

智能数字人赋能虚拟展馆的交互体验

李丹彤 董一喆

吉林动画学院虚拟现实与电影技术学院, 吉林 长春 130000

摘要:近些年来,虚拟现实技术不断发展,同时为推动展览行业的发展,我国发布了多项行业政策,提出支持各地打造超高清、沉浸式、互动式演艺新空间和数字展览新空间。UE5引擎的应用使得各地展馆与虚拟现实技术进一步紧密、便捷地关联,给用户带来了全新的虚拟体验。文章选取UE5(Unreal Engine 5)作为实现技术的软件平台,以VR天文展馆为例,创建搭载有AI数字人导游的交互式虚拟展馆。以期待用户可以沉浸其中,并以此路线为基础,提出其他实现步骤,降低数字人导游的实现难度。

关键词: 虚幻引擎; 人工智能; 数字人; 虚拟现实

DOI: 10.64649/yh.shfzykjc.issn3078-8994.202605006

0 引言

虚拟现实(VR)技术通过计算机构建出一个可交互的虚拟世界,可给参与者带来包括视觉、听觉以及触觉等沉浸式体验,使用户身临其境,该技术融合了计算机图形学、立体显示以及人机交互技术,把沉浸感、交互性、多感知性与构想性当作核心特征,为用户模拟出多感官的体验场景。随着硬件设备不断迭代升级以及5G甚至更新一代网络迅速发展,VR技术正逐渐朝着下一代通用计算平台迈进。在这个过程中,UE5引擎的MetaHuman技术呈现出强大的应用潜力,其高精度建模能力与面部表情捕捉系统,游戏开发给予了革新,还为虚拟展馆中虚拟数字人的应用开拓了新途径,在虚拟展馆中,AI导游系统也正日益成熟,这些智能导游凭借自然流畅的语音交互、精准的图像识别以及先进的大模型语义理解能力,为参观者提供个性化导览服务。

1 关键技术分析

1.1 智能数字人的本体构建

1.1.1 高精度建模与绑定

对于符合展馆调性的虚拟导游部分,以创造高保真、可驱动的数字人形象为目的。需要考虑高精度建模与绑定、材质与光照渲染以及实时驱动技术。

对于高精度建模与绑定部分,可基于扫描建模或依靠建模师熟练技术进行数字人基础建模,再后续进行建模参数化,如设置基础尺寸、进行角度以及位置坐标的调整等。

对于材质与光照渲染部分,在不同引擎中的表现效果可能不同,但主要包含PBR材质、实时光线追踪、全局光照以及环境光遮蔽。其核心是解决大规模场景中的材质(光照)计算负载。

实时驱动技术部分,便是基于关键帧动画、动作捕捉驱动、语音驱动口型同步以实现程序

化动画或者AI驱动动画。在其中使用各类工具解决动作如何表现流畅等技术难题,这是交互可以“活”起来的基础,AI驱动便可实现更适应的非预设动作。

1.1.2 智能交互能力核心(AI驱动)

在AI驱动部分,关键技术实现的核心目标是赋予数字人感知环境、理解用户意图、自主决策和自然表达的能力。

首先需要数字人能够进行自然语言处理与理解包含语音识别、意图识别、大型语言集成等,使数字人能听懂用户的问题,进行上下文的自然对话;接着便是对知识图谱与内容的生成,进行结构化知识库的构建,可赋能数字人丰富的知识支撑,使其回答更有深度;最后是情感的计算与表达,通过情感驱动的语音合成、表情以及肢体动作的生成来使数字人能做出恰当的情感回应,通过语调、表情、肢体语言传递自身“情感状态”,大幅提升交互的亲感和代入感。

1.2 AI集成方案

系统中使用的AI技术需要进行整合为优化方案,并要对交互体验产生实际提升。在实现AI与现有技术的接口设计时,AI可以实现对于数字人表情驱动的参数化,甚至可以优化HDR环境贴图以及其它的一些功能,主要分为三个层次:输入层、智能处理引擎以及渲染与交互输出层。通过三个层次来形成环境感知—智能决策—动态呈现的增强现实闭环。

1.2.1 AI输入层

对于输入层次来讲,其思路便是对用户信号接收并将其传入下一层进行响应。在信号接收方面,应更偏向融合型的深度学习架构。

1.2.2 处理层

作为连接输入层和输出层的中间层次,要对多模态感知内容进行深度整合、理解和转化,分析用户意图,进行内部推理与生成来调整需要输出的内容,驱动智能化响应的生成。

1.2.3 输出层

输出层的技术实现路径包括通过个性化行为实现参数化建模驱动动作,通过自适应光照材质实现实时光照参数调整,以及通过场景化解说实现语音加全息投影的联动。

2 系统设计与实现

2.1 系统总体设计

实现搭载的 AI 数字人导游的虚拟展馆,其基础设计思路可分为四部分采用“端—边—云”三级协同架构,在这四部分上可选择性对其进行拓展,例如,我们选择在 UE 中对其进行实现:

终端层: UE5 渲染引擎 (UE5.6.0) 插件集成 MetaHuman 数字人,负责进行多模态交互的呈现。

边缘层: 部署轻量化 AI 模块 (语音端点检测、音频预处理、实时唇形同步),降低云端依赖。

云端层: DeepSeek-V3-0324 模型提供语义理解,百度 TTS V1.5 API 实现语音合成

数据流: 基于 HTTP/2 协议构建低延迟通道。

我们将计算机密集型任务如 LLM 推理等推至云端,通过 Varest 插件实现非阻塞式的大模型 API 调用,避免了 UE5 主线程的阻塞,实现了异步通信机制。

2.2 核心模块实现

2.2.1 语音输入与识别模块

针对虚拟展馆嘈杂、游客流量不稳定的场景,传统流式识别“全程开机”模式存在三大问题:模型常驻显存造成浪费、无效语音涌入增加并发成本、静音误判导致交互中断。我们提出两段式闸门控制与流式识别优化进行静态模型加载与动态音频过滤。可实现的优化方法例如 Runtime Speech Recognizer 插件实现端到端识别。

2.2.2 AI 大模型问答板块

系统仅在收到前端校验后的文本 prompt 时启动问答大模型推理流程:问题文本与场景 ID、语言标签被打包为 JSON 格式,通过 HTTP/2 POST 发送至云端,首包数据进行返回 {answer, tts_url, emote}。该模块通过三项设计提升展馆场景适用性:采用显式异步接口,使问答与渲染线程完全解耦,避免阻塞程序主循环。

2.2.3 语音输出板块

语音输出板块以“瞬时同步”为核心目标,将云端返回的 TTS 音频流直接映射为 StreamingSoundWave 对象,并在 Event Begin 触发时通过 FPlay Sound 节点进行零缓存播放。该设计通过时间耦合确保音频帧与数字人面部动画、字幕更新共用同一渲染 Tick,实现口型、声波与字幕的精准对齐;同时,利用空间解耦机制,流式对象的生命周期由引用计数管理,支持多路并发播放而互不干扰。此外,数据流

单向,下游仅消费上游已校验的音频流 URL,不进行反向请求或状态回写,将语音输出作为问答链路的末端闸门,将任何潜在延迟锁定在播放前一刻,从而确保语义与听觉的瞬时一致性。

2.2.4 数字人唇形同步

数字人唇形同步技术是实现自然交互体验的关键组成部分,它确保了数字人导游在提供信息时的表达自然且富有表现力。在 UE5 中,通过采用 Runtime MetaHuman Lip Sync 插件,本系统能够为 MetaHuman 角色以及自定义角色实现实时、离线和跨平台的口型同步。

2.3 AI 导游逻辑

在本研究中, AI 导游逻辑的设计目标是创建一个以用户为中心的智能化导览系统,该系统集成了语音识别、自然语言处理和语音合成等先进技术,以提升虚拟展馆的互动性和教育价值。系统的工作流程从用户通过语音输入板块发出指令开始,这些指令随后被 Runtime Speech Recognizer 插件实时转换为文本。利用 AI 问答大模型,理解用户的查询意图并生成相关回答。接着, TTS 服务将文本回答转换为自然流畅的语音,通过 UE5 的 StreamingSoundWave 对象进行播放,同时 Runtime MetaHuman Lip Sync 插件确保数字人导游的口型与语音输出精确同步,增强了交互的自然性和可信度。

2.4 与现有研究的不同及创新点

现有研究中对于搭载 AI 导游的虚拟展馆方案有两种:一种是基于 Unity 引擎的轻量化交互设计,其优点在于可向移动设备端进行移植,打造 AR 软件进行导游讲解,但是受限于渲染精度, AI 数字人在表情细节以及沉浸感上表现较弱;另外一种则是基于 UE4 开发的常规数字人,其在场景中的表现力较强,但 AI 交互部分强烈依赖本地轻量化模型,问答的准确率以及响应速度无法做到较低延迟同步。

在本研究中,为解决高保真渲染与实时响应无法同步进行的问题,我们将 MetaHuman 与语音问答大模型进行结合,通过 Varest 插件的实时调用 API 机制与 UE5 中的 Nanite 几何技术结合,使 AI 问答响应速度再次得到提升,同时 AI 数字人导游的唇形动画可以与语音保持毫秒级延迟同步。

在语音识别方面,提出“两闸式”策略,即静态模型加载与动态音频检测相结合的流式语音输入。较传统非流式输入方案,我们的语音识别可做到连续识别与无时间限制,为虚拟展馆场景下的自然对话提供技术支持。

