

智能家居中虚拟人的设计与生成

姚智皓^{1,2)}, 乔玥涵^{1,2)}, 徐千尧^{1,2)*}

¹⁾ (清华大学美术学院 北京 100084)

²⁾ (清华大学未来实验室 北京 100084)

(xuqy@mail.tsinghua.edu.cn)

摘 要: 随着智能家居的快速发展, 国内外学者已经对交互界面进行了多方面的研究. 立足虚拟人技术, 重点探索智能家居中虚拟人的潜在需求. 基于设计师的人物设计经验, 将各异的人脸五官特征归类为有限个解剖维度, 以此作为设计规则, 设计人脸五官的卡通风格化模型模板, 比对人脸图像特征构建模板库. 在此基础上, 结合人脸特征识别及基于美学计算的特征——卡通形象模板库映射关系, 实现从二维真实人脸图像到三维卡通人脸的自动生成. 该生成方法轻量且能够满足用户的个性化需求, 并已投入实际应用.

关键词: 人机自然交互; 虚拟人; 智能家居

中图分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023.20052

Research on Virtual Human Design in Smart Home

Yao Zhihao^{1,2)}, Qiao Yuehan^{1,2)}, and Xu Qianyao^{1,2)*}

¹⁾ (Academy of Arts & Design, Tsinghua University, Beijing 100084)

²⁾ (The Future Laboratory, Tsinghua University, Beijing 100084)

Abstract: With the rapid development of smart home, researchers around the world have carried out numerous studies on its interactive interface. Based on virtual human technology, this paper explores the potential demand of virtual humans in smart home scenarios. Through the character design experience of designers, different facial features are grouped into limited anatomical dimensions as design rules. Under the guidance of these rules, cartoon stylized facial model templates can be made via comparing face image features to build a template library. Combined with face feature recognition as well as the mapping relationship between the face features and the cartoon model template-based aesthetic computing, automatic generation from two-dimensional real face images to three-dimensional cartoon faces is realized. This generation method is lightweight, meets users' personalized needs, and has been put into practical application.

Key words: natural human-computer interaction; virtual human; smart home

当前的智能家居通常采用触摸式交互结合语音交互的方式, 通过触摸屏的中控面板或智能手机结合如“小爱同学”“小度”“Alexa”“Siri”等语音助手实现用户与智能家居环境的控制和反馈. 近

年来, 随着虚拟人技术的发展, 虚拟人技术已经在各个领域展现了巨大的潜力, 融入人们的日常生活, 包括虚拟人直播、虚拟试衣和展馆导航等. 在智能家居领域, 为智能语音助手增加生动的虚

收稿日期: 2022-11-08; 修回日期: 2022-12-06. 基金项目: 国家自然科学基金(62172252); 中国博士后科学基金(2021M701914). 姚智皓(1996—), 男, 博士研究生, 主要研究方向文化遗产数字化、虚拟人设计等; 乔玥涵(1999—), 女, 硕士研究生, 主要研究方向美学计算、嗅觉计算等; 徐千尧(1990—), 女, 博士, CCF 会员, 论文通信作者, 主要研究方向移动终端用户界面、美学计算、虚拟人设计、用户体验等.

拟人形象也变得更加可行. 文献[1-5]的研究表明, 与传统的语音交互相比, 在语音交互之外加入精心设计的虚拟人形象作为视觉界面呈现, 可以增加用户对于计算机系统的信任, 有助于建立更高效、舒适和友善的沟通交互过程. 文献[6]研究表明, 虚拟人这种拟人化代理的社会影响可以与人类相当. 综上所述, 当可自行设计的虚拟人被用作智能家居中的交互界面时, 交互界面可能具有独特的互动潜力和新的效果. 因此, 虚拟人交互助手或虚拟管家将会成为一个新的智能家居场景中的研究及尝试方向.

本文首先通过设计工作坊调研了用户对于智能家居虚拟人界面的潜在需求, 据此提出并实践了基于单目摄像头的卡通虚拟人脸设计及生成方法, 生成结果达到了较高的用户满意度. 基于本文方法, 智能家居厂商可以为用户提供个性化的虚拟人形象生成平台, 并将用户设计的虚拟人形象结合到智能语音助手上进行呈现, 为未来的智能家居交互界面提供了可能性.

1 智能家居中的交互界面

1.1 虚拟人在交互中的应用

虚拟人作为一种结合了视觉、听觉的拟人化技术, 可以为智能系统加入类人的形象、形态、动作和表情等信息, 使得用户与智能系统的交互更加自然、舒适, 对智能系统的信任感和满意度等评价得到一定程度的提升. 这种优势在诸多领域内均得到了体现: 在智能医疗方面, Bickmore 等^[2]为即将出院的病人开发了用于介绍自我护理方案的虚拟护士, 患者对这种新呈现方式的接受和喜爱程度甚至超过了真实的医生或护士. 除了医院患者, 特殊人群和老年人也是虚拟人交互研究中经常关注的群体. Ali 等^[7]为自闭症青少年开发了一套可以随时随地进行线上对话的虚拟人系统, 并总结出部分针对自闭症群体的虚拟人界面设计建议. Chattaraman 等^[8]则关注老年人网上购物问题, 通过在网商店中嵌入了一个具有搜索和导航等功能的虚拟代理, 提升了老年用户对网商店的了解、信任程度和光顾意愿. Vardoulakis 等^[9]设计的虚拟管家可以与老年用户讨论天气、家庭、未来活动计划以及与老年群体相关的社会话题, 对老年人进行陪伴. 另外, 虚拟人技术还常常应用于教育与生活当中. Ramachandiran 等^[10]为本科生开发了以学习者为中心的虚拟教师. 通过实验发现, 这种

具有高情感吸引力的虚拟教师, 会让用户在互动中给予更积极的回应. 虽然虚拟人已经在诸多领域里进行了尝试及研究, 并普遍取得了较积极的研究成果, 但这些研究中使用的虚拟人形象普遍较为粗糙, 与当今生活中常用的界面显示效果相差甚远. 因此, 在当下智能系统中, 生成视觉效果较好的虚拟人仍是一个值得研究的问题.

