# 叙事工坊:交互式叙事场景的自动生成

朱晗希[1) ] 高可隽[1)] 陈小雨[1)] 李怡珂[2)]
覃诗睿[1)] 赵乐朋[1)] 张少魁[1)] 张松海[1)]

[1)](清华大学计算机系 北京 100084)
[2)](乔治梅森大学计算机系 费尔法克斯 22030 美国)

**摘 要** 近年来,随着交互式数字环境的快速进步,现代技术使用户能够完全沉浸在交互式叙事场景和游戏体验中。尽管静态 3D 场景构建技术已相当成熟,但人们对于叙事场景中能够增强叙事体验的需求却日益增长。本文提出了一种创新方法,通过将构成叙事情节的单个事件(故事点)巧妙地融入场景中,从而构建出引人入胜的叙事场景。本方法通过优化故事点的空间和时间布局,综合考虑视觉显著性、叙事节奏和场景布局合理性,从而构建一个动态且交互性强的场景。这不仅使用户能够更有效地探索、交互叙事场景,还极大地增强了沉浸的故事体验。

**关键词** 场景合成;叙事场景;视觉显著点;模拟退火算法;交互式 3D 体验
**中图法分类号** TP391 **DOI 号** 10.11897/SP.J.1016.2025.01517

# StoryCraft: Automatic Generation of Interactive Storytelling Scenes

ZHU Han-Xi[1)] GAO Ke-Jun[1)] CHEN Xiao-Yu[1)] LI Yi-Ke[2)]
QIN Shi-Rui[1)] ZHAO Le-Peng[1)] ZHANG Shao-Kui[1)] ZHANG Song-Hai[1)]

[1)](Department of Computer Science and Technology, Tsinghua University, Beijing 100084)
[2)](Department of Computer Science, George Mason University, Fairfax VA 22030 USA)

**Abstract** With significant and rapid advancements in interactive digital environments in recent years, modern technology now offers users the ability to fully immerse themselves in interactive narrative scenarios and gaming experiences. Although existing literature can successfully create static 3D scenes, there is a growing demand for more dynamic features that can support storytelling. In this paper, we propose an approach to incorporate the individual events which make up the plot, or "story points", into scenes, thus generating storytelling scenes. Our approach leverages an optimization-based method to create a dynamic interaction scene from these story points, while taking into account the timing of events, as well as the spatial distribution and visual saliency of the story points. This enables users to explore and engage with the generated narrative and storytelling scene effectively. By seamlessly integrating individual story points into dynamic scenes and optimizing their spatial and temporal arrangement, our approach enhances immersive storytelling experiences, offering users engaging and compelling virtual environments.

**Keywords** scene synthesis; storytelling scenes; visual saliency; simulated annealing; interactive 3D experiences

收稿日期:2024-08-27;在线发布日期:2025-03-19。本课题得到国家重点研发计划(2023YFF0905104)、国家自然科学基金(62402262,62361146854)、中国博士后科学基金资助项目(2024M751696)、国家资助博士后研究人员计划(GZB20230353)和清华大学腾讯互联网创新技术联合实验室项目资助。**朱晗希**,博士研究生,主要研究方向为三维场景生成与深度学习。E-mail:13651325353@sina.cn。**高可隽**,本科生,主要研究方向为三维场景生成。**陈小雨**,硕士研究生,主要研究方向为三维场景生成、重定向行走。**李怡珂**,博士研究生,主要研究方向为计算机视觉。**覃诗睿**,本科生,主要研究方向为三维场景生成。**赵乐朋**,本科生,主要研究方向为三维场景生成。**张少魁**(通信作者),博士,助理研究员,中国计算机学会(CCF)会员,主要研究方向为三维场景生成与交互。E-mail:shaokui@tsinghua.edu.cn。**张松海**,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为计算机图形学与虚拟现实、图像/视频处理。

# 1 引 言

近年来,随着三维场景构建领域的发展,有关场景叙事的研究呈新态势。然而,目前的研究多偏向于为三维场景生成讲述式的叙事,即用户以第三人称的方式观看场景中发生的故事。在讲述式的叙事中,用户被动接收信息,限制了叙事的效果[1]。为了打破这一局限,交互式叙事允许用户探索虚拟世界,主动触发事件,积极参与叙事过程,最终获得既互动又引人入胜的情感体验。

交互式叙事作为一种前沿叙事范式,横跨游戏、教育、健康等多元领域,彰显出广阔且极具挖掘价值的应用潜能。在此叙事模态下,玩家深度嵌入虚拟角色的冒险进程,借由一系列具备决定性意义的抉择,切实干预故事演进脉络。高度的交互性不仅赋予玩家体验以鲜明个性化烙印,契合个体独特诉求,还加深了玩家与游戏世界的情感纽带,强化叙事感染力。除此之外,互动性对于提高学生对教育性虚拟现实(VR)叙事内容的参与度至关重要[2]。例如,通过提供身临其境的叙事体验,VR 为大学生提供了一种交互式的方式来探索复杂的概念[3]。另外,文献[4]通过结合传统手工艺和现代技术,为儿童提供了一个互动性强、富有创意的学习环境。在本文中,将这种可以为用户提供交互式叙事体验的三维场景称为叙事场景。

然而,叙事场景创作是一项兼具复杂性与挑战性的任务。设计师需对环境详尽的分析,安排场景与用户交互的方式,同时确保叙事结构的连贯性。设计实践中,常常需要针对各细节施以精细化设计与建模,以在交互性和叙事性之间找到完美的平衡。设计者需要反复调整构成情节的“故事点”,并精心调整叙事节奏,以优化用户体验。相较而言,自动构建叙事场景流程可大幅缩减创作耗时,提升工作效率。

如图 1 所示,本文提出叙事工坊,根据故事剧本自动将故事点布置到一个静态的 3D 场景,从而生成一个动态的叙事场景。根据静态场景配置、叙事风格等信息,本系统将该场景中的关键元素提炼出,并转化为 ChatGPT 能理解的请求格式。随后,本系统将这个请求发送给 ChatGPT,要求它基于给定的关键字生成一个详细的故事图(第 3 节)。故事图是一个有向图:图中的节点是故事元素,其中包括交互物体、声音效果和动画等。在本文中,这些元素称为故事点;图中的边指示用户经历故事点的前后顺序。最后,在场景中安排故事点的位置,使代价函数(第 4.2 节)根据模拟退火算法(第 4.3 节)最小化。具体来说,代价函数由三个叙事场景设计原则组成:故事点视觉显著性、叙事节奏和场景合理性。结合这三个约束,本系统对故事点的位置进行优化,最终得到一个交互式的叙事场景(见图 2)。当用户探索动态场景时,每个故事点都会揭示故事的一个片段,并在用户触发故事点时提供引导,从而增强叙事探索。图 3 展示了本算法的实现细节。

