

仿人面部表情机器人研究进展

张国亮,王展妮,王田,吴昊

(华侨大学 计算机学院,福建 厦门 361021)

摘要: 首先,梳理近年来国内外这一领域的发展现状,并从心理学维度空间角度探讨了表情机器人的情感建模方法.其次,从机构设计、人脸表情的计算模型、表情特征的提取与分类的识别等多个方面,对表情机器人研究中的关键技术展开综述.最后,在综述分析的基础上,对表情机器人研究现状及其未来发展进行系统的评价,给出今后发展的 4 个方向.

关键词: 仿人机器人;表情机器人;面部表情;面部动作编码;心理学

中图分类号: TP 391.4 **文献标志码:** A

工业机器人在制造业中的快速发展为机器人进入民用服务领域提供了良好的契机.与传统的固定工位、固定工序和固定操作场景的工业机器人不同,服务机器人对于交互性、智能性和自主性的要求更高.在服务机器人领域中,仿人面部表情机器人因其人机交互友好、学科交叉融合度大等特点,成为当前这一研究领域中的热点问题.表情机器人的研究不仅涉及到信息传感、控制等自然学科,而且还需要考虑其内在的情感及情绪等心理变化.因此,本文从表情机器人的心理学建模、物理建模和计算模型等多个角度,详细阐述了该领域的相关理论与技术,展望了未来的发展方向.

1 国内外表情机器人研究现状

目前,对具有仿人面部表情机器人的研究主要集中在欧美、日本的大学及研究机构^[1-2],近年来,国外具有代表性的仿人面部表情机器人的概括,如图 1 所示.

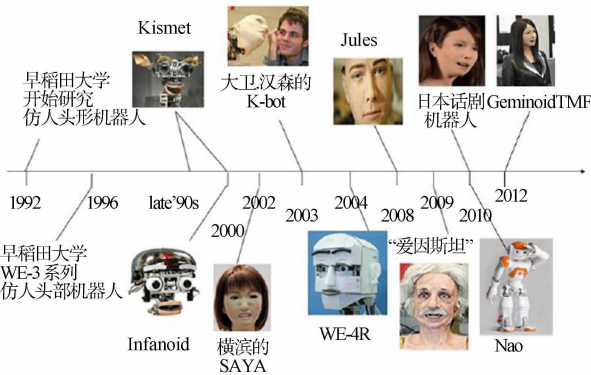


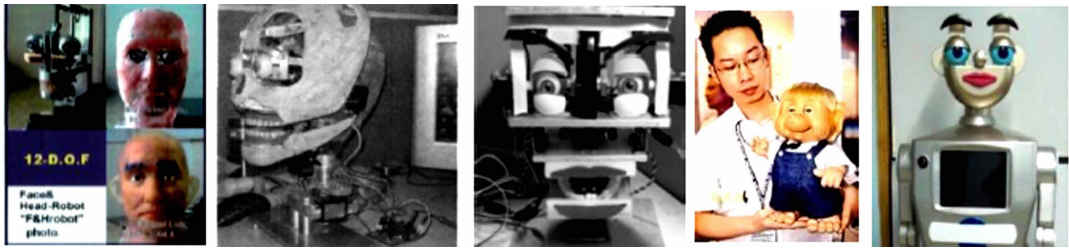
图 1 国外面部表情机器人研究进展

Fig. 1 Development of foreign facial expression robot

20 世纪 90 年代,美国麻省理工学院人工智能实验室的辛西娅·布雷齐尔博士等研发了婴儿机器人 Kismet.它是较早的具有面部表情的仿人机器人,能以类似婴儿的行为方式表达出 Ekman 定义

的喜、怒、哀、乐等 6 种基本情感状态. 日本早稻田大学开发的 WE-3RV 的情感机器人将 6 种基本情感扩充到 9 种, 并建立了具有外界刺激的三维情感空间模型. 东京理工大学开发的 SAYA 机器人较早采用了人工肌肉辅助运动机构, 从而产生面部表情. SAYA 女性的外观使得其具有较好的亲和力, 现已尝试应用于指挥交通、授课及售货等日常生活领域. 英国赫特福德大学的设计 Nao 机器人, 具有 1 岁婴儿的情感及行为学习模式. 机器人在与人类的接触过程中, 通过观察身边人的表情、动作来分析人们的情感变化, 从而与人建立双向情感联系. 日本大阪大学智能机器人教授 Hiroshi 相继研究了 ReplieeQ1, ReplieeQ2 和 GeminoidTMF 等女性仿人机器人. 其中, GeminoidTMF 以日本新闻女主播藤井雅子为原型, 动作表情与真人极其相似, 使人即便在近处也很难分辨真伪.