1.2 虚拟人形象在交互中的作用

在人与人的交流中, 传递信息的途径有 2 种: 一种是通过口头、书面等方式, 明确地进行消息传输的显式通信渠道; 另一种则是涉及表情、手势和肢体动作等方式, 间接地进行信息传递的隐式通信渠道^[1]. 对应到人机交互中, 传统的语音交互方式只涉及其中的显式通信渠道. 而在语音交互的基础上加入虚拟人技术之后, 虚拟人的表情、肢体语言等可以有效地弥补隐式通信渠道的空缺, 辅助人机之间的信息和情感传达, 提高交互效率, 提升交互体验^[1,3,5]. 这种能力显著地体现在人机之间构建信任和融洽关系的方面. 虚拟人的人格化界面可以消除交互中的匿名性, 从而促进人们对计算机系统产生信任感^[3]. Bickmore 等^[2]认为, 在用户与虚拟人的面对面互动中, 虚拟人的对话和非语言行为可以让其更好地获得用户的信任, 并达到较高的交流满意度. Vardoulakis 等^[9]同样肯定了虚拟人可以通过对话交流与用户建立融洽关系, 从而让用户愿意长期、持续地与之交流, 并实现陪伴功能.

在交流行为之外, Bergmann 等^[4]的工作表明, 用户对于虚拟人的外观同样十分敏感, 与机器人形的虚拟人相比, 人形虚拟人更能够令用户在第一次和后续的长时间接触中感到更加舒适. 而精心设计的人形虚拟人形象可以使用户与计算机系统的交互更加愉快^[3]. 在各种使用场景下, 虚拟人的外表和身份对于使用者同样存在影响. 有些实验选用不同年龄、不同吸引力程度和不同外貌表现的虚拟人对学生进行学业建议指导^[11-12], 实验结果表明, 受试学生更容易受到与自己外表等特征更相似的虚拟人的影响^[13]. Shibani 等^[14]使用了吸引力不同的男性和女性虚拟导师对学生进行指导, 虽然接受不同虚拟导师教学的学生在成绩上没有显著的区别, 但学生对课程展现出的兴趣受到了虚拟导师外表的影响. 当虚拟人作为自我的化身时, 虚拟自我的外表特征也会导致使用者在现实中产生行为变化^[15]. 让用户观看一个形似自己的虚拟形象在锻炼和减肥, 会使用户在现实世界中

进行更多锻炼的同时, 饮食也更健康^[16]。除此之外, 当前智能语音助手在复杂指令或嘈杂环境下的交互中依然存在一定量的理解和执行错误, 通过语音的单一模态反馈会让用户对于智能系统的错误产生直接的负面感知。Cheng 等^[17]的研究工作表明, 加入可爱外表的虚拟人可以极大地减轻用户对系统错误的负面感知, 有效地提升交互中面对错误反馈时的用户体验。

1.3 虚拟人生成技术

三维虚拟人脸的制作流程复杂且乏味, 设计师需要花费大量的精力来制作三维形象。由于不同人脸之间具有结构相似性, 因此自动生成三维虚拟人脸具有可行性和产业化的能力。在真实的虚拟人脸生成方面, 已经能够基于单幅或者多幅图像进行三维重建^[18-19]。如果结合深度信息, 就可以实现更高精度的建模^[20]。然而, 真实人脸的“恐怖谷”效应较为明显, 用户对于其效果往往喜爱程度不高甚至反感^[21]。

近年来, 越来越多的研究工作关注到了非真实(风格化)的三维人脸生成方法。基于已有的三维模型, Zhou 等^[22]利用局部映射变形, 提出了一种把真实三维人脸转化为三维卡通模型的方法。为了具有更高的创作自由度, Sucontphunt^[23]把真实的三维人脸和目标风格化人脸混合起来, 让用户可以交互式地调整三维模型的卡通化程度。虽然这些方法已经可以便捷地产生卡通人脸, 但前提是必须拥有用户的真实三维模型, 成本较高, 通常难以满足。

由真实照片直接得到卡通人脸的研究中, 作为较早的三维形象重建工作, PICASSO 系统^[24]通过提取面部区域特征点来进行人脸的三维重建, 但是其需要多角度照片且算力要求较高, 但产生的模型精度较低。Liu 等^[25]把人脸分类为 5 个模板, 并在全局和局部级别上夸大输入图像来实现卡通化头像, 并为模型映射了纹理。Liu 等^[26]先用主成分分析重建算法对大量的二维漫画进行重建, 然后将其作为训练集建立学习回归模型, 实现二维到三维的生成工作。上述基于训练的方法均需要大量的三维模型, 不仅不易获得且不易迁移卡通风格。Lim 等^[27]利用移动设备拍摄单张照片进行简单的面部重建, 并映射上面部纹理, 但是其对纹理的处理比较粗糙, 且面部网格精度不够, 不足以支持进一步的动画制作。