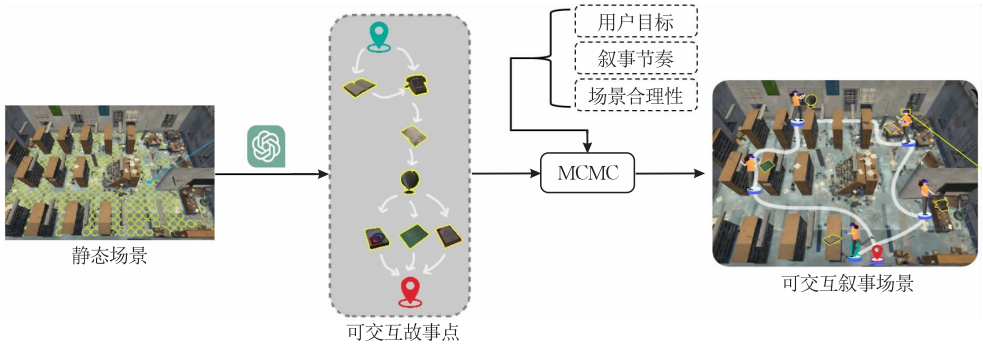


图 1 叙事工坊通过将构成情节的单个事件(故事点)放置在给定静态场景(左)中的最佳位置来创造交互式 3D 体验。根据故事的流程和进展,将这些故事点分类组成一个故事图(中),以产生一个互动的故事场景(右)


图 2 交互式叙事场景示例(当用户探索一个动态场景时,物体可能会被他/她的动作触发。每次触发显示故事的一个片段,并引导(提示)用户到达另一个故事点。用户将通过沉浸式的场景探索来体验故事)

总的来说,本文的核心贡献如下:

(1)提出了生成叙事场景的框架,基于静态三维场景配置,使用大语言模型自动地生成相应的可交互故事点和故事图,并根据叙事场景设计原则将其自动布局到场景中,从而生成一个用户可交互的动态叙事场景。

(2)根据故事叙述的部分原则量化了用户可交互叙事场景的评估规则,包括故事点的视觉显著性、叙事节奏和叙事场景布局的合理性。

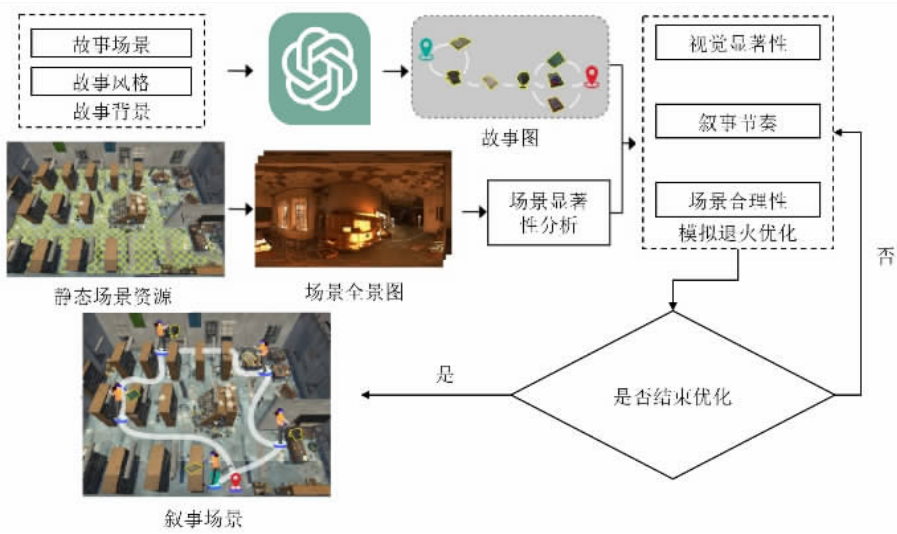


图 3　输入包含故事剧本和静态 3D 场景(输出是一个沉浸的交互式 3D 叙事场景。本文将叙事场景的四个设计原则作为优化约束,基于约束条件对场景进行优化,生成交互式叙事 3D 场景)

为验证所提出算法在生成交互叙事场景的性能和可用性,本方法与随机方法、人工方法、VirtualHome[5]进行了对比实验(第 5.2 节)。在实验中,特别关注了三个关键指标:生理数据(第 5.2.2 节)、用户主观评估(第 5.2.3 节)以及模拟器晕动症(第 5.2.4 节)。实验结果显示,本方法在叙事效果上优于随机方法和 VirtualHome,与人工方法效果相近。

## 2　相关工作

### 2.1　叙事场景

在 AR/VR 互动叙事系统方面,部分工作关注于叙事系统中非玩家角色的布局[6],也有工作为虚拟宠物生成典型行为序列[7]。MARAT[8]实现了一种基于上下文的创建复杂交互式 AR 内容的工具,使得没有编程技能的用户也能够直接在现场使用移动设备上创建交互式 AR 演示。StoryMakAR[9]将设备与虚拟角色融合在一起,从而生成叙事内容。SceneAR[10]提出了一个用于在增强现实(AR)中创建基于连续场景的微型叙事的移动应用程序。Li 等人[11]基于场景语义信息将虚拟人嵌入到 AR 场景中,增强 AR 场景的真实性,但无法对场景探索提供帮助。Puig 等人[5]实现了多智能体模拟器 VirtualHome,该模拟器可以基于用户指定的故事为虚拟场景创建叙事动画。

TNFT VR 提供了戏剧角色扮演的互动体验[12]。而文献[13]则深入探索了叙事场景中的第三人称视角。此外,文献[14]的研究着眼于利用实体道具在 VR 中讲述人物的生活故事。为了提供更真实的戏剧表演体验,文献[15]介绍了一种新系统,该系统允许演员实时驱动虚拟角色的动画。

这些研究并未关注第一视角下用户可交互场景的构建,而集中在第三人称视角的故事讲述上。另外,这些系统需要人为设计好的故事剧本作为输入,而无法基于给定的场景,自动地生成与场景匹配的故事。

相比之下,本文关注用户在第一视角下的场景叙事体验。具体而言,本方法为静态场景配置自动生成故事线,并基于故事点视觉显著性、叙事节奏和场景布局合理性[16],优化叙事场景中故事点的布局,以提升用户体验的质量。

### 2.2　场景合成

场景合成在计算机图形学领域中占据着举足轻重的地位,尤其是在室内场景合成方面[17-22]。为了优化家具布局,研究者们已经开发出基于人类工程学因素[23]和室内设计准则[24]的方法。这些方法利用真实数据训练的贝叶斯网络,通过优化算法生成一组平面布置图,从而构建出切实可行的室内布局方案[25]。

随着深度学习技术的迅猛发展,其在室内场景生成领域的应用也日益广泛。Wang 等人[26]提出了一种基于空间实例化关系图的生成模型,该模型能够高效地进行场景合成。此外,文献[27]提出了 Scene Mover 这一基于 Monte Carlo 树搜索(MCTS)的生成模型,该模型由深度强化学习网络驱动,能够指导 MCTS 的搜索过程,从而生成高质量的室内场景。

在深度卷积网络方面，文献[28]提出了一种利用深度卷积网络先验来合成室内场景的方法，该方法通过捕捉图像中的空间结构和语义信息，实现了室内场景的自动生成。文献[29]则采用了一种基于深度卷积生成模型的处理流程，该流程通过操作自上而下的基于图像的表示，并利用单独的神经网络模块预测并迭代插入对象，从而构建出逼真的室内场景。