国内仿人表情机器人研究工作起始于哈尔滨工业大学吴伟国教授领导的团队, 2004 年, 该项目组在国内首次研制了能够表达 8 种基本表情的 H&Frobot-I, 随后又相继研究了具有面部动作协调、口形与语音识别协调功能的 H&Frobot-II 和 H&Frobot-III^[3]. 北京科技大学王志良教授^[4-5] 等较早地对表情机器人的人工情感、多信息融合交互等方面展开研究, 设计了具有人机共同注意、动态情绪调节的仿人机器人头部. 2010 年, 上海大学的柯显信等^[6] 基于 AT89S52 单片机和 7 个舵机, 设计并实现了具有 11 个自由度的结构相对简单的 SHFR-1 表情机器人. 除此之外, “百智星”幼教机器人及“童童”3 机器人也是近年来国内研发的具有代表性的表情机器人. 国内面部表情机器人, 如图 2 所示.



(a) H&F 机器人 (b) 北科大仿人头部 (c) SHFR-1 (d) 百智星幼教机器人 (e) 童童机器人

图 2 国内面部表情机器人

Fig. 2 Domestic facial expression robot

2 仿人机器人表情的心理学模型

外在的刺激可以由传感器获取, 但内在的情感及心理活动的变化却属于心理学范畴. 因此, 建立与表情机器人相适应的情感模型是实现预期表情的理论基础. 在心理学研究领域, 对于情感的研究非常广泛, 常用的方法主要有基本情感论、维度空间理论、认知情感论、个性化和情绪化情感论等. 其中, 维度空间理论由于便于将复杂的情感多维量化, 在情感机器人建模领域中受到了广泛的关注, 目前多数的具有情感模型的表情机器人都采用了类似的理论模型.

维度空间论认为, 人的所有情感分布在由若干维度构成的某一空间中, 不同情感之间不是独立的, 而是连续可渐变的. 最早提出情感可进行维度量化的是实验心理学奠基人 Wundt, 他利用愉快-不愉快、紧张-松弛、兴奋-沉静 3 个维度来表示情感. 美国心理学家 Schlosberg 提出了与 Wundt 的方法类似的倒立圆锥情感空间. Schlosberg 认为, 任何的情感都可以在椭圆锥上找到自己合适的位置, Schlosberg 的椭圆锥是西方心理学提出的第一个具有维度情感量化表的情感模型. Plutchik^[7] 发展了 Schlosberg 的模型理论, 他利用因素分析法假设了情感之间存在着 3 个基本特征: 强度、相似性和两极性. Plutchik 认为原始的情感是从进化过程中分化出来的, 居于顶端最强烈, 沿着扇形面向下, 越靠近底端情感越微弱^[8]. 同一扇面上越临近, 情感性质上越接近, 互为对角的两个扇形中的情感则是相互对立的, 距离越远, 差异越大. Plutchik 情感模型具有典型的多维量表的模型结构, 在表情机器人等应用领域得到深入的研究, 如 MIT 的 Kismet 表情机器人就采用了与 Plutchik 情感模型类似的模型结构. 将 Plutchik 的扇形空间转换为对立性和外界刺激程度二维平面, 而将强度转换为对外界刺激程度的反应. 除 Plutchik 的倒立圆锥模型, 由 Russell 等提出的 PAD 维度观测模型也是目前心理学界公认的一种普适的三维情感模型. 国内毛峡等^[9-10] 研究的表情机器人及仿生代理中都采用了相似的模型.

3 仿人表情机器人的关键技术

3.1 表情机器人机构设计

表情机器人的头部机构是表情交互的物理载体,良好的机构设计是实现表情自然交互的前提条件.人类的脸部不仅运动器官众多,而且器官运动幅度的细微瞬间变化都可能表达不同的情感.因此,为达到表情正常输出,表情头部机构的设计不仅需要从生物学的角度模仿,还需要对整个系统综合考虑,在满足机构的运动学、动力学、控制能力、驱动系统和传感器需求的前提下,机构设计一般至少需要满足以下两点要求.