上述的生成方法中, 设计师和艺术家参与的部分极少, 最终产生的卡通人脸难以满足用户的

审美需求, 并且算力要求较高, 不利于多种设备使用。基于此, 本文提出了一种轻量级三维卡通虚拟人脸生成方法, 基于三维标准人脸模型和计算机视觉算法进行卡通模型生成, 只需要单张照片就可以快速地计算出对应的卡通人脸。研究结果表明, 生产的模型与本体相似度较高, 用户喜爱程度较高, 同时模型布线标准, 可方便地进行面部动画制作以及进一步使用。

2 面向智能家居的虚拟人设计研究

2.1 智能家居中的虚拟人应用探索

为探索虚拟人在智能家居中的应用场景, 本文邀请了 12 位背景不同的设计师参与针对智能家居中虚拟人应用的工作坊。工作坊首先介绍了智能家居中交互界面以及虚拟人技术的发展情况, 以便于参与者对此问题有更好的理解; 其次, 请参与者用简单的形容词来描述虚拟人在此情景下的优势, 并进行归类和第一轮自由讨论; 最后, 在此基础上, 参与者会更进一步地描述虚拟人出现的具体场景和形式, 并在此轮结束后详细地陈述构想的动机和理由。图 1 所示为虚拟人潜在应用场景工作坊便签。

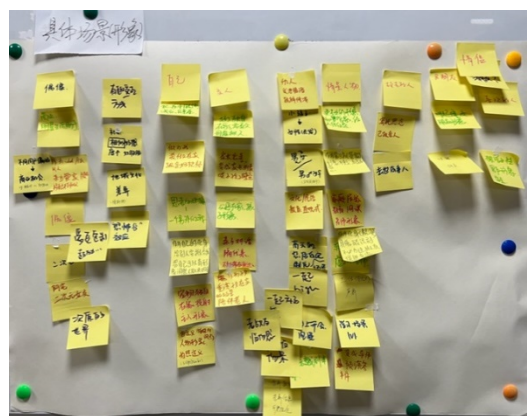


图 1 虚拟人潜在应用场景工作坊便签

2.1.1 虚拟人的优势

本文分析并归类了参与者提出的智能家居中虚拟人的优势(一共 47 个观点), 数量从高到低依次是情感、身份、系统和功能等 4 个方面。在情感方面, 大部分参与者认为, 虚拟人可以提供更自然的人机交互, 使用者会非常自然地与人形代理进行交流, 并且虚拟人可以通过肢体动作和表情等非常有人类特征的方式进行信息传达; 在身份方面, “代入感”与“临场感”是讨论的核心, 一致认为

虚拟人具有定制不同身份和外形的潜力,可以提供多样性的交互体验;在系统和功能 2 个方面,参与者更偏重于虚拟人作为智能管家的有用性和多功能性.最后,参与者均表示对于虚拟人在智能家居中作为交互界面抱有较高的期待.

2.1.2 虚拟人的潜在应用场景

此环节主要着眼于具体的虚拟人形象和对应的应用场景,共收到 47 份回答.根据虚拟人的身份不同,大致分为了自己、亲人、偶像和情景人物等 4 类对象.当虚拟人是自己的形式时,参与者更希望虚拟人能集成自己的身体数据,同时结合饮食进行健康检测;当对象是亲人时,老人和孩子则成为了关注的重点,普遍认为虚拟人在家里可以进行陪伴和提供信任,如化身为孩子的父母或者老人的孩子;当虚拟人化身为偶像或者具体情景人物时,参与者则期望虚拟人能够在居家的环境下提供各种各样的专业服务.由此可以看出,大家对于虚拟人在智能家居环境下的需求各不相同,并且存在可变的动态需求,即在不同的时间和场合可能需要虚拟人改变相貌.同时,也注意到参与者对于虚拟人陷入“恐怖谷”的担忧.

2.2 虚拟人风格调研

以往研究的卡通人脸生成风格多种多样,其中绝大部分工作均未研究所生成的人脸风格是否符合用户的需求,用户是否愿意在各种场合下使用.有实验表明,人们对不同的风格的卡通虚拟人接受程度大不相同,在不同环境下关注的侧重点也有较大差异.在家居环境下,用户所处的时间较长,如何提高用户对虚拟人形象的接受程度和偏好尤为重要.据此,在前期工作中,本文以问卷形式调研了大量人群对于虚拟人风格的态度和喜好,调研共收集到 202 份有效问卷,问卷对年龄、性别 2 个变量采取了配额规则,保证了男女比均衡为 1:1,在 18 岁及以下、19~33 岁、34~48 岁、48 岁以上 4 个年龄段中,作答者的年龄段比如图 2 所示.

针对喜好分析,作答者需要在 33 个词中选出 5 个词形容其理想中的虚拟形象,词频统计结果如表 1 所示.其中,词频最高的 5 个词:有吸引力的(0.55)、有创新性的(0.50)、有品质的(0.45)、可爱的(0.55)和印象深刻的(0.40).女性比男性更在意“可爱的”“有品质的”,男性比女性更在意“有创新性的”“个性化的”,男性群体的偏好比女性更多元,词频呈现的方差较大.而在年龄差异上,全年龄段对“有吸引力的”有显著偏好,“可爱的”“亲切的”的词频与年龄正相关,即年龄越大,词频越高.