### 2.3 场景感知分析

最近的研究在模拟人-物交互（Human-Object Interaction，HOI）结构方面取得了显著的成果。其中，基于深度学习和迁移学习的新方法被提出，用于检测人与物体之间的交互。例如，文献[30]的研究采用了图解析神经网络（GPNNs）来学习人-物交互的表示，该方法能够有效地捕捉人与物体之间的语义交互关系。另一方面，VSGNet 模型[31]作为基于图卷积神经网络的视觉模型，不仅引入了空间注意机制来关注人体和物体在图中的重要区域，还利用图卷积神经网络学习人与物体之间的语义关系。

受到 HOI 图和场景感知概念的启发，文献[32]进一步提出了场景感知（Scene Perception，SP）图的概念。这种图模型能够模拟观察者对附近物体及其相对空间关系的感知，其序列表示形式类似于计算机视觉中广泛使用的词袋模型[33]，由发生频率向量组成的统计特征。

受到这些研究的启发，本方法对 3D 场景中用户可到达的区域进行采样，并计算用户在每个采样点观察场景时的视觉注意力。本文采用文献[34]提出的基于显著性的视觉注意模型，用以表征故事点的视觉显著性。

## 3 从故事剧本到场景对象

本方法根据故事剧本自动将故事点布局到一个静态 3D 场景中，生成一个动态的叙事场景。首先，使用大语言模型根据静态场景配置生成故事剧本。静态场景配置包括叙事设置、类型以及其他叙事元素（例如，物品、情节等）。

首先为 ChatGPT 提供设定："我们期望创建一个交互的叙事场景，希望它输出可以进行用户交互的故事点和与场景相关的叙事故事"。接下来，为了确保 ChatGPT 完全理解需求，需要提供示例来帮助理解，例如"在一个废弃的酒店大厅里，接近轮椅会引发轮椅的突然移动和倒下，引起恐慌。随后，屏幕上出现一个提示，引导玩家在中央桌子上寻找一个钥匙。在找到钥匙后，玩家需要拾取，然后继续解锁并探索其他房间。"同时，为 ChatGPT 详细的说明本文希望在 Unity3D 中呈现文本提示和场景物体的方式。用户、文本提示和场景物体之间的关系分为如下两大类：

（1）用户进入某个特定的空间，将弹出提示，引导用户找到物体或前往某地。这些提示既推动情节发展，也提高了沉浸感。

（2）用户接近与场景物体关联的特定位置，即与物体关联的触发点，出现声音效果与动画，以增加用户体验。物体和用户之间的交互相互联系，其中用户是被动的接收者。当用户接近某些物体的包围盒附近时，触发物体的移动或环境的变化，例如门自动地打开或关闭、物体突然掉落、闪烁的灯光等。

最后，为 ChatGPT 输入潜在的空间和故事点。继续以废弃酒店场景为例，示例房间包括：一个大堂、一个走廊、一个房间；故事点的例子包括：轮椅、闪烁的壁灯、门、台灯。同时，允许 ChatGPT 添加自己的房间和故事点。具体而言，场景中的每个房间都可以被视为故事剧本中的不同游戏关卡。只有当用户探索并获得当前房间中的所有线索时，他们才能进入下一个房间。最终，ChatGPT 将根据给定的关键字生成故事剧本。

故事的顺序逻辑是所创造的故事场景最终呈现的决定性因素。在叙事场景设计中，设计师通过调整故事点之间的顺序关系，从而调节叙事的节奏，控制事件的流程，并创造出能够与玩家产生共鸣的连贯叙事体验。一个故事包含了一系列先后发生的事件（故事点），这些故事点依照发生的次序依次连接，可以形成一个有向故事图。故事剧本被抽象为故事图，指示用户在场景中经历各个故事点的先后顺序，其中包括叙事中的故事点的顺序结构化数据，引导用户走向有故事点的提示或线索，潜在的交互物体，以及相关的声音效果和动画。该故事图将被布局在静态场景中。

## 4 叙事场景合成

叙事场景设计是叙事背景（包括环境、物体和物体的性质等）和叙事元素（包括情节、主题、人物等）的结合[35]。本文将部分叙事原则量化为叙事场景评估规则，并基于此规则将故事图自动布局到三维场景中。

本方法使用模拟退火方法对故事点位置进行优化，代价函数分为视觉显著性、叙事节奏与场景合理性。其中视觉显著性基于故事点位置的视觉注意力得分得到，视觉注意力得分的定义与具体计算方式将在第 4.1 节中阐释；代价函数的具体实现将在第 4.2 节中阐释；模拟退火算法的计算过程将在第 4.3 节中阐释。

### 4.1　故事点的视觉注意力得分

在整个叙事过程中，故事中发生不同事件的各个地点具有不同的三维场景视觉显著性，该变量描述了用户对三维场景中周围物体的视觉感知。例如，当下一个故事点的引导事件发生在用户当前位置的视觉显著点时，用户更有可能被该事件所吸引，从而前进到该位置并触发下一个事件。事件位置表示事件发生的特定位置，并且由地图上可达区域的网络地图表示，通过地图的索引确定。

虚拟室内场景中采样点的视觉注意力方向，称为基于显著性的视觉注意力（Visual Attention，VA）方向。通过基于采样点渲染的 RGB 图像中分析场景视觉显著性结果，建模视觉注意力图（VA 图），可以确定每一个位置的视觉注意力方向。VA 图是以用户位置和场景视觉显著位置为节点的无向图，用户与场景视觉显著位置之间连有一条无向边，VA 图通过场景视觉显著性分析建模，并用于 VA 分数的计算。在 VA 分数的计算中，需要基于全景图像中的场景信息和用户的视点来评估候选位置的预期值，帮助系统在虚拟室内场景中选择最具吸引力的候选位置。在本节中，视觉显著性[34]用于评估场景的预期值。

关于视觉注意力方向图的提取，本节未采用人类-物体交互（HOI）中的物体检测器。游戏场景中高视觉显著性区域可能非实体，如光和空白区，它们是设计中的关键引导线索。光是重要视觉元素，可营造不同氛围和情感，如阳光传达温暖，昏暗光线则营造紧张。设计师利用光线引导注意，突出重点或设置情感氛围。空白区则是细节较少区域，有助于集中注意，突出其他区域，吸引玩家向特定方向前进，或通过对比使重点更突出。因此，计算场景显著性时，本节用向量表示非实体对象的视觉注意力，而非全景图像中的对象识别方法[32]。

本方法从第一人称视角对 VA 图建模。首先，对于给定采样点，通过 VR 全景相机拍摄从该点观察场景的全景 RGB 图片；然后根据该全景图进行 VA 图的建模。在全景图像中，观察者站在中心位置，通过 360°的视场角的观察和理解周围环境。VA 图包含两种类型的节点：(1) 人类节点，即用户节点，独立于场景感知检测，并且在初始化过程中是隐式的，可达区域内的每个采样点都是一个位置；(2) 视觉注意力节点，即场景感知检测的输出。每个检测到的视觉注意力显著点在 VA 图中生成一个对象节点，并通过有向边和人类节点连接。图 4 展示了基于某采样点渲染的全景图像中提取的 VA 图的一个例子。关于场景感知检测，本方法采用了文献[34]中提取显著图（Saliency Map）的方法，通过亮度、颜色和纹理三个特征提取一张全景图片的视觉显著点；其中亮度设定为输入图片的 RGB 三通道的均值，颜色特征使用该方法中扩展的四个颜色通道，纹理特征采用四个角度的定向 Gabor 滤波器得到。通过对比中心和邻域的差异计算上述特征的显著性。