1) 外观上与人类头部外形、大小、样貌相似,给人以亲切感和可信任感.人的面部器官包括眼睛、鼻子、眉毛、嘴巴和下颚等.在这些器官中,眼睛是人类最为重要的面部器官.然而,对于人类的表情,仅仅依靠人眼是无法独立表达情感的,眼部表情动作一般需要结合眼球、眼睑和眉毛一起来表达.眼球具有上、下、左、右共 4 个自由度(DOF),每一个眼睛包含上、下两个眼睑,其设计难点在于眼球运动与眼睑运动可能存在空间干涉,且由于下眼睑运动幅度较小,多数研究从简化机构设计角度,只做了一侧眼睑,但考虑到眼睑睁开的幅度对表情有很大影响,双侧眼睑运动的设计能够表达更丰富的情感.眉毛和嘴部涉及的表情变化非常灵活,目前已出现有 1,2,4,8 等不同自由度的眉毛,而嘴部运动的变化更加丰富,不仅包含平面的运动,还涉及了空间运动.因此,在表情机器人设计中,从 2~6 个自由度的形式均有出现.下颌和颈部的承载力度最大,通常需要选择载荷较大的电机加以控制.

2) 在机构设计中,需要针对表情机器人的应用形式,整体考虑空间的布局 and 自由度的分配,为驱动和传感器预留足够的空间.在表情机器人中,为使每个器官运动的幅度与人类相近,需要把大量传感器和驱动装置装入头部中,这就需要在有限空间完成复杂机构设计,整个系统具有运动空间狭窄、运动范围小的特点.针对这种特殊的应用需求,国内外学者在机构设计方面提出了大量可行的研究方案,如德国 Kaiserslautern 大学的 ROMAN 机器人利用 8 个金属板支撑骨架,依靠 10 个舵机推拉线控制硅胶面皮,实现表情动作,并加入 1 个大扭矩输出舵机控制下颌动作,实现 6 种基本表情.日本东京通信研究实验室设计了 15 个自由度的 Infanoid,其突出特点是在眼部加了 5 个 DOF,眼球能在正负 45°范围内扫视.日本早稻田大学 WE-4R 机器人共有 25 个 DOF,其设计也非常重视眼部和眉部的运动,眼球的俯仰运动与眼睑的开合能够联动.WE-4R 与其他表情机器人最大不同之处是运用蓝色和绿色的发光板来表达脸颊面色,面色并将其与眉毛、嘴唇、下颌和声音一起来表达情感.美国汉森公司研发的“爱因斯坦”是近年来具有最复杂机构的机器人表情头,它能够实现 31 个自由度,可表达 28 个脸部表情.国内北京科技大学的李娜等^[11]设计了 11 个 DOF 的表情头,眼球、眼睑、眉毛、颈部和下颚的自由度分别为 4,2,2,2,1.哈工大的“H&F”系列机器人共有 12 个自由度,面部表情的驱动装置采用了电机与绳索相结合的驱动方式,而眼睛机构采用线传动.

3.2 人脸表情的计算模型

人类的面部表情非常丰富,脸部细小的微变可能表达完全不同的情感.据统计,人脸约可产生 1 万种表情,而其中的 3 000 种是具有情感意义的.因此,为使表情机器人正确表达情感,需要对这些表情进行编码、统计,从而使其在不同的情境做出合适的表情,其中具有代表性的表情编码系统包括:面部动作编码系统(facial action coding system,FACS)、情感面部动作编码系统(emotional facial action coding system,EMFACS)、最大判别面部运动编码系统(maximally discriminative facial movement coding system,MAX)、MPEG-4 三维人脸动画标准和面部表情编码系统(facial expression coding system,FACES)^[12]等.

FACS 是美国心理学教授保罗·艾克曼于 20 世纪 70 年代末提出的,FACS 根据解剖学定义了 44 个运动单元,及运动描述符来描述面部运动,每个运动单元具有不同的位置和强度.FACS 是迄今为止最为详尽、最为精细的面部测量系统,很多后续的方法都是以它作为基础展开的,例如,2002 年,Ekman 在 FACS 系统的基础上,定义了 6 种基本表情及 33 种不同的表情倾向,确定了改进的简化 FACS 系统.EMFACS 则是将情感相关的面部运动与原始的 FACS 结合,对面部表情重新进行编码.MAX 系统由 Izard 提出,该系统提供一个有效和可靠的系统测量婴儿和儿童的面部行为中的情感信号.MPEG-4 三