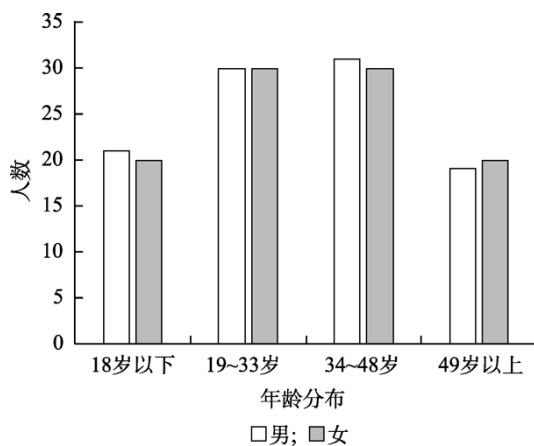


图 2 年龄段统计

表 1 对虚拟人风格喜好的描述词词频统计

描述词	词频	描述词	词频
有吸引力的	0.55	亲切的	0.20
可爱的	0.55	耳目一新的	0.15
有创新性的	0.50	年轻的	0.15
个性化的	0.50	令人信任的	0.10
高品质的	0.45	私人的	0.10
印象深刻的	0.40	简单的	0.05
喜欢的	0.35	复杂的	0.05
时尚的	0.30	缺乏人情味的	0.05
自由的	0.25	其他	0.05

在风格调研部分,共调研了 5 种常见的卡通人物风格如图 3 所示,人群最偏向组 4 和组 5;且对组 4 和组 5 的偏好不相上下(2.81/5, 2.79/5).组 4(2.81) > 组 5(2.79) > 组 1(2.47) > 组 3(2.14) > 组 2(2.11).人群普遍共识组 4 是“有吸引力的”或“时尚的”,对组 4 的评价褒义词和外观词较多;组 5 是“可爱的”或“时尚的”,对组 5 的评价褒义词和外观词较多.人群对组 4 风格有显著偏好,在“美化”“简约”词频中有明显偏好,在简约-复杂维度中无明显偏好.

2.3 卡通风格化面部特征设计研究

人脸特征复杂而多样,已经有美学研究对人脸特征进行了较为明确的划分^[28],而风格化的人脸包含大量主观的视觉感知特征(如风格、语义等).如何将这些特征量化为结构化的数据以表达设计的特点,是设计师解读和资源挖掘过程中需要探究的问题.据此,面向专业设计师开展卡通虚拟形象设计工作坊,从设计卡通虚拟形象的风格特征、五官特征等方面进行设计和探索,最终总结归纳出了面向智能家居这种私人空间的设计准则和设计指导,构建了本生成系统的美学基础.

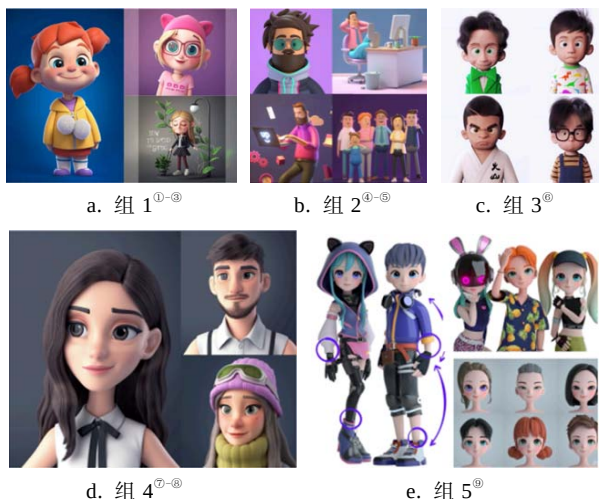


图3 风格调研

设计师的工作是将其以往的人脸卡通形象经验总结为可量化的设计知识, 对不同年龄段、不同性别、不同种族和不同面部特征的人物设计的表现方法进行总结, 把人脸中五官的特征总结为如图4所示130多个解剖维度, 这些维度能够基本覆盖人物五官的特征。例如, 在眉眼特征上, 将外眼角大于内眼角, 高度高于内眼角且眼裂细长的眼型定义为丹凤眼; 眼睛长宽比保持在3:2内, 外眼角相连成线的眼型定义为杏眼; 眼裂高宽眼白露出较多, 且瞳仁色深圆润的眼型定义为圆眼等。眉毛也遵循三庭五眼比例, 以眉头、眉峰、眉尾三点定位的形状确定眉眼的比例和眉形与用户的眉形相

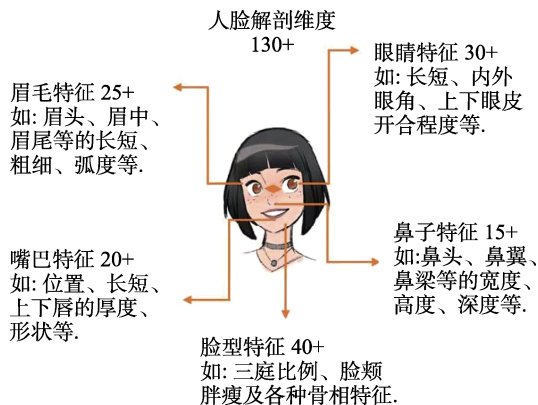


图4 五官维度

匹配。其他脸部特征也依据此设计流程, 将抽象形态量化为可计算的数据。根据此方法将总结出的人脸面部特征规律排列组合出卡通人脸, 并与用户的实际长相在五官上一一匹配, 这些维度能够基本覆盖人物五官的长相特征。例如, 总结人物的眉眼特征: 柳叶眼、杏眼、下垂眼、丹凤眼、桃花眼、肿泡眼, 可以排列组合出 n 种长相的人物, 有了这一部分的设计特征总结工作, 才有可能实现最终的“千人千面”和“自动生成”。

3 卡通虚拟人脸生成方法

本系统的框架如图5所示, 分为特征库建立和识别生成2大部分。用户拍摄照片后, 利用计算机视觉技术对其进行分析, 得到多维度的特征向量, 并与特征库中的向量进行最小差值匹配; 确认最相似模板后, 针对细节差异进行 blendshape(后称形变系数)的微调, 从而得到最终的模型。