3 应用场景与最终效果

3.1 示例场景

AI 导游作为智能化的参观辅助工具,在虚

拟展馆里的应用前景受到广泛关注,在实际应用时发现,随着虚拟展馆复杂度与特色性增加, AI 导游对于游客提出问题的回答会不精准,交互性效果也会受到一定程度影响,现阶段达成 AI 导游在虚拟展馆中的完美搭载与运行,还存在不少挑战,基于此,我们认为应在相关技术成熟或者出现响应更快、准确率更高的 AI 大模型时,再开展大规模测试。

本次实验选取了团队精心搭建的 VR 天文展馆作为测试对象,该展馆包含宇宙大爆炸、太阳系介绍、艺术介入太空、硅基生命形成等 7 个很有吸引力的板块,内容丰富且有教育意义,前期测试中,我们依靠 VR 设备做了全面交互测试,结果显示展馆各项功能均可正常运行,交互体验良好,为后续 AI 导游搭载测试奠定了基础。

本次实验中,我们依靠 AI 大模型问答板块的 Make Json 功能,给 AI 导游设置了明确身份,通过以此类对于限制身份的设置,希望 AI 导游在与游客互动中呈现专业又亲和的形象,提升游客游览体验。

3.2 实验数据

经过多轮测试,最后得到 AI 响应延迟以及唇形同步延迟数据。因 DeepSeek 调用 API 的响应速率受时间段影响,在实际应用中可使用本地部署问答大模型来解决此问题,除去本影响因素外, AI 响应延迟在 815ms 左右;唇形同步延迟在 160ms 左右,若通过时间耦合机制,理论上唇形同步与语音输出的延迟可再进一步降低至 120ms 内。

3.3 最终效果

本研究实现了 AI 导游在 VR 天文展馆中的具体应用,实验表明:经过限定的身份设定与交互流程优化, AI 导游在与游客的互动中可以展现出良好的适应性与友好性。在交互体验上, AI 导游可对指令进行快速处理,并在输出层以简洁幽默的语言回答问题,能有效激发游客交流积极性,使游览过程更具互动性。

参考文献:

- [1] 池春阳, 蒋永贵. AI 时代职业教育高质量发展的逻辑要义、范式危机与实践路径 [J]. 温州职业技术学院学报, 2025, 25 (2): 44-49.
- [2] 黄国荣, 刘炜. AI 虚拟导游在高职虚拟仿真实训基地的应用研究——以广州番禺职业技术学院为例 [J]. 现代信息科技, 2024, 8 (2): 92-95.
- [3] 赵越, 孙旭. ChatGPT 助力虚拟岩石展馆数字资源建设的实践探索 [J]. 中国现代教育装备, 2025 (2): 19-21.
- [4] 高博. VR 技术在数字虚拟展馆中的应用 [J]. HomeTheater, 2024 (2): 94-97.
- [5] 罗列异, 张帆. 沉浸式数字虚拟展馆设计研究 [J]. 中国有线电视, 2021 (8): 847-850.
- [6] 刘华锋. 基于 MetaHuman 和 UE 的数字人制作 [J]. 软件, 2024, 45 (11): 99-101.

作者简介: 李丹彤 (1997.06—), 女, 汉族, 吉林省, 硕士研究生, 助教, 研究方向: 虚拟现实技术。

在技术层面, 现阶段 MetaHuman 插件仍存在部分限制, 但通过精细化动画蓝图的设计并连接, 可以实现数字人唇形与语音输出的自然同步, 显著提升沉浸感与交互真实感。但在复杂场景中唇形同步的流畅性仍需提升。

4 结论与展望

4.1 成果总结

在本次研究中, 我们对 AI 导游在 VR 天文展馆中的应用效果进行了全面评估。实验结果表明, 经过细致的身份设置和交互流程优化, AI 导游在与游客的互动中展现出了显著的适应性和友好性。在交互体验方面, AI 导游能够迅速响应游客的语音指令, 并以简洁明了且幽默风趣的语言回答问题。这种开朗幽默的回答风格极大地激发了游客的交流欲望, 使得游客在游览过程中更加积极主动地与 AI 导游互动。

在技术层面, 尽管 MetaHuman 插件存在一定的限制, 但我们通过精细化的动画蓝图设计, 使得数字人的唇形能够与语音输出实现较为自然的同步效果。这一优化显著提升了游客的沉浸感, 增强了交互的真实感。然而, 我们也注意到, 在复杂的交互场景中, 唇形同步的流畅性仍有待进一步提高。

4.2 未来方向

在未来, AI 数字人导游的发展方向会着重于提升互动体验、拓展应用场景以及优化知识传达途径, 例如, 可在未来智能图书馆的终端搭载 AI 数字人导游。随着 AI 技术不断进步, 它的交互能力有望得到较大提升, 语言表达会变得更加自然且精准, 可为游客提供更为贴心的个性化服务, 知识库会持续充实, 从现有的文化、历史、科学领域延伸到政治、经济等更为广泛的领域, 构建起更加全面的系统的知识体系, 最后可借助跨领域知识串联来满足游客多样化的需求。