维人脸动画标准利用人脸定义参数(facial definition parameter,FDP)描述特定人脸纹理形状模型与通用人脸模型之间的差别,利用人脸动画参数(facial animation parameter,FAP)描述特定人脸表情与中性表情的变化关系。MPEG-4中共定义了68个FAP,每个FAP分别定义人脸某一小区域的运动。在FAP基础上,MPEG-4还利用了FAP之间的空间关系,定义了人脸动画参数单元(facial animation parameter unit,FAPU),及其84个特征点,用于校准人脸和产生人脸动画。美国加利福尼亚大学的Kring等提出的FACES提供了表情的频率、强度、诱惑力和持续时间等信息,能对分布在情感维度模型上的表情进行度量,并且所需时间较短。FACS考虑了身体位置的移动、手部的遮挡、眼睛、讲话、面部抽搐等对表情的影响,增强了系统的可靠性。

3.3 人脸表情识别技术

人脸表情识别是表情机器人应用的关键支撑技术。虽然近年来人脸检测技术已逐渐成熟,但与之密切相关的表情分析识别方法仍有许多难点有待克服,而表情数据库的建立、人脸特征的检测、分类的识别是其核心的研究内容。

3.3.1 人脸表情库 目前人脸表情数据库大致分为3类。第一类是以6种基本情感为基础,直接建立具有不同情感的表情库,例如马里兰大学人脸表情库^[13],包含40种不同文化和种族的测试者6种基本情感的人脸表情库。荷兰代尔夫特理工大学MMI人脸表情数据库包含50名测试者的6种基本情感,共500个静态图像和2000个视频。德国慕尼黑理工大学FG-NET数据库包含18名测试者的6种基本情感表情和中性表情的序列。伊朗谢里夫理工大学DML-SUT人脸表情数据库包含10名被试者6种情感表情和中性表情的可见光和红外视频序列。五邑大学建立了包含6名男性和5名女性的6种基本情感表情和中性表情的CED-WYU人脸表情库^[14]。中国科技大学的建立了包含6种基本情感的可见光和红外人脸表情图像库USTC-NVIE^[15]。第二类是以FACS是编码为基础,建立基于量化数据的细微表情库。例如,卡内基梅隆大学机器人研究所建立的包含单一运动单元和复合运动单元共23种表情人脸表情数据库CK+。第3类是针对较大人脸表情姿态变化的三维表情数据库。如纽约州立大学宾汉姆顿分校建立了BU-3DFE库^[16],包含100名不同种族被试者6种基本情感表情和中性表情的三维信息。英国中央兰开夏大学建立的ADSIP3-D动态人脸表情数据库包含10名被试者的害怕、嫌恶、生气、高兴、惊讶、悲伤和疼痛7种表情,并分为轻微、一般和极度3种强度^[17]。

除了以上3种典型表情库之外,混合型表情数据库也受到学者的关注,如北京航空航天大学建立的共包含18种单一表情、3种混合表情和4种复杂表情的BHU数据库^[18]。中科院计算技术研究所人脸数据库建立了具有姿势、表情、饰品、光照、背景、距离和时间7种不同环境类别的人脸表情库CAS-PEAL^[19],其特殊性是考虑了环境因素的影响。

3.3.2 表情特征提取与分类识别 人的面部表情涉及了大量复杂的肌肉运动变化,在考虑特征选择及分类模型时,必须考虑柔性形变对于识别性能的影响。因此,中性表情脸的差异化形变特征选取是表情分析识别领域最常采用的特征提取方法。常采用的方法有活动外观轮廓、点分布模型、Gabor图像变换和采用主成分分析方法(PCA)等。

表情的分类器模型是建立在特征选择的基础之上的,主要包括隐马尔科夫模型法、神经网络法、时空运动能量模板法、支持向量机、Adaboost、基于规则推理的方法、PCA等。如胡敏等^[20]提出了一种基于梯度Gabor直方图的人脸表情特征识别方法。蒋斌^[21]在判别分量分析算法的基础上,提出了针对人脸表情识别的局部判别分量分析算法。希腊雅典国家技术大学Caridakis等^[22]提出了对用户和上下文自适应的神经网络方法进行表情识别的方法。北航大学的Wang等^[23]引入流行学习中较为成熟的LLE代替AAM中的PCA方法描述面部运动,在减小图像残差的基础上,实现了面部较大变化的识别。