建立特征库是基于第2节虚拟形象设计工作坊提炼的人脸特征维度。首先, 针对亚洲平均人脸通过计算机视觉的方法进行特征识别, 并让三维设计师对其进行卡通化建模并制作对应的形变系数。其次, 输入200幅各具特征的真人人脸图像作为预制数据集, 与平均人脸的特征作差, 从而得到相应的形变系数, 并由设计师进行调整, 最终生成200组特征值-形变系数的特征库。

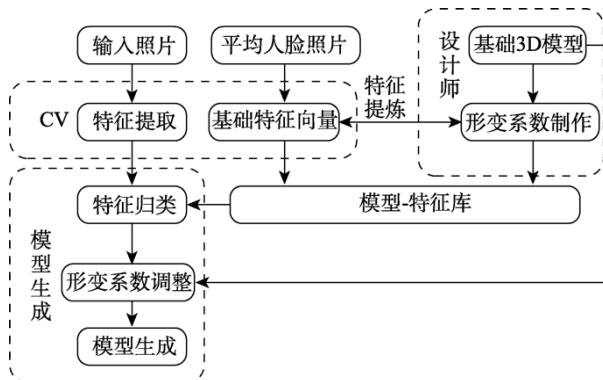


图5 系统框架

- ① https://www.behance.net/gallery/145224799/cute-girl?tracking_source=project_owner_other_projects
- ② <https://www.artstation.com/artwork/4x2dq>
- ③ <https://www.artstation.com/artwork/XEDDY>
- ④ <https://dribbble.com/tags/vago>
- ⑤ <https://dribbble.com/shots/14126676-Working>
- ⑥ <https://www.behance.net/gallery/44299737/The-Little-Brats>
- ⑦ <https://www.behance.net/gallery/57332553/Hostess-characters>
- ⑧ <https://www.artstation.com/artwork/KZ3Ry>
- ⑨ <https://www.behance.net/gallery/86918231/-3D>

3.1 特征提取

3.1.1 特征点检测

本文采用了华为 HIAI 的人脸检测技术, 对输入图像进行人脸五官特征检测, 返回代表五官轮廓位置的关键点坐标(目前数量为 276 个特征点). 可以为美颜处理、人脸建模、表情识别等后续算法处理提供输入.

由于用户的拍摄距离一般各不相同, 且难以保证脸部的绝对水平, 容易导致识别到的特征点整体偏转或者缩放, 因此, 需要对特征点数据进行旋转和归一化处理. 首先, 进行旋转操作, 以双眼中心为基础得到旋转矩阵, 可以算出旋转后的每一个特征点的坐标. 归一化的目的是为了把人脸图像裁剪成统一大小, 便于匹配各种特征的大小; 经过旋转操作后, 只需要以脸部边缘特征点为边界进行裁剪得到脸部图像, 进而把脸部宽度缩放到一个定值, 得到宽度统一的图像. 归一化操作会损失掉人脸的宽度信息, 详细阐述见第 3.1.2 小节.

3.1.2 特征抽象

在特征抽象时, 将特征分为全局和局部两部分. 其中, 全局特征描述脸部整体的大致比例, 局部特征则包括五官的大小和形状细节. 这 2 类特征相互独立, 分别对应模型库中的每一张人脸, 每输入一幅图像后也会从这两方面单独进行特征匹配.

对于全局特征, 关注整张脸的长宽比以及三庭的比例. 在上述的归一化阶段, 每张人脸已经被缩放到了预设的统一宽度 d , 重点提取人脸长度进行比对. 图 6 选取了人脸边缘的 4 个特征点 P_1 , P_2 , P_3 , P_4 , 计算得到长宽比, 即

$$w_{\text{face}} = (P_1 \cdot y - P_2 \cdot y) / d.$$

关于三庭比例, 可以先获取图像中中间竖直

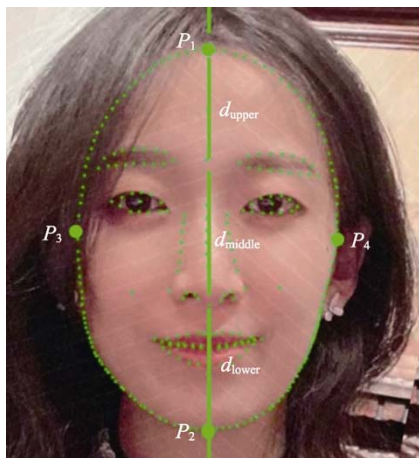


图 6 全局特征特征点

线上、中、下庭的竖直距离 d_{upper} , d_{middle} , d_{lower} , 同时除去人脸宽度 d , 得到 w_{upper} , w_{middle} , w_{lower} . 最终得到了全局特征向量, 即

$$W_{\text{overall}} = (w_{\text{face}}, w_{\text{upper}}, w_{\text{middle}}, w_{\text{lower}}).$$

局部特征包括眼部、眉毛、鼻子和嘴巴 4 个局部特征, 下面详细描述如何由特征点计算得到这些对应的特征值.

