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4): 269-306 [DOI: 10.1016/j.cogsys.2004.02.002]
- Han J X, Xu R Y and Chen L Y. 2020. Convolutional neural network fusing ranking and regression for expression intensity estimation. *Journal of Computer-Aided Design and Computer Graphics*, 32(8): 1228-1235 (韩加旭, 徐如意, 陈靓影. 2020. 融合排序与回归的卷积神经网络用于表情强度估计. *计算机辅助设计与图形学学报*, 32(8): 1228-1235) [DOI: 10.3724/SP.J.1089.2020.17753]
- Jiang Y Y, Li W, Hossain M S, Chen M, Alelaiwi A and Al-Hammadi M. 2020. A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. *Information Fusion*, 53: 209-221 [DOI: 10.1016/j.inffus.2019.06.019]
- Kanjo E, Younis E M G and Ang C S. 2019. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion*, 49: 46-56 [DOI: 10.1016/j.inffus.2018.09.001]
- Kessous L, Castellano G and Caridakis G. 2010. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 3:33-48 [DOI: 10.1007/s12193-009-0025-5]
- Ko B C. 2018. A brief review of facial emotion recognition based on visual information. *Sensors*, 18(2): # 401 [DOI: 10.3390/s18020401]
- Lee J and Marsella S C. 2010. Predicting speaker head nods and the effects of affective information. *IEEE Transactions on Multimedia*, 12(6): 552-562 [DOI: 10.1109/TMM.2010.2051874]
- Li S and Deng W H. 2019a. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transactions on Image Processing*, 28(1): 356-370 [DOI: 10.1109/TIP.2018.2868382]
- Li S and Deng W H. 2019b. Blended emotion in-the-wild: multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning. *International Journal of Computer Vision*, 127(6): 884-906 [DOI: 10.1007/s11263-018-1131-1]
- Li S and Deng W H. 2020. Deep facial expression recognition: a survey. *IEEE Transactions on Affective Computing* [DOI: 10.1109/TAF-FC.2020.2981446]
- Ma Y X, Hao Y X, Chen M, Chen J C, Lu P and Košir A. 2019. Audio-visual emotion fusion (AVEF): a deep efficient weighted approach. *Information Fusion*, 46: 184-192 [DOI: 10.1016/j.inffus.2018.06.003]
- Mao X and Xue Y L. 2011. *Human-Computer Emotional Interaction*. Beijing: Science Press (毛峡, 薛雨丽. 2011. 人机情感交互. 北京: 科学出版社)
- Moussa M B and Magnenat-Thalmann N. 2013. Toward socially responsible agents: integrating attachment and learning in emotional decision-making. *Computer Animation and Virtual Worlds*, 24(3-4): 327-334 [DOI: 10.1002/cav.1515]
- Pferifer R and Bongard J. 2009. *Body Intelligence*. Yu W W, Chen W D, Yang J G, Xu M and Jin D, trans. Beijing: Science Press. (Pferifer R, Bongard J. 2009. 身体的智能: 智能科学新视角. 俞文伟, 陈卫东, 杨建国, 许敏, 金丹, 译. 北京: 科学出版社)
- Quan C Q, Zhang B, Sun X and Ren F J. 2017. A combined cepstral distance method for emotional speech recognition. *International Journal of Advanced Robotic Systems*, 14(4): 1-9 [DOI: 10.1177/1729881417719836]
- Queiroz R B, Barros L M and Musse S R. 2008. Providing expressive gaze to virtual animated characters in interactive applications. *Computers in Entertainment*, 6(3): 1-23 [DOI: 10.1145/1394021.1394034]
- Rao Y, Wu L W, Wang Y M and Feng C. 2018. Research progress on emotional computation technology based on semantic analysis. *Journal of Software*, 29(8): 2397-2426 (饶元, 吴连伟, 王一鸣, 冯聪. 2018. 基于语义分析的情感计算技术研究进展. *软件学报*, 29(8): 2397-2426) [DOI: 10.13328/j.cnki.jos.005564]
- Ren F J, Yu M L, Hu M and Li Y Q. 2018. Dual-modality video emotion recognition based on facial expression and BVP physiological signal. *Journal of Image and Graphics*, 23(5): 688-697 (任福继, 于曼丽, 胡敏, 李艳秋. 2018. 融合表情和BVP生理信号的双模态视频情感识别. *中国图象图形学报*, 23(5): 688-697) [DOI: 10.11834/jig.170401]
- Rosales J H, Rodríguez L F and Ramos F. 2019. A general theoretical framework for the design of artificial emotion systems in autonomous agents. *Cognitive Systems Research*, 58: 324-341 [DOI: 10.1016/j.cogsys.2019.08.003]
- Roxas J C, Richards D, Bilgin A and Hanna N. 2018. Exploring the influence of a human-like dancing virtual character on the evocation of human emotion. *Behaviour and Information Technology*, 37(1): 1-15 [DOI: 10.1080/0144929X.2017.1386714]
- Samara A, Galway L, Bond R and Wang H. 2019. Affective state detec-