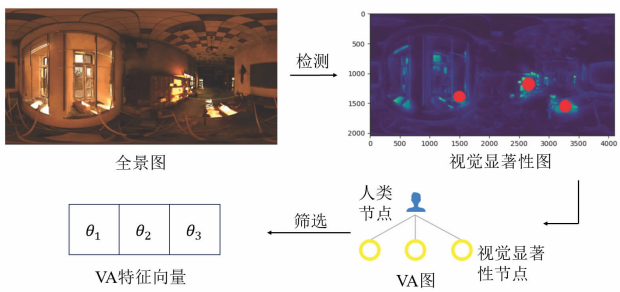


图 4　渲染场景中每个可达区域的全景图像（左上），通过视觉显著性模型[34]得到全景图片中的视觉显著点（右上），将显著性区域抽象成 VA 图（右下），并选取显著性最高的三个点的值作为后续计算的向量（左下）

对于每个采样点对应的显著性图，选择其中显著性最高的三个点的值作为后续计算的向量。即每个采样点的显著性节点被表示为 $N=\{n_i \mid i=1,2,3\}$，其中，$n_i$ 为第 $i$ 大的 VA 的点。

图像上每个点的二维坐标表示为 $n_i=(x_i,y_i)$。使用每个点相对于观察者的方向角来表示视觉注意力点在空间中相对于观察者的位置，则可将采样点的注意力向量定义为 $\boldsymbol{A}_{p_i}=\{\theta_1,\theta_2,\theta_3\}$。每个显著性节点 $N$ 与采样点之间的角度 $\theta_i$ 计算如式（1）所示：

$$\theta_i=\frac{2\pi(x_i-x_0)}{W} \tag{1}$$

式(1)中 $x_0$ 是相机在采样点拍摄全景图像样本时的初始位置，而 $W$ 是输出全景图像的像素宽度。对于给定的采样点 $p_i$ 及其在路径图上的前一个采样点 $p_{i-1}$，视觉注意力得分如式(2)所示，其中相机的初始渲染角度被定义为 $\theta_0$，两个采样点之间的相对角

度被定义为$\Delta\theta$。

$$R(p_i)=\sum\left|\frac{\Delta\theta-\theta_j}{2\pi}\right|,\ \theta_j\in A_{p_i} \tag{2}$$

$R(p_i)$表示采样点 $p_i$的 VA 得分。VA 得分是通过将每个显著节点 $N$ 在采样点 $p_i$处的方向角 $\theta_j$与相对角度$\Delta\theta$之间的差值,除以 $2\pi$ 进行归一化计算得到的。

### 4.2 代价函数

将故事图中的节点(故事点)映射到场景的平面图中,根据故事图中的有向边将节点相连,可得到关于场景的路径图。需要注意的是,路径图中故事点之间是通过线段相连的,该路径仅仅指示用户在场景中移动的大致方向,而非用户实际的移动路线。路径与场景中物体碰撞将被考虑到路径图的优化中。

路径图的代价函数,用于优化用户在叙事场景中的行动路径,该函数记作 $C_{\text{total}}(M')$。这个代价函数包括三个代价项,用以评估路径图的三个属性:场景中事件的视觉显著性、故事的叙事节奏、故事场景布局的合理性。

事件的视觉显著性:记作 $C_{\text{VA}}(M')$。场景中故事事件的视觉显著性与故事事件位置的分布是否能有效吸引用户的注意力有关。为了保持叙事的连贯性,故事事件应该分布在能够轻松吸引用户注意力的地方。

故事的叙事节奏:记作 $C_{\text{PC}}(M')$。故事的叙述遵循"引言、发展、转折和结论"的结构。在场景中体验故事事件时,每个故事事件之间应存在非故事元素加以间隔,以确保用户的心理体验在整个场景中均匀分布,避免叙事情感节奏的破坏,影响叙事效果[36]。

故事场景布局的合理性:记作 $C_{\text{AS}}(M')$。对于一个场景,物体布局的合理性十分重要,一个不合理的场景布局(比如所有物体摆放在房间中的同一个角落)削弱了虚拟场景的真实感,同时影响着用户的体验。因此,本文在故事图优化的过程中,增加了对场景布局合理性的考量。

最终的代价函数的定义如式(3)所示。

$$C_{\text{total}}(M')=w_{\text{VA}}C_{\text{VA}}(M')+w_{\text{PC}}C_{\text{PC}}(M')+w_{\text{AS}}C_{\text{AS}}(M') \tag{3}$$

其中,$w_{\text{VA}}$、$w_{\text{PC}}$和 $w_{\text{AS}}$分别代表三个约束项的权重。本文提出的算法通过最小化代价函数 $C_{\text{total}}(M')$来优化路径图 $G'$。

#### 4.2.1 用户目标

给定场景中可达区域的一个采样路径图 $M$,本文评估路径图 $M'$上每个点 $P_i$相对于前一个点 $P_{i-1}$的视觉显著性,以确保故事中的每个事件都能在整个场景中激发用户的兴趣,并促进故事的整体进展。为了评估在路径图 $M'$中设置的每个事件的视觉注意力显著性,本节定义了一个故事事件注意力显著性的代价函数,如式(4)所示。

$$C_{\text{VA}}(M')=\frac{1}{|P'|}\sum_{p_i\in P'}R(p_i) \tag{4}$$

即在故事路径图中,故事事件产生的位置在整个场景中视觉显著性越高,越可能吸引用户的视觉上的注意力,该代价函数的值越低。

#### 4.2.2 叙事节奏

叙事节奏用来描述玩家在场景体验中推进的速度。一般来说,故事节奏会遵循一系列高峰和低谷的顺序,如图 5 中所示。

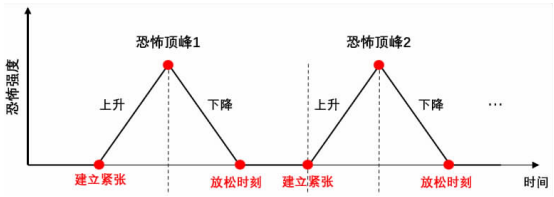


图 5 叙事节奏示例

叙事节奏应该在整个讲故事过程中相对均匀地分布。分布不均可能导致用户长时间处于恐惧或平静的状态。为了达到这种效果,故事情节对象不应集中在场景的特定区域,而应该与非情节对象混合在整个场景中。因此,两个故事情节事件之间应该有障碍物,迫使用户绕过障碍物到达下一个情节事件,如图 6 所示。叙事节奏代价函数的具体值由式(5)、(6)计算。其中,$m$ 表示场景中物体数量,$o_i$表示场景中第 $i$ 个物体,$T(o_i)$表示第 $i$ 个物体在场景平面图中的投影区域,$T_{\text{obj}}$表示场景中所有物体的投影。从故事点中随机采样得到采样点 $p_i$,$p_j$。$E'=(p_i,p_j)$表示采样点之间的直接连接。其中 $T_{\text{obj}}$表示场景中所有物体的投影。