(1) 眼部特征. 宏观特征量是眼距和眼睛大小. 眼距和眼睛大小(以左眼为例)计算公式分别为

$$w_{\text{eyesdistance}} = |P_{\text{eyeL}} \cdot x - P_{\text{eyeR}} \cdot x| / d,$$

$$w_{\text{eyessize}} = |P_{\text{eyeLtop}} \cdot y - P_{\text{eyeLbottom}} \cdot y| / d.$$

其中, P_{eyeL} 和 P_{eyeR} 分别是左右眼瞳孔中心的位置; P_{eyeLtop} 和 $P_{\text{eyeLbottom}}$ 分别是左眼的最高点和最低点位置; 右眼类似. 眼部的细节特征量是建立在眼部宏观特征的基础上. 如图 7 所示, 选取双眼的中心点, 拟合出一条水平的眼部基准线, 可以得到内眼角 $P_{\text{eyesinner}}$ 和外眼角 $P_{\text{eyesouter}}$ 到基准线的距离, 用于描述眼角和外眼角的纵向偏移, 进而可以共同描述眼睛的斜度. 眼皮弧度可以近似由眼皮上的 3 个点得到, 即

$$w_{\text{eyestopbend}} = \overline{P_{\text{eyetop}} P_{\text{eyesinner}}} \cdot \overline{P_{\text{eyetop}} P_{\text{eyesouter}}}.$$

相应地, 下眼皮弧度也如此描述.



图 7 眼部特征点

(2) 眉毛特征. 图 8a 与眼部特征相似, 需要先计算宏观量眉毛宽度和眉毛间距, 确定一条经过眉心的水平基准线. 然后, 分别计算各个特征点到基准线的竖直距离, 从而可以精准地得到眉尖、眉中和眉尾的各种特征.

(3) 鼻子特征. 鼻子的特征较为简单, 可由鼻梁处的一系列特征点拟合出一条竖直基准线, 通过计算两侧特征点与此线距离的均值得到鼻翼宽度、鼻头宽度等特征值. 特别地, 由图 8c 中鼻头处的 3 点可以粗略地计算出鼻子鹰钩的程度.

(4) 嘴部特征. 嘴部的宏观特征包含了嘴部宽度和嘴唇厚度, 宽度可以由嘴角的 2 点快速得到, 建立起嘴部的水平基准线, 如图 8b 所示. 据此, 分别计算 2 个唇峰到基准线的距离和嘴唇最低点到基准线的距离, 可以计算出上下嘴唇的厚度. 细节特征同样也基于嘴部基准线来计算.

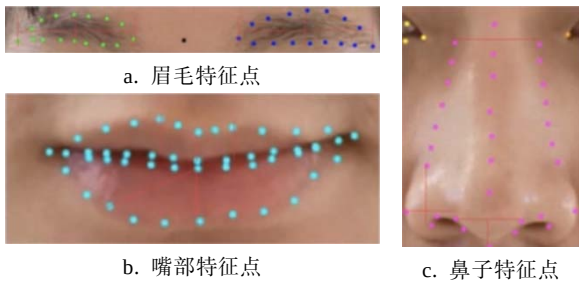


图8 其他五官特征点

3.2 模型特征库

3.2.1 基础模型制作

以亚洲平均人脸为基础, 邀请资深的三维设计师进行了卡通化的三维模型制作工作. 其中, 制作的主要依据是在第2节提出的风格设计标准和二维卡通参考. 在此模型基础上, 由一位专业的动画师基于上述的特征维度, 制作了一一对应的形变参数. 值得注意的是, 每个维度的形变系数需要保证面部模型拓扑结构足够平滑, 否则在后面的组合中面部容易出现严重毛刺或者混乱的结构.

3.2.2 特征-形变模板库

基于第3.1节描述的特征提取方法, 预先输入了人工挑选的200幅各具特征的人脸图像, 得到200组特征值. 比较输入照片的各组特征值和标准照片的特征值差异, 得到对应的差值系数, 再将差值系数映射到卡通模型的形变系数, 从而可以得到200个相应的卡通模型. 据此开发了一个工具供设计师使用, 设计师可以根据主观感受对生成的模型进行形变程度的调整, 工具会记录下调整的结果, 并以权重的方式反馈调整形变系数的极值. 以眼睛斜度为例, 如图9所示. 设计师对生成模型进行夸张调整后极值扩大为原来的120%, 在该模板下眼睛斜度的系数都需要乘上这个系数. 最终, 在保证美观和相似性的前提下, 建立一个具有200对特征值-形变参数的基础模板库. 其中, 每个模板内均包含全局特征和局部特征, 局部特征又包含4个特征向量, 彼此解耦独立, 以供后续的特征匹配使用.

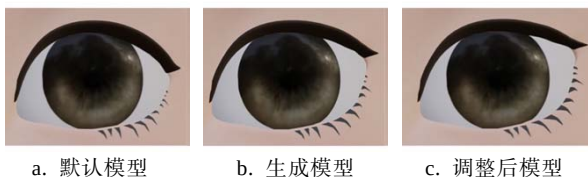


图9 眼睛斜度调整过程

3.3 模型生成

3.3.1 特征匹配与矫正

在本系统中, 风格化的卡通人脸生成包含特征匹配和形变系数调整2部分. 每一个人脸的照片均可以计算出一组对应的形变系数, 应用到最优模板上后, 即可得到最终的三维人脸. 对于每一张输入的人脸照片提取特征, 会产生1组全局特征向量 $\mathbf{W}_{\text{overall}}$ 以及4组局部特征向量 $\mathbf{W}_{\text{eyes}}$, $\mathbf{W}_{\text{eyebrows}}$, $\mathbf{W}_{\text{mouth}}$ 和 $\mathbf{W}_{\text{nose}}$, 分别以最小差值函数 E_{diff} 在模板库中寻找最相似的模板, 即 $E_{\text{diff}} = \sum_{s \in Q} \|\mathbf{w}_s - \mathbf{w}_o\|^2$.

