

中图法分类号: TP391.9 文献标识码: A 文章编号: 1006-8961(2024)09-2471-23

论文引用格式: Yue L, Tan H, Huang J K and Zhang S K. 2024. Review of digital 3D indoor scene synthesis. Journal of Image and Graphics, 29(09):2471-2493(岳亮, 谈皓, 黄俊凯, 张少魁. 2024. 数字室内三维场景构建综述. 中国图象图形学报, 29(09):2471-2493)[DOI:10.11834/jig.230712]

数字室内三维场景构建综述

岳亮, 谈皓, 黄俊凯, 张少魁*

清华大学计算机系, 北京 100084

摘要: 在计算机图形学发展过程中, 数字三维场景长期对于学术界和工业界都起着至关重要的作用。其作用体现在展示渲染效果、支持应用环境以及充当交互载体等多个方面。然而, 三维场景本身作为一种数据形式, 其结构复杂且没有统一的数据结构, 故难以像诸如图像、文本数据集一样被大量获取与应用。已有一些工作尝试让计算机自动构建场景或是让计算机辅助构建场景。而在众多场景之中, 室内场景尤为重要。本文总结归纳了这些数字室内三维场景构建工作。提出从3个方面调研和总结场景构建的主要工作: 场景自动构建、基于用户交互的场景辅助构建以及基于多通道与丰富输入的场景构建。自动构建在于让计算机直接基于当前三维内容给出场景构建结果; 交互构建在于让用户控制计算机并辅助构建场景; 多通道构建在于让构建的场景参考输入的图像、文本和点云。最后, 本文总结了科研工作中的应用场景和关键技术, 并介绍了一些其他应用情景与未来面对的诸多挑战。数字室内三维场景构建的发展前景十分广阔, 随着新算法的不断提出以及三维场景数据集日益完善, 数字室内三维场景构建领域也将持续发展。

关键词: 室内场景; 三维场景构建; 三维场景交互; 三维场景智能编辑; 计算机图形学

Review of digital 3D indoor scene synthesis

Yue Liang, Tan Hao, Huang Junkai, Zhang Shaokui*

Department of Computer Science, Tsinghua University, Beijing 100084, China

Abstract: During continuous development of computer graphics and human-computer interaction, digital three-dimensional (3D) scenes have played a vital role in the academy and industry. 3D scenes show graphical rendering results, supply the environment for applications, and provide a foundation for interaction. Despite being common occurrences, indoor scenes are important. To increase players' gaming experience, indoor game designers require all kinds of aesthetic digital 3D scenes. In online scene decoration, designers also need to predesign the decoration and furniture layout preview by interacting with 3D scenes. In studies of virtual reality, we can synthesize virtual space from a digital 3D scene, such as the synthesis of training data for wheelchair users. However, a number of difficulties still need to be overcome to obtain ideal digital 3D scenes for the applications mentioned above. First, manually synthesized 3D scenes are usually time consuming and require considerable experience. Designers must add objects to a scene and adjust their location

收稿日期: 2023-10-08; 修回日期: 2024-04-01; 预印本日期: 2024-04-08

* 通信作者: 张少魁 shaokui@tsinghua.edu.cn

基金项目: 国家重点研发计划资助(2023YFF0905104); 国家自然科学基金项目(62132012, 62361146854); 北京市重大科技专项项目(Z221100007722001); 清华大学腾讯互联网创新技术联合实验室项目; 清华大学“水木学者”计划项目(2023SM061); 国家资助博士后研究人员计划项目(GZB20230353); 北京市科协2024-2026年度青年人才托举工程项目(BYESS2024242)

Supported by: National Key R&D Program of China(2023YFF0905104); National Natural Science Foundation of China(62132012, 62361146854)