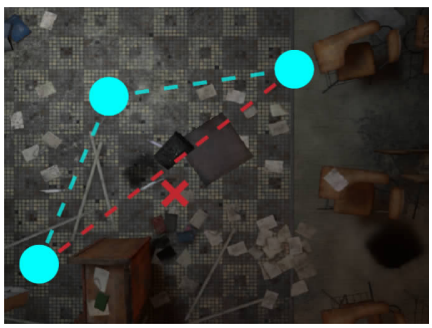

图 6 叙事点间添加障碍物示例

$$T_{\text{obj}}=T(o_1)\bigcup T(o_2)\bigcup\cdots\bigcup T(o_m) \tag{5}$$

$$C_{\text{PC}}(M')=\frac{E'-E'\bigcap T_{\text{obj}}}{E'} \tag{6}$$

物体投影与 $E'$ 相交，代表该物体在两个故事点间形成障碍，因此物体投影与 $E'$ 相交部分 $E'\bigcap T_{\text{obj}}$ 占 $E'$ 总长的比例可以准确量化故事点间的分隔程度。故事点间的分割程度越大，用户抵达下一故事点的时间越长，以达到叙事中缓解的效果，平衡叙事节奏。

4.2.3　场景布局的合理性

物体除了作为叙事的触发器，其布局还具有美学意义。因此，在优化故事图的过程中，考虑物体布局的美感也很重要。不同房间内不同物品的布局可以暗示不同的互动和活动，这为场景的叙事潜力增添了深度。在本文中，场景合理性 $C_{\text{AS}}(M')$ 由物体可达性 $C_{\text{acc}}(M')$ 和物体关联关系 $C_{\text{pair}}(M')$ 组成。一般地，场景合理性通过式(7)衡量。

$$C_{\text{AS}}(M')=\lambda C_{\text{acc}}(M')+(1-\lambda)C_{\text{pair}}(M') \tag{7}$$

物体可达性和物体关联关系通过 $\lambda$ 进行加权得到场景合理性 $C_{\text{AS}}(M')$。

物体可达性 $C_{\text{acc}}(M')$：构建用户可交互叙事场景，需要确保场景中的所有故事点对用户可达。因此，为每个物体 $o_i$ 周围都分配相应的可达空间。一般地，这个空间是一个以物体为中心的圆，其半径 $r_i$ 基于物体尺寸 $L_{o_i}$ 和人体尺寸 $H$ 的比例进行计算，如式(8)所示。

$$r_i=\frac{L_{o_i}}{H}S+L_{o_i} \tag{8}$$

其中，$L_{o_i}$ 为物体包围盒外接圆半径，$S$ 为一个定值，即当物体尺寸和人体尺寸相当时，需要为该物体留出 $S$ 的空间。物体可达性 $C_{\text{acc}}(M')$ 可通过式(9)和式(10)计算。

$$C_{\text{acc}}(M')=\frac{1}{n^2}\sum_{i=0}^{n}\sum_{j\neq i}^{n}a(o_i,o_j)obs(o_i,o_j) \tag{9}$$

$$obs(o_i,o_j)=\max\left(1-\frac{d(o_i,o_j)-L_{o_i}-L_{o_j}}{r_i-L_{o_i}},0\right) \tag{10}$$

其中 $d(o_i,o_j)$ 为 $o_i$ 和 $o_j$ 在水平平面上的距离。$obs(o_i,o_j)$ 计算了场景中物体 $o_j$ 相对于物体 $o_i$ 的阻碍关系，一般情况下，如果物体 $o_j$ 与物体 $o_i$ 的可达空间不重合，则物体 $o_j$ 相对于物体 $o_i$ 的阻碍为 0，如果物体 $o_j$ 与物体 $o_i$ 的可达空间重合，则物体 $o_j$ 相对于物体 $o_i$ 的阻碍会相应地增加。在两物体相邻时，阻碍为 1。对于一些物体来说，并不需要在每个方向都对用户可达，例如放置在墙边的书架，在这种情况下，该物体的背面不需要对用户可达。因此，本文通过 $a(o_i,o_j)$ 检测物体 $o_j$ 相对于物体 $o_i$ 需不需要可达，如果需要，则 $a(o_i,o_j)$ 返回 1，如果不需要则 $a(o_i,o_j)$ 返回 0。式(9)计算了场景中物体的平均阻碍程度，当阻碍程度低时，说明大部分物体的可达空间没有和其他物体重合，因此可达性较好。

物体关联关系 $C_{\text{pair}}(M')$：特定的物品如桌子和椅子、床和床头柜、沙发和咖啡桌，是成对根据特定的方向和距离限制进行互动的家具。确保这些物品适当地配对，可以创造一个统一的视觉构图，提升环境的整体设计质量。在本方法中，有关联关系的物体组成无序二元组 $p_{ij}=(\hat{o}_i,\hat{o}_j)$，所有二元组构成集合 $P$。在优化的过程中 $C_{\text{pair}}(M')$ 可根据式(11)计算。

$$C_{\text{pair}}(M')=\sum_{p_{ij}\in P}|d(o_i,o_j)-d(\hat{o}_i,\hat{o}_j)|+|rot(o_i,o_j)-rot(\hat{o}_i,\hat{o}_j)| \tag{11}$$

其中 $rot(o_i,o_j)$ 计算物体 $o_j$ 相对于物体 $o_i$ 的朝向。式(11)将保证优化后有关联关系的物体间相对位置与相对朝向的改变尽可能小。

### 4.3　故事点的优化

模拟退火是一种概率性优化算法，用于寻找优化问题的近似解[37]。从理论上讲，该算法确保根据每个随机调整解的接受/拒绝标准，以对数速度收敛到全局最小值[38]。模拟退火算法的灵感来源于冶金学中的退火物理过程，每个场景可以被视为被加热的材料(初始解)，然后慢慢冷却(探索邻近解)，以消除缺陷(对对象的布局进行小的修改)并最小化势能(代价函数)。

在 Metropolis-Hastings 算法中使用的 Metropolis-Hastings 标准，即马尔可夫链蒙特卡洛(MCMC)技术中采用的概率标准，被用来确定在每个温度下每个解的接受概率[39]。Metropolis-Hastings 标准的目标函数如式(12)所示。

$$f(\phi)=\mathrm{e}^{-\beta C(\phi)} \tag{12}$$

在式(12)中，$\phi=\{(p_i,\theta_i)\mid i=1,2,\cdots,n\}$ 是指场景中参与代价函数计算的物品集合，其中 $p_i$ 和 $\theta_i$ 分别代表物品的位置和方向(详见第 4.2 节)，而 $\beta$ 与温度成反比。

对于每次迭代，算法将任意选择一个物品，然后通过一个随机值将其平移或旋转，从而创建一个新的场景配置。每个提议的接受概率如式(13)所示。

$$\alpha(\phi\mid\phi')=\min\left(\frac{f(\phi')}{f(\phi)},1\right)=\min(\mathrm{e}^{\beta(C(\phi)-C(\phi'))},1) \tag{13}$$

故事点优化过程的具体算法如算法 1 所示。

**算法 1.** 基于模拟退火的故事点优化算法

输入：初始场景物体集合 $\phi_0=\{(p_i,\theta_i)\}$

输出：优化后场景物体集合 $\phi'=\{(p_i',\theta_i')\}$

1. 读入场景，令 $\phi'=\phi_0$. 设置初始温度 $T_0$ 与迭代轮次 $R$. 计算当前状态 $C_{\text{total}}(\phi_0)$.

2. 随机挑选场景中的一个物体. 随机选择平移或旋转. 得到新的场景集合 $\phi_1$. 计算当前状态 $C_{\text{total}}(\phi_1)$.