其中, $\mathbf{w}_s$ 表示输入图像的特征值向量; 集合 Q 表示上述4个特征的集合; $\mathbf{w}_o$ 表示模板图像的特征值向量. 全局特征和局部特征分别匹配到最相似的模板后, 通过比较输入特征值和模板特征值的差异, 在对应模板的基础上进行形变系数的调整后组合, 如图10所示.

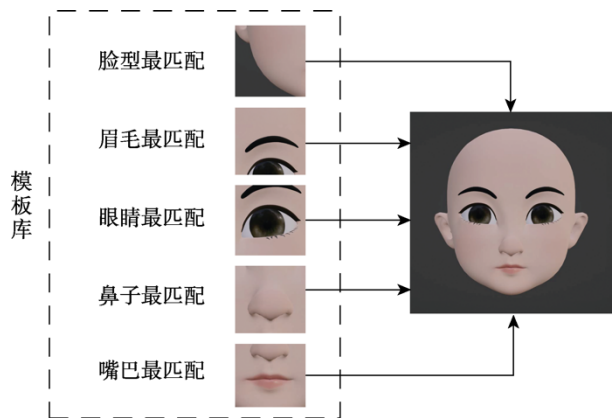


图10 模板匹配生成过程

在匹配到多个模板后, 在各个模板中逐个比较得到特征值的差异, 计算差异比例并将此比例映射到对应的形变系数上, 即

$$b_s = qb_o(w_s - w_o / w_o) + b_o.$$

其中, w_o 表示模板中的基础特征值; w_s 表示输入图像的特征值; b_o 表示基础形变系数; q 表示夸张系数.

最终, 得到完全适配输入图像的多组形变系数, 直接应用到所属模板的模型上, 即可得到相似度较高的卡通风格化三维人脸.

值得注意的是, 在多次测试中发现, 局部特征之间的边界值容易形成混乱的结果. 例如, 某个测试样本中的眉眼的距离远小于平均脸且眉毛较粗, 导致模型的下边缘下陷入眼眶之中. 对于另一个样本的鹰钩鼻则与嘴巴上沿冲突, 尝试在这些局

部特征之间添加规则和约束,虽然导致一部分人脸特征不够明显,但是保证了最终的模型被接受程度较高。

3.3.2 材质和可调参数

基于前期风格调研,本系统在人脸材质方面的处理较为简单,采用肤色识别叠加基础材质的方式生成。输入图像后直接对整张脸进行肤色识别得到一个平均色值,与平均脸比对了红色通道和明度,以得到红润程度和黑白程度不同的参数。然后,在平均脸卡通风格化模型的材质中,调整漫反射通道贴图的颜色偏移数值,便可以形成不同肤色的材质。其中,颜色差异映射的曲线采用了平滑函数并钳制了极值,以消除边缘情况带来的让人不适的结果。

3.4 人脸生成结果

本系统的虚拟人脸生成结果如图 11 所示,经过采访,用户对于生成结果满意度较高,在智能家居的环境下使用的意愿较高,能够应对在各种场合下的虚拟人脸生成需求。



图 11 虚拟人脸生成效果

在技术层面,虚拟人脸生成的模型拓扑结构规范,面数较低,如图 12 所示。因此,它可以较地进行后续使用,集成在各种智能家居设备上。

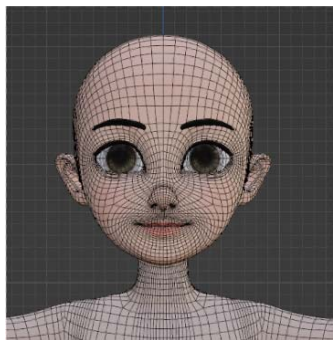


图 12 虚拟人脸模型布线

4 应用场景拓展

本文阐述的虚拟人生成方法目前已进入市场应用阶段,这在很大程度上证明了其可行性和在生成成本、生成效果等方面的优越性。借助本文虚拟人生成方法的优势,该方法可以更广泛地应用于各种智能家居场景当中。在居家健康方面,通过生成形似家庭成员的虚拟人形象,将相关健康数据、健身效果等标注在相应虚拟人上,可以得到更直观、生动的显示效果。在模拟家庭成员之外,虚拟人常以多功能的虚拟管家形式出现。在日常家居场景中,虚拟管家可以辅助家庭成员与快递员、外卖员等陌生人进行交流,在获得较高沟通效率的同时,保障家庭成员的隐私与安全。在面向老人、儿童等特殊群体,虚拟管家可以自动生成更多样化的虚拟形象,如家人、朋友和家庭教师等不同角色,以更好地实现陪伴和教育等不同功能。当前,元宇宙应用场景也在逐步拓展,此虚拟人生成方法同样可以适用于虚拟主播、虚拟偶像以及数字娱乐等场景。结合相关人工智能技术,本文方法有望在有效地降低虚拟人视频生成成本的同时,获得更生动的虚拟人形象。

5 结 语

本文探究了虚拟人在智能家居中的潜在应用场景,并据此提出了卡通风格化虚拟人脸设计方法,搭建了相应的三维模型生成系统。仅需单目摄像头即可快速生成个性化的卡通虚拟人脸;生成的人脸具有较高的接受程度,用户愿意长时间在家居或者公开场合下使用;人脸模型标准,方便进一步开发,能够满足智能家居中多样的交互界面需求。

此外,本系统存在一些不足,如前期准备工作较多,难以快速适配不同的风格需求;没有估测深度信息,难以刻画脸颊鼻梁高度等特征。未来的工作会对这 2 点做进一步研究和提升。

致谢 感谢华为移动终端 BG, UX 部门在模板数据库与生成算法中的帮助与支持。

参考文献(References):