4 未来展望

表情机器人集成了人工智能、认知科学、信息科学等多学科的研究成果,近年来取得了较快的发展,但距离完全自然的人机交互还有一段距离,其研究仍然面临许多挑战性课题,从目前的研究状况来看,以下3个方向将是今后发展的趋势。

4.1 特殊材料的应用

从近年来已面世的表情机器人来看,头部机械结构虽然较为复杂,但在机构设计方面基本已经成熟,并无太多实质性的研究进展.与之对比,表情机器人的外观和内部传感领域却硕果颇丰,其突出代表是汉森公司发明的人造皮肤和东京理工大学研制的气动人工肌肉.近十年来,汉森公司在人造皮肤领域取得了令人惊叹的成就,他们利用人造橡胶和泡沫混合物制造的人造皮肤外观仿真度极高,不仅可完整地模仿、表达人类的 28 种面部表情,且面部会随着年龄的变化而出现皱纹.东京理工大学研究的人工肌肉当被充气 and 吸气时,可以像人的肌肉一样伸缩,这种气动结构的肌肉组织具有质量小、运动平滑、动力强劲等特点,适合较大幅度的表情运动.

4.2 特殊情境下的表情识别研究

传统的表情分析、识别针对的是人类的正常状态,而特殊情境包括环境存在遮挡、被识别者本身受到某种外界因素干扰,如疲劳和疼痛等.鲁棒化识别和细微化识别是这一类特殊应用必须面对的课题.与鲁棒性识别相比,表情的细微化识别难度更大,目前,主要有概率模型和强度模型两种方法.由于人脸细微化表情在运动速度、幅度和时间上特殊性,目前的研究方法仅能对特定细微表情建模分析,很难对有意为之的伪表情和不经意流露出的微表情进行识别.因此,如何提高细微化表情的自动识别的准确性和鲁棒性,将是今后研究者深入思考的内容.

4.3 多信息、多方法的融合应用

考虑到表情建模,尤其是细微表情变化度量的复杂性,多信息、多方法的融合受到学者的广泛关注,这一策略包含两方面内容:一是方法策略间的融合;二是通过方法间的互补性来提高复杂表情的识别效率.Zhao 等^[24]提出了结合模糊核分类和支持向量机的面部复杂表情识别方法;Littlewort 等^[25]结合 Adaboost 和 SVM 建立了自发疼痛表情识别系统;为提高痛苦表情识别的准确率,张伟等^[26]提出基于监督保局投影与多核线性混合支持向量机的识别方法.

5 结束语

类人表情机器人研究涉及心理学、认知科学、计算机智能、自动控制、机械设计等多领域知识,具有典型的多学科交叉特征.文中系统分析了该领域国内外研究工作的最新进展,并从心理学理论模型和应用关键技术两个方向详细阐述了当前亟需解决的问题,展望了表情机器人未来的发展方向.

参考文献:

- [1] HIRTH J, BRAUN T, BERNS K. Emotion based control architecture for robotics applications[C]// Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Italy: IEEE Press, 2007: 2150-2155.
- [2] HIRTH J, BRAUN T, ERNS K. Emotion based control architecture for robotics applications[C]// Proceedings of 30th Annual German Conference on Artificial Intelligence, Osnabrück: IEEE Press, 2007: 464-467.
- [3] 吴伟国, 宋策, 孟庆梅. 仿人头像机器人“H&F Robot-III”语音及口型系统研制与实验[J]. 机械设计, 2008, 25(1): 15-20.
- [4] 王巍, 王志良, 郑思怡, 等. 一类具有相同结构的表情机器人共同注意方法[J]. 机器人, 2012, 34(3): 265-274.
- [5] 刘欣, 谢仑, 杨文祥, 等. 表情机器人的动态情绪调节过程研究[J]. 控制理论与应用, 2011, 28(7): 936-947.
- [6] 柯显信, 唐文彬, 陈玉亮, 等. 仿人面部表情机器人的控制系统设计[J]. 上海大学学报(自然科学版), 2011, 17(1): 79-84.
- [7] PLUTCHIK R. Emotion: A psychoevolutionary synthesis [M]. New York: Harper & Row Publishers, 1980: 30-52.
- [8] RUSSELL S J, SUBRAMANIAN D. Provable bounded-optimal agents[J]. Journal of Artificial Intelligence Research, 1995, 2(1): 575-609.
- [9] MAO Xia, XUE Yuli, CHENG Leelun, et al. Harmonious Graphics Generating Based on the $1/f$ Function Theory [J]. Chaos, Solitons and Fractals, 2007, 32(2): 521-525.
- [10] 包海燕. 智能仿生代理的情感建模方法研究[D]. 北京: 北京航空航天大学, 2010: 45.
- [11] 李娜, 陈工, 王志良. 表情机器人设计与实现[J]. 机器人技术, 2007, 23(12): 231-234.
- [12] KRING A M, SHOAN D M. The facial expression coding system (FACES), development, validation, and utility