3. 根据式(13)的接受概率决定当前状态是否被接受. 若被接受，则 $\phi'=\phi_1$. 若不接受，则 $\phi'$不变.

4. $T=T_{\text{new}}$. 若到达迭代轮次 $R$，则算法结束；否则，跳至步 2.

## 5 实验与结果

为验证本方法在实际应用中的可行性和效果，我们设计了如下用户实验。实验旨在比较本方法生成的场景、专业设计的场景以及 VirtualHome[5] 生成的场景在叙事体验上的差异。

本实验分为两个不同的部分。在实验 1 中，使用恐怖叙事场景进行实验。本方法之前，并未有研究关注第一人称视角下叙事场景生成。因此在实验 1 中，比较本方法、专业设计、随机布局三种故事点布局方法的生成效果。相比于日常场景，恐怖场景能为用户带来更加明显的生理数据变化，便于量化场景带给用户的叙事体验，因此选用恐怖叙事场景进行实验，用以测量客观生理数据的变化。VirtualHome 是当前最先进的基于居家场景的第三人称叙事生成方法，由于此前并未有工作对第一视角下的叙事场景进行研究，因此实验 2 使用 VirtualHome 作为对照组。为对比本方法与 VirtualHome，实验 2 使用日常居家场景进行实验，在给定故事剧本的情况下比较两种方法所生成的故事中剧本主人公动线上的差异，量化两种方法在叙事体验上的差别。

在实验 1 中，用户将先后随机实验三个场景——一个是由本方法生成的，一个是专业游戏设计师的原创作品，另一个是通过将故事点随机布局在房间中生成的。鉴于部分用户可能缺乏电脑游戏实际操作经验，实验采用预先录制好的第一人称视角短片展示场景①，以直观易懂的方式帮助用户感知和理解场景细节。在实验 2 中，用户将先后观看 VirtualHome 生成的居家场景叙事视频和通过本方法对其中的故事点位置进行优化后生成的居家场景叙事视频。实验后，用户需要在所给的两个场景中进行对比，反馈他们最直观的感受。实验用场景之一如图 7 所示。

图 7 一个废弃宅邸(该场景是本文用户实验使用的场景之一，其中故事点包括灯光氛围改变、门自动打开、僵尸出现等)

### 5.1 用户实验设计

在恐怖场景实验中，一共构建了九个不同的场景——其中三个由游戏设计者搭建(手工)，三个通过本方法生成(叙事工坊)，另外三个则是随机生成的(随机)。叙事工坊场景将被用于实验 1 和实验 2；手工场景只用于实验 1；随机场景只用于实验 2。所有场景均通过 Unity3D 游戏引擎和 C# 编程语言生成；而视频则是由一台配备 64 位 Windows 11 家庭版操作系统、16 GB 内存、20 核 5.0 GHz Intel i9-12900H 处理器，英伟达 GeForce RTX 3060 显卡的计算机录制和播放。在实际实现中，部分参数的取值如下：全景图像素宽度 $W$ 为 4096、人体尺寸 $H$ 为 1.7、$S$ 为 0.5、$w_{\text{VA}}$ 为 0.01、$w_{\text{PC}}$ 为 0.01、$w_{\text{AS}}$ 为 2、$\lambda$ 为 0.5。在优化故事点算法的过程中，$T_0$ 设置为 1000，$R$ 设置为 1000。在迭代轮次不大于 400 时，$T_{\text{new}}$ 为 1000；迭代轮次不大于 800 时，$T_{\text{new}}$ 为 100；迭代轮次大于 800 时，$T_{\text{new}}$ 为 10。

观看每段视频前，用户需要静息 3 min。然后，用户需要填写模拟器疾病问卷(SSQ)[40]，旨在记录并评估游戏设计对用户可能产生的模拟器疾病的影响及程度。接下来，用户在右手食指上佩戴指夹式血氧饱和仪，用于实时监测用户的心率变化；在右手大臂上佩戴血压计。观看完一个实验视频后，用户将摘下指夹式血氧饱和仪，并再次填写 SSQ 以记录当下的感受。在整个实验结束后，用户将对刚才的两个场景的恐怖程度和真实性进行 1 到 10 的评价。

在居家场景实验中，我们对三个不同场景生成了相应的叙事视频。每个场景根据同一个故事剧本分别生成两段第一人称视角下的叙事视频，其中一段使用 VirtualHome 方法生成，另外一段使用本方法对场景中故事点布局进行优化从而生成。用户需要观看这两段视频，并对其中主人公的动线进行

① https://cloud.tsinghua.edu.cn/d/fa16cb60a08a49f192d7/

评价。

整个实验过程对于每位用户而言，大约需要30min的时间。

### 5.2　实验指标与结果

#### 5.2.1　用户信息

本实验共招募了 24 名被试者，其中 16 名被试者为在校学生，8 名被试者已经参加工作。

#### 5.2.2　生理数据

恐惧会引起多种生理表现的变化，包括心率变化、血压变化、面部表情抽动、呼吸频率、皮肤电导率变化等。然而大部分生理表现不易测量或量化，因此选择心率和血压作为验证叙事效果的生理指标。

心率的变化经常用于评估恐惧对个体心理和生理状态的影响[41]。然而，个体差异导致恐惧时心率反应不同，一些人恐惧时心率加快[42]，因“战斗或逃跑”反应释放肾上腺素，使心率加快①。反之，有一些人在恐惧时会使副交感神经活动增强，从而使心率显著下降[43]。因此，对于心率，先记录用户静息心率，分析观看视频时最高和最低心率，计算与静息心率差的绝对值，取较大者量化恐惧程度。

对于血压，分别测量用户观看视频前、观看视频时、观看视频后的血压。

实验 1 的结果如表 1 所示。由表 1 可以看出，被试在观看本方法所生成场景的漫游视频时，心率波动明显高于观看随机方法所生成场景的漫游视频，略低于观看手工搭建的场景的漫游视频。同时，被试在观看本方法和手工方法所生成场景的漫游视频时，收缩压和舒张压均升高，而在观看随机方法生成的场景漫游视频时，收缩压和舒张压均降低。

**表 1　实验 1 心率、血压波动结果**

| 场景类型 | 心率变化（次/min） | 收缩压变化（mmHg） | 舒张压变化（mmHg） |
|---|---|---|---|
| 手工 | 12.67 | 0.00 | 1.40 |
| 叙事工坊 | 11.78 | 1.40 | 0.30 |
| 随机 | 10.00 | −0.70 | −1.60 |

可以认为，本方法所生成的场景相较于随机方法可以为用户提供更加紧张刺激的叙事体验，同时，本方法所生成的场景与设计师搭建的场景之间略有差距，但差距不大；可以认为，本方法所生成的场景能够达到与手工方式相近的效果。

#### 5.2.3　用户评价

在实验 1 中，用户需要对所体验两种场景的恐怖程度和真实性打分。在实验 2 中，用户需要对叙事视频中主人公动线的合理性进行评价。实验 1 中，用户对于手工、本方法、随机布局所生成场景的恐怖程度和真实性的反馈如表 2 所示。实验 2 中，用户对于 VirtualHome、本方法所生成的叙事视频的反馈如表 3 所示。表格中的每个单元展示了评分的均值与方差。