- [1] Guerrero-Vázquez L F, Chasi-Pesantez P A, Castro-Serrano R, et al. AVATAR: implementation of a human-computer interface

- based on an intelligent virtual agent[C] //Proceedings of the IEEE Colombian Conference on Communications and Computing. Los Alamitos: IEEE Computer Society Press, 2019: 1-5
- [2] Bickmore T W, Pfeifer L M, Jack B W. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents[C] //Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2009: 1265-1274
 - [3] André E, Pelachaud C. Interacting with embodied conversational agents[M] //Chen F, Jokinen K. Speech Technology. Heidelberg: Springer, 2010: 123-149
 - [4] Bergmann K, Eyssel F, Kopp S. A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time[C] //Proceedings of the 12th International Conference on Intelligent Virtual Agents. Heidelberg: Springer, 2012: 126-138
 - [5] Cassell J. Embodied conversational interface agents[J]. Communications of the ACM, 2000, 43(4): 70-78
 - [6] Baylor A L. The design of motivational agents and avatars[J]. Educational Technology Research and Development, 2011, 59(2): 291-300
 - [7] Ali M R, Razavi S Z, Langevin R, *et al.* A virtual conversational agent for teens with autism spectrum disorder: experimental results and design lessons[C] //Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. New York: ACM Press, 2020: Article No.2
 - [8] Chattaraman V, Kwon W S, Gilbert J E. Virtual agents in retail web sites: benefits of simulated social interaction for older users[J]. Computers in Human Behavior, 2012, 28(6): 2055-2066
 - [9] Vardoulakis L P, Ring L, Barry B, *et al.* Designing relational agents as long term social companions for older adults[C] //Proceedings of the International Conference on Intelligent Virtual Agents. Heidelberg: Springer, 2012: 289-302
 - [10] Ramachandiran C R, Mahmud M M, Jomhari N. A revolutionary approach in virtual learning: user-centered kansei virtual agent[C] //Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning. New York: ACM Press, 2019: 7-11
 - [11] Baylor A L, Ashby Plant E. Pedagogical agents as social models for engineering: the influence of agent appearance on female choice[C] //Proceedings of the Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology. New York: ACM Press, 2005: 65-72
 - [12] Rosenberg-Kima R B, Baylor A L, Plant E A, *et al.* Interface agents as social models for female students: the effects of agent visual presence and appearance on female students' attitudes and beliefs[J]. Computers in Human Behavior, 2008, 24(6): 2741-2756
 - [13] Bailenson J N, Blascovich J, Guadagno R E. Self-representations in immersive virtual environments[J]. Journal of Applied Social Psychology, 2008, 38(11): 2673-2690
 - [14] Shibani Y, Schelhorn I, Jobst V, *et al.* The appearance effect: influences of virtual agent features on performance and motivation[J]. Computers in Human Behavior, 2015, 49: 5-11
 - [15] Baylor A L. Promoting motivation with virtual agents and avatars: role of visual presence and appearance[J]. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 2009, 364(1535): 3559-3565
 - [16] Fox J, Bailenson J N. Virtual self-modeling: the effects of vicarious reinforcement and identification on exercise behaviors[J]. Media Psychology, 2009, 12(1): 1-25
 - [17] Cheng Y, Qiu L Y, Pang J. Effects of avatar cuteness on users' perceptions of system errors in anthropomorphic interfaces[C] //Proceedings of the 7th International Conference on Human-Computer Interaction. Heidelberg: Springer, 2020: 322-330
 - [18] Moschoglou S, Ploumpis S, Nicolaou M A, *et al.* 3DFaceGAN: adversarial nets for 3D face representation, generation, and translation[J]. International Journal of Computer Vision, 2020, 128(10): 2534-2551
 - [19] Kemelmacher-Shlizerman I, Basri R. 3D face reconstruction from a single image using a single reference face shape[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(2): 394-405
 - [20] Zeng X X, Peng X J, Qiao Y. DF₂Net: a dense-fine-finer network for detailed 3D face reconstruction[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 2315-2324
 - [21] Shin M, Kim S J, Biocca F. The uncanny valley: no need for any further judgments when an avatar looks eerie[J]. Computers in Human Behavior, 2019, 94: 100-109
 - [22] Zhou J Y, Tong X, Liu Z C, *et al.* 3D cartoon face generation by local deformation mapping[J]. The Visual Computer, 2016, 32(6): 717-727
 - [23] Sucontphunt T. A practical approach for identity-embodied 3D artistic face modeling[J]. International Journal of Computer Games Technology, 2014, 2014: Article ID 781950
 - [24] Fujiwara T, Koshimizu H, Fujimura K, *et al.* A method for 3D face modeling and caricatured figure generation[C] //Proceedings of the IEEE International Conference on Multimedia and Expo. Los Alamitos: IEEE Computer Society Press, 2002: 137-140
 - [25] Liu S K, Wang J J, Zhang M M, *et al.* Three-dimensional cartoon facial animation based on art rules[J]. The Visual Computer, 2013, 29(11): 1135-1149
 - [26] Liu J F, Chen Y Q, Miao C Y, *et al.* Semi-supervised learning in reconstructed manifold space for 3D caricature generation[J]. Computer Graphics Forum, 2009, 28(8): 2104-2116
 - [27] Lim Y K, Fedorov A, Kim S D. 3D caricature generation system on the mobile handset using a single photograph[C] //Proceedings of the International Conference on Parallel Processing Workshop. Los Alamitos: IEEE Computer Society Press, 2007: 37
 - [28] Bruce V, Young A. Face perception[M]. London: Psychology Press, 2013