- [J]. *Psychological Assessment*, 2007, 19(2): 210-224.
- [13] BLACK M J, YACOB Y. Recognizing facial expressions in image sequences using local parameterized models of image motion [J]. *International Journal of Computer Vision*, 1997, 25(1): 23-48.
- [14] 朱明早, 李树涛, 叶华. 基于稀疏表示的遮挡人脸表情识别方法[J]. *模式识别与人工智能*, 2014, 27(8): 708-712.
- [15] WANG Sangfei, LIU Zhilei, Lyu Siliang, et al. A natural visible and infrared expression database for expression recognition and emotion inference[J]. *IEEE Transactions on Multimedia*, 2010, 12(7): 682-691.
- [16] RABIEE H R. DML-SUT facial expression database[EB/OL]. [2011-03-10]. <http://fedb.dml.ir>.
- [17] FROWD C D, MATUSZEWSKI B J, SHARK L K, et al. Towards a comprehensive 3D dynamic facial expression database[C]// *Proceedings of the 9th WSEAS International Conference on Multimedia, Internet & Video Technologies*. Hungary: IEEE Press, 2009: 113-119.
- [18] 薛丽丽, 毛峡, 张帆. BHU 人脸表情数据库的设计与实现[J]. *北京航空航天大学学报*, 2007, 33(2): 224-228.
- [19] GAO Wen, CAO Bo, SHAN Shiguang, et al. The CAS-PEAL large-scale Chinese face database and base-line evaluations[J]. *IEEE Transactions on Systems*, 2008, 38(1): 149-161.
- [20] 胡敏, 朱弘, 王晓华, 等. 基于梯度 Gabor 直方图特征的表情识别方法[J]. *计算机辅助设计与图形学学报*, 2013, 25(12): 1856-1862.
- [21] 蒋斌, 贾克斌. 局部判别分量分析算法[J]. *电子学报*, 2014, 42(1): 155-159.
- [22] CARIDAKIS G, KARPUZIS K, KOLLIAS S. User and context adaptive neural networks for emotion recognition [J]. *Neurocomputing*, 2008, 71(13/14/15): 2553-2562.
- [23] WANG Xiaokan, MAO Xia, CALEANU C D. Nonlinear shape texture manifold learning[J]. *IEICE Transactions on Information and Systems*, 2010, E93-D(7): 2016-2019.
- [24] ZHAO Hui, WANG Zhiliang, MEN Jihui. Facial complex expression recognition based on fuzzy kernel clustering and support vector machines[C]// *Proceedings of 3rd International Conference on Natural Computation*. Haikou: Natural Computation, 2007: 562-566.
- [25] LITTLEWORT G C, BARTLETT M S, LEE K. Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain[C]// *Proceedings of the 9th International Conference on Multimodal Interfaces*. Haikou: ACM, 2007: 15-21.
- [26] 张伟, 黄炜, 夏利民, 等. 基于 SLPP 与 MKSVM 的痛苦表情识别[J]. *计算机工程*, 2013, 39(12): 196-199.

Development of Humanoid Facial Expression Robot

ZHANG Guoliang, WANG Zhangni, WANG Tian, WU Hao

(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract: In this paper, the development status of this field has been investigated recently and the method of expression robot modeling has been explored for expression robot. Firstly the key technologies of expression robot are surveyed in detail, such as mechanism design, expression computation model, expression feature extraction and classification and so on. Furthermore, the research status and future development of facial expression robot are evaluated based on the survey analysis and four directions of future development are presented.

Keywords: humanoid; expression robot; facial expression; facial action coding; psychology

(责任编辑: 陈志贤 英文审校: 吴逢铁)