and orientation one by one. These trivialities but heavy works cause difficulty in focusing on core ideas. Second, digital 3D scene is a data structure with extremely complex structure, and no unified consensus has been given to its data structure. Thus, digital 3D scenes are difficult to obtain and apply in large quantities compared with traditional data structures, such as image, audio or text. To solve the problems mentioned above, some existing work attempted to allow computers to automatically synthesize 3D scenes or interactively help synthesize scenes. This survey summarizes these works. This survey also investigates and summarizes 3D digital scene synthesis methodologies from three aspects: automatic scene synthesis, scene synthesis with multichannel and rich input, and interactive scene synthesis. The automatic synthesis allows the computer to directly build an indoor layout based on few inputs, such as the contour of the room or the list of objects. Initially, the scene is synthesized by manually setting rules and applying optimizers in an attempt to satisfy these rules. However, the situation increases in complexity during the synthesis practice, and thus, listing all the rules becomes impossible. As the amount of digital indoor scene increases, more works are introducing machine learning methods to study priors from the digital scenes of the 3D indoor scene dataset. Most of these works organize the furniture with graph to apply algorithms on the graph to process with the information. The results outperform those of former works. Researchers have been applying deep learning (DL) technology, such as convolutional neural network and generative adversarial network, to indoor scene synthesis, which strongly improves the synthetic effect. The synthesis with multichannel and rich input aims to synthesize a digital indoor 3D scene with unformatted information, such as image, text, RGBD scan, point cloud, etc. These algorithms enable the convenient formation of digital copies of scenes in the real world because they are mainly recorded by photos or literal description. Compared with the works on automatic synthesis, the scene syntheses with multichannel and rich input do not require diversity or aesthetics. However, this type of synthesis needs an algorithm for the accurate reconstruction of the indoor scene in the digital world. The interactive synthesis aims to let users control the process of computer-aided scene synthesis. The related works can mainly be divided into two parts: active and passive interactive syntheses. Active interactive synthesis simultaneously provides designers with suggestions while they synthesizing a scene. If the scene syntheses program can analyze the designers' interaction and recommend the options with higher possibility to be chosen, considerable workload can be saved. During passive interactive synthesis, the system learns the user's personal preferences from aspects, such as their behavior trajectory, personal abilities, work habits, and height information and automatically synthesize scenes that match the user's preferences as much as possible. Eventually, this survey will also summarize the application scenario and core technology of the papers and introduce other typical application scenarios and future challenges. We summarized and classified the recent studies on applications of digital 3D scene synthesis to form this survey. Digital 3D indoor scene synthesis has attained great progress and has a wide prospect. The automatic scene synthesis has generally achieved its goal, and more attention should be focused on the proposal and resolution of sub-problems and related issues afterward. For scene synthesis with rich input, existing work has explored inputs, such as image, RGBD-scan, text, and sketches. In the future, more potential input forms, such as music and symbols, should be explored. For scene interactive synthesis, current interactions are still limited to mouse and keyboard inputs, and methods based on interactive scenes, such as virtual reality, augmented reality, and hybrid reality, still need to be explored. Scene synthesis algorithm has continuously broadened its application. Industries normally require the automatic synthesis of a large amount of indoor scenes. The synthetic efficiency can be strongly increased if a computer can provide suggestions regarding an object and its layout. In academic studies, 3D scenes are usually applied to form all kinds of dataset. By rendering a scene's photos from various perspectives and channels, researchers can easily obtain images. However, the study on indoor scene synthesis is still facing a number of limitations. The dissimilarity of data structure causes difficulty in extending the work of others. Copyright issues prevent a scene dataset from being freely used by researchers and coders. In the future, indoor scene datasets with additional furniture model and room contour will serve as the basis of indoor scene synthesis studies. Numerous related fields, such as style consistency and automatic photography, are also showing progress.

Key words: indoor scene; 3D scene synthesis; 3D scene interaction; 3D scene intelligent editing; computer graphics

0 引言

随着计算机、多媒体和互联网等技术的飞速发展,三维场景的应用与需求日益增长。为了增加玩家的游戏体验,制造商对于三维场景的需求也在不断增加;家装设计中,设计师也需要通过与三维场景交互,预先设计出装修与家具布局预览;在虚拟现实领域,同样可用不同的三维场景构建虚拟空间的样例,如构建轮椅使用者的训练数据。可以说,数字三维场景对于当今学术研究、工业生产与日常生活都起着至关重要的作用,并且还将进一步影响着人们。因此,如何通过计算机辅助设计师构建出美观合理的数字室内三维场景也是一个非常重要的问题。

在引入室内三维场景构建算法之前,室内设计师在构建场景时往往需要将物体逐个添加到场景中。这些前期繁重的工作大大消耗了设计者的精力,使其无法着眼于自己最主要、最核心的设计构思。另外,室内设计作为一项很需要经验的操作,普通人往往难以驾驭,还需要高价聘请专业的设计师来帮助自己设计。

在引入室内三维场景构建算法之后,构建算法可以预先构建出一个场景草图,设计师只需要在此基础上做一点微调即可,不需要从零开始构建数字三维场景。在调节的过程中,辅助构建算法也可以向设计师进行有针对性的推荐。这些都大大降低了室内设计的门槛,使得更多人参与到室内三维场景设计中来。另外,随着深度学习技术的不断发展,也需要很多设计完善的三维场景来构成数据集。

由此可见,辅助构建的算法凭借其众多的优势吸引了越来越多的关注。学者借助深度学习等技术对场景构建算法进行了各种各样的探索和尝试,在各种层面上不断降低设计师的工作量。本文将从自动构建、交互构建和基于丰富输入构建3个角度,对数字室内三维场景构建工作进行总结归纳。

1 问题背景

三维场景在家装设计、渲染展示和游戏制作等领域有着广阔的应用与需求。随着互联网和人机交互等技术的飞速发展,这些需求也在不断拓宽。如图1所示,家装设计中,设计师通过与三维场景交互,



图1 三维场景构建平台

Fig. 1 3D scene constructing platform((a)3D scene sythesis platform(Zhang et al. , 2022);(b)Kujiale(b.kujiale.com);(c)Plamer 5D layout design platform(www.autodesk.com))