**表 2　实验 1 用户评价结果**

| 场景类型 | 恐怖程度 | 真实性 |
|---|---|---|
| 手工 | 6.33(3.51) | 6.67(2.24) |
| 叙事工坊 | 6.50(2.45) | 6.75(1.48) |
| 随机 | 4.33(5.07) | 6.17(3.77) |

**表 3　实验 2 用户评价结果**

| 方法 | 叙事工坊 | VirtualHome |
|---|---|---|
| 动线合理性 | 3.8(0.62) | 3.4(1.60) |

由表 2 可看出，手工和本方法所生成场景之间的得分差距不大，即本方法所生成的场景能够达到与手工方式相近的效果。同时，本方法所生成的场景在恐怖程度与真实性的方差均低于手工生成的场景，这表明本方法所生成的场景在恐怖程度与真实性方面均具有更小的离散性，即生成的场景在恐怖程度和真实性上更为一致。本方法所生成的场景明显优于随机生成的场景。因此可以认为，本方法所生成的场景明显优于随机生成的效果，说明本方法在叙事场景构建中的有效性。

由表 3 可以看出，在动线合理性的评估上，本方法优于 VirtualHome 方法，说明在故事点的安排上，本方法比 VirtualHome 更为合理。

综上所述，通过对比手工、本方法、随机布局方法在“恐怖程度”和“真实性”两个维度上的表现，可以发现本方法在生成具有恐怖氛围和真实感的场景时，明显优于随机生成的场景，并且其效果与手工场景相当，具有更好的一致性和更小的离散性。在故事点的安排上，本方法比 VirtualHome 更为合理。这表明论文所提出的本方法具有较高的实用性和可靠性，在未来的场景生成研究中具有一定的参考价值。

#### 5.2.4　模拟器晕动症

在三维场景中，模拟器晕动症是视觉与身体感知的不匹配而引发的眩晕症状，严重影响玩家的体验。该眩晕表现为恶心、头痛以及方向迷失等症状。为了科学地评估模拟器晕动症的程度，研究者们常常采用模拟器晕动症问卷[40]（Simulator Sickness Questionnaire，SSQ）作为工具。该问卷共 16 个问

① https://www.ncbi.nlm.nih.gov/books/NBK541120/

题,全面覆盖了从恶心、眼部不适到方向迷失,再到整体的网络晕动症等各个方面的症状。参与者需根据自身的实际感受,在 0(无明显症状)到 3(严重)的范围内,为每一项症状打分,以量化他们的不适程度。

根据用户在每个视频前后填写的 SSQ 所收集的结果,计算前后的 SSQ 总分(Total Score,TS)以及前后的 SSQ TS 差。TS 差的结果如表 4 所示。

**表 4 实验 1 用户评价结果**

| 场景类型 | 场景 1 | 场景 2 | 场景 3 |
|---|---|---|---|
| 手工 | 20.57 | 10.29 | 17.77 |
| 叙事工坊 | 24.31 | 4.68 | 19.64 |

虽然无法得出长时间体验下用户的眩晕情况,但是上述数据可以验证本方法生成的叙事场景在短时体验中不但不会造成额外的眩晕,还可能较手工场景造成更少的眩晕。

## 6 结 语

本文提出了一种通过将交互式叙事元素融入静态三维场景中来自动生成叙事场景的方法。此方法使用大语言模型为预布局的静态 3D 场景生成故事剧本,并自动确定场景中故事点的最佳放置位置。在布局阶段,根据故事叙述的部分原则量化了用户可交互叙事场景的评估规则,包括故事点的视觉显著性、叙事节奏和叙事场景布局的合理性。通过综合这些要素,本方法能够生成具有良好叙事体验的叙事场景。实验结果展示了本方法的有效性。生成的场景不仅满足了叙事要求,而且在视觉效果和用户体验上与专业设计师创作的场景相媲美。这些结果证明了本方法的实用性。总体而言,本文节省了手动构建交互式叙事场景所需的时间和人力成本,并且保证了较高质量的场景构建效果。

虽然本文在场景设计方面的考量已经可以适用于大部分叙述性场景,但本文所述方法难以覆盖全部场景,例如对于空间结构复杂的场景进行优化或对于同一场景的故事点进行多样化的布局。复杂化和多样化叙事场景的是未来工作中可以进行探究的方向之一。同时,虽然本文算法在故事点布局方面可以与人工布局相媲美,但故事内容创作上的能力有待提升,这是由于单纯使用大语言模型生成的故事剧本与专业创作者之间仍有差距。若想要得到更好的叙事内容,仍需要人与 AI 的协同创作。

## 参 考 文 献


[1] Zeman N B. Storytelling for Interactive Digital Media and Video Games. Massachusetts, USA: AK Peters/CRC Press, 2017

[2] Zhang Lei , Bowman D A, Jones C N. Exploring effects of interactivity on learning with interactive storytelling in immersive virtual reality//Proceedings of the 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games). Vienna, Austria, 2019: 1-8

[3] Zhang Lei, Bowman D A, Jones C N. Enabling immunology learning in virtual reality through storytelling and interactivity //Proceedings of the 21st HCI International Conference (HCII 2019). Orlando, USA, 2019: 410-425

[4] Song Y, Yang C, Gai W, et al. A new storytelling genre: Combining handicraft elements and storytelling via mixed reality technology. The Visual Computer, 2020, 36: 2079-2090

[5] Puig X, Ra K, Boben M, et al. VirtualHome: Simulating household activities via programs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8494-8502

[6] Lang Y, Liang W, Yu L F. Virtual agent positioning driven by scene semantics in mixed reality//Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Osaka, Japan, 2019: 767-775

[7] Liang W, Yu X, Alghofaili R, et al. Scene-aware behavior synthesis for virtual pets in mixed reality//Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Yokohama, Japan, 2021: 1-12

[8] Rumiński D, Walczak K. Creation of interactive AR content on mobile devices//Proceedings of the Business Information Systems Workshops: BIS 2013 International Workshops. Poznań, Poland, 2013: 258-269

[9] Glenn T, Ipsita A, Carithers C, et al. StoryMakAR: Bringing stories to life with an augmented reality & physical prototyping toolkit for youth//Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, USA, 2020: 1-14

[10] Chen M, Monroy-Hernández A, Sra M. SceneAR: Scene-based micro narratives for sharing and remixing in augmented reality//Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Bari, Italy, 2021: 294-303

[11] Li C, Li W, Huang H, Yu L F. Interactive augmented reality storytelling guided by scene semantics. ACM Transactions on Graphics (TOG), 2022, 41(4), Article No. 91: 1-5

[12] Gupta S, Tanenbaum T J, Muralikumar M D, Marathe A S. Investigating roleplaying and identity transformation in a virtual reality narrative experience//Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, USA, 2020: 1-13

[13] Plaxen B, Qi Z, Schoeller M R, You S. One of the family: An exploratory 3rd person branching narrative for virtual reality//Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal, Canada, 2018: 1-4

[14] Leo C J, Tsai E, Yoon A, et al. An ant's life: Storytelling in virtual reality//Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play. London, UK, 2015: 779-782

[15] Wu Q, Boulanger P, Kazakevich M, Taylor R. A real-time performance system for virtual theater//Proceedings of the 2010 ACM Workshop on Surreal Media and Virtual Cloning. Firenze, Italy, 2010: 3-8