- tion via facial expression analysis within a human-computer interaction context. *Journal of Ambient Intelligence and Humanized Computing*, 10(6): 2175-2184 [DOI: 10.1007/s12652-017-0636-8]
- Sharp H, Preece J and Rogers Y. 2003. *Interaction Design*. Liu X H, trans. Beijing: Publishing House of Electronics Industry (Sharp H, Preece J and Rogers Y. 2003. 交互设计: 超越人机交互. 刘晓晖, 译. 北京: 电子工业出版社)
- Shiratori T, Nakazawa A and Ikeuchi K. 2006. Dancing-to-music character animation. *Computer Graphics Forum*, 25(3): 449-458 [DOI: 10.1111/j.1467-8659.2006.00964.x]
- Shu L, Xie J Y, Yang M Y, Li Z Y, Li Z Q, Liao D, Xu X M and Yang X Y. 2018. A review of emotion recognition using physiological signals. *Sensors*, 18(7): #2074 [DOI: 10.3390/s18072074]
- Tao J H, Yang M H, Wang Z L, Ban X J, Xie L, Wang Y H, Zeng Q, Wang F, Wang H Q, Liu B, Han Z S, Pan H and Chen W Z. 2019. Non contact multi-channel natural interactive surgical environment under sterile conditions. *Journal of Software*, 30(10): 2986-3004 (陶建华, 杨明浩, 王志良, 班晓娟, 解仑, 汪云海, 曾琼, 王飞, 王红迁, 刘斌, 韩志帅, 潘航, 陈文拯. 2019. 无菌条件非接触式多通道自然交互手术环境. 软件学报, 30(10): 2986-3004) [DOI: 10.13328/j.cnki.jos.005785]
- Tewell J, Bird J and Buchanan G R. 2017. The heat is on: a temperature display for conveying affective feedback//*Proceedings of 2017 CHI Conference on Human Factors in Computing Systems*. Colorado, USA: ACM: 1756-1767 [DOI: 10.1145/3025453.3025844]
- Tinwell A, Grimshaw M, Nabi D A and Williams A. 2011. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior*, 27(2): 741-749 [DOI: 10.1016/j.chb.2010.10.018]
- Volonte M, Anaraky R G, Knijnenburg B, Duchowski A T and Babu S V. 2019. Empirical evaluation of the interplay of emotion and visual attention in human-virtual human interaction//*ACM Symposium on Applied Perception 2019*. Barcelona, Spain: ACM: 1-9 [DOI: 10.1145/3343036.3343118]
- Volonte M, Babu S V, Chaturvedi H, Newsome N, Ebrahimi E, Roy T, Daily S B and Fasolino T. 2016. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Transactions on Visualization and Computer Graphics*, 22(4): 1326-1335 [DOI: 10.1109/TVCG.2016.2518158]
- Wisessing P, Zibrek K, Cunningham D W, Dingliana J and McDonnell R. 2020. Enlighten me: importance of brightness and shadow for character emotion and appeal. *ACM Transactions on Graphics*, 39(3): #19 [DOI: 10.1145/3383195]
- Wojdel A and Rothkrantz L J M. 2005. Parametric generation of facial expressions based on FACS. *Computer Graphics Forum*, 24(4): 743-757 [DOI: 10.1111/j.1467-8659.2005.00899.x]
- Zhang K L, Liu T T, Liu Z, Zhuang Y and Chai Y J. 2020. Multimodal human-computer interactive technology for emotion regulation. *Journal of Image and Graphics*, 25(11): 2451-2464 (张凯乐, 刘婷婷, 刘箴, 庄寅, 柴艳杰. 2020. 面向情绪调节的多模态人机交互技术. 中国图象图形学报, 25(11): 2451-2464) [DOI: 10.11834/jig.200251]
- Zhang Y, Zhang L and Hossain M A. 2015. Adaptive 3D facial action intensity estimation and emotion recognition. *Expert Systems with Applications*, 42(3): 1446-1464 [DOI: 10.1016/j.eswa.2014.08.042]
- Zibrek K, Martin S and McDonnell R. 2019. Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception*, 16(3): 1-20 [DOI: 10.1145/3349609]

作者简介



刘婷婷, 1980年生, 女, 副教授, 主要研究方向为虚拟现实和人工智能。

E-mail: liutingting@nbu.edu.cn



刘箴, 通信作者, 男, 研究员, 主要研究方向为虚拟现实和人工智能。

E-mail: liuzhen@nbu.edu.cn

柴艳杰, 女, 讲师, 主要研究方向为虚拟现实和人工智能。

E-mail: chaiyanjie@nbu.edu.cn

王瑾, 女, 讲师, 主要研究方向为虚拟现实和人工智能。

E-mail: wangjin2@nbu.edu.cn

王媛怡, 女, 讲师, 主要研究方向为应用数学和人工智能。

E-mail: wangyuanyi@nbu.edu.cn