[16] Riedl M O, Bulitko V. Interactive narrative: An intelligent systems approach. AI Magazine, 2013, 34(1): 67-67

[17] Zhang S H, Zhang S K, Liang Y, Hall P. A survey of 3D indoor scene synthesis. Journal of Computer Science and Technology, 2019, 34: 594-608

[18] Zhang S H, Zhang S K, Xie W Y, et al. Fast 3D indoor scene synthesis by learning spatial relation priors of objects. IEEE Transactions on Visualization and Computer Graphics, 2021, 28(9): 3082-3092

[19] Zhang S K, Li Y X, He Y, et al. MageAdd: Real-time interaction simulation for scene synthesis//Proceedings of the 29th ACM International Conference on Multimedia. Chengdu, China, 2021: 965-973

[20] Zhang S K, Liu J H, Li Y, et al. Automatic generation of commercial scenes//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada, 2023: 1137-1147

[21] Zhang S K, Tam H, Li Y, et al. SceneDirector: Interactive scene synthesis by simultaneously editing multiple objects in real-time. IEEE Transactions on Visualization and Computer Graphics, 2023, 30(8): 4558-4569

[22] Zhang S K, Xie W Y, Zhang S H. Geometry-based layout generation with hyper-relations among objects. Graphical Models, 2021, 116: 101104

[23] Yu L F, Yeung S K, Tang C K, et al. Make it home: Automatic optimization of furniture arrangement. ACM Transactions on Graphics (TOG), 2011, 30(4), Article No. 86: 1-12

[24] Merrell P, Schkufza E, Li Z, et al. Interactive furniture layout using interior design guidelines. ACM Transactions on Graphics (TOG), 2011, 30(4), Article No. 87: 1-10

[25] Merrell P, Schkufza E, Koltun V. Computer-generated residential building layouts//Proceedings of the ACM SIGGRAPH Asia 2010. Seoul, Republic of Korea, 2010: 1-12

[26] Wang K, Lin Y A, Weissmann B, et al. PlanIT: Planning and instantiating indoor scenes with relation graph and spatial prior networks. ACM Transactions on Graphics (TOG), 2019, 38(4), Article No. 134: 1-5

[27] Wang H, Liang W, Yu L F. Scene mover: Automatic move planning for scene arrangement by deep reinforcement learning. ACM Transactions on Graphics (TOG), 2020, 39(6), Article No. 233: 1-5

[28] Ritchie D, Wang K, Lin Y A. Fast and flexible indoor scene synthesis via deep convolutional generative models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA, 2019: 6182-6190

[29] Zhang Z, Yang Z, Ma C, et al. Deep generative modeling for scene synthesis via hybrid representations. ACM Transactions on Graphics (TOG), 2020, 39(2), Article No. 17: 1-21

[30] Li Y L, Zhou S, Huang X, et al. Transferable interactiveness knowledge for human-object interaction detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA, 2019: 3585-3594

[31] Ulutan O, Iftekhar A S, Manjunath B S. VSGNet: Spatial attention network for detecting human object interactions using graph convolutions//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 13617-13626

[32] Li C, Huang H, Lien J M, Yu L F. Synthesizing scene-aware virtual reality teleport graphs. ACM Transactions on Graphics (TOG), 2021, 40(6), Article No. 229: 1-5

[33] Jamalpur B, Sudheer K K. Implementation of BoVW model towards obtaining discriminative features of the images. International Journal of Advanced Science and Technology, 2019, 28(17): 205-213

[34] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259

[35] Jenkins H. Game design as narrative architecture. Computer, 2002, 44(3): 118-130

[36] David I. The emotional rhythm of storytelling. Journal of Screenwriting, 2014, 5(1): 47-57

[37] Suman B, Kumar P. A survey of simulated annealing as a tool for single and multi-objective optimization. Journal of the Operational Research Society, 2006, 57(10): 1143-1160

[38] Cardoso M F, Salcedo R L, de Azevedo S F, Barbosa D. A simulated annealing approach to the solution of MINLP problems. Computers & Chemical Engineering, 1997, 21(12): 1349-1364

[39] Madrigal-Cianci J P, Nobile F, Tempone R. Analysis of a class of multilevel Markov chain Monte Carlo algorithms based on independent Metropolis-Hastings. SIAM/ASA Journal on Uncertainty Quantification, 2023, 11(1): 91-138

[40] Kennedy R S, Lane N E, Berbaum K S, Lilienthal M G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The International Journal of Aviation Psychology, 1993, 3(3): 203-220

[41] Rogers R W. Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection

motivation. Social Psychology: A Sourcebook. New York, USA: Guilford, 1983: 153-176

[42] Cacioppo J T, Klein D J, Berntson G G, Hatfield E. The Psychophysiology of Emotion. New York, USA: Guilford, 1993

[43] Rainville P, Bechara A, Naqvi N, Damasio A R. Basic emotions are associated with distinct patterns of cardiorespiratory activity. International Journal of Psychophysiology, 2006, 61(1): 5-18

**ZHU Han-Xi**, Ph. D. candidate. His research interests include 3D scene synthesis and deep learning.

**GAO Ke-Jun**, undergraduate. His research interest is 3D scene synthesis.

**CHEN Xiao-Yu**, M. S. candidate. Her research interests include 3D scene synthesis and RDW.

**LI Yi-Ke**, Ph. D. candidate. Her research interest is computer vision.

**QIN Shi-Rui**, undergraduate. His research interest is 3D scene synthesis.

**ZHAO Le-Peng**, undergraduate. His research interest is 3D scene synthesis.

**ZHANG Shao-Kui**, Ph. D., assistant researcher. His research interests include 3D scene synthesis and interaction.

**ZHANG Song-Hai**, Ph. D., associate professor. His research interests include computer graphics and virtual reality, image/video processing.

**Background**

The study presents research within the domain of interactive 3D scene synthesis, specifically focusing on the automatic generation of narrative scenes that enhance storytelling experiences in 3D spaces. This issue is of significant interest in the fields of computer graphics, human-computer interaction, and artificial intelligence, as it intersects the creation of immersive virtual reality (VR) and gaming experiences.

Internationally, there has been considerable progress in developing systems that can generate static 3D scenes. However, the dynamic integration of narrative elements that respond to user interactions and enhance the storytelling aspect is an area that is still under active development. Researchers are exploring various techniques, including the use of machine learning, natural language processing, and optimization algorithms to create more engaging and interactive narratives.

This research contributes to the field by proposing an innovative approach that leverages optimization-based methods to integrate "story points" into a scene. The method takes into account the timing of events, spatial distribution, and visual saliency to create a dynamic and interactive narrative scene from a static scene. The authors have developed an algorithm that uses simulated annealing to optimize the spatial and temporal arrangement of story points, aiming to generate scenes that are as compelling as those designed by human creators.

This work was supported by the National Key Research and Development Program of China (No. 2023YFF0905104), the National Natural Science Foundation of China (Nos. 62402262, 62361146854), the China Postdoctoral Science Foundation (No. 2024M751696), the Postdoctoral Fellowship Program of CPSF (No. GZB2023-0353), and the Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

The implications of this work extend beyond the academic domain. If successful, the outcomes could support major initiatives related to the advancement of VR and interactive media, aligning with projects that aim to innovate and expand the capabilities of digital storytelling.