预先设计出装修与家具布局预览(Ching 和 Binggeli, 2018)。为了增加玩家的游戏体验,制造商对于三维场景的需求同样日益增加,如游戏虚拟人生 4(The Sims 4)就需要大量的场景。一部分设计完善的三维场景可用于大量构建计算机视觉中的图像多通道数据集(Handa 等, 2016)。在虚拟现实领域,可用不同的三维场景构建虚拟空间的样例,如构建轮椅使用者的训练数据(Li 等, 2020)。此外,在 3D-AIGC (3-dimension artificial intelligence generated content) 领域,大规模的三维场景数据集亦能够以不同形式用于模型训练的输入,进而助力不同的涉及三维场景的任务(Foo 等, 2023)。因此,数字三维场景在学

术研究、工业生产与日常生活等领域中都至关重要。

图 2 展现了利用 3D-Front 数据集(Fu 等, 2021)渲染的一些常见三维场景。每个场景整体为一个户型图,无论是从户型还是到房间都有自身的形状,形状通常是任意多边形。每个房间(区域)还包含若干物体,这些物体的选择与布局就是数字三维场景构建当前主要的研究对象。图 3 展示了三维场景的常见数据组织形式,每个户型中的若干房间会分开存储,其房间本身会包括墙体、地板、天花板几何的索引,房间通常会以数组的形式包含一批物体,每个物体包括对于自身几何与材质的索引、放置的位置(平



图 2 基于 3D-Front 数据集渲染的数字三维场景

Fig. 2 Digital 3D scene rendered based on 3D-Front dataset

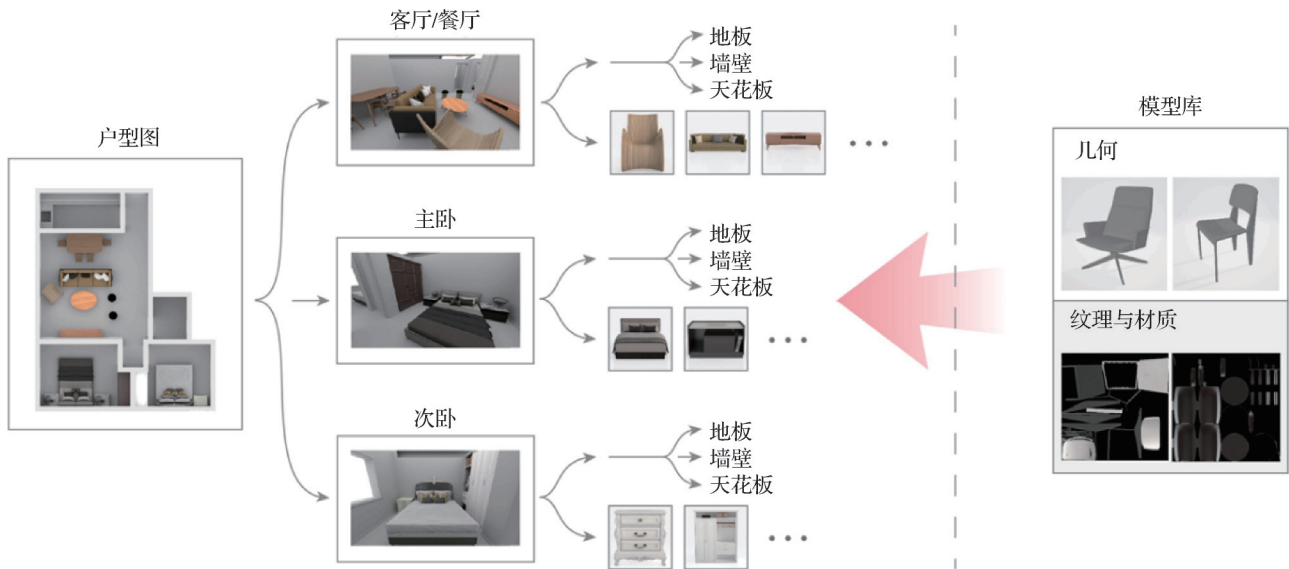


图 3 三维场景的常见层级结构

Fig. 3 Common hierarchy structure of 3D scene

移)和放置的朝向(旋转)等。这里的“索引”表示物体的几何与材质通常在数据库中另外存放,这些信息不会出现在场景的配置中。如一个场景中存在相同的 4 把椅子,则不需要将 4 把椅子的几何复制 4 次,只需要记录每个椅子在数据库中的索引即可。故三维场景构建主要构建图 3 中虚线左侧的内容。

这些三维场景能否自动构建?如图 4 所示,计算机首先推断一批物体,再将这批物体合理地摆放入房间。推断的一批物体可以由用户给定、也可以在后续摆放(布局)的过程中动态推断。“摆放”要求物体相对于墙、门、窗、地板以及其他物体等房间的要 素均合理。因此三维场景的构建是一个有挑战性并需要深入研究的课题。本文从 3 个角度介绍数字三维场景的若干构建形式。首先,介绍三维场景的自动布局方法,其输入均为三维内容,可以是一个空房间、也可以是一个没有布局或较为凌乱的房间、甚至可以是一个不存在任何物品的三维空间。自动的场景构建方法专注于在纯粹的三维条件下直接构建

出符合人类美学认知与功能需求的场景。常见的技术手段包括基于纯规则的模型、数据驱动的模型、图结构和神经网络。其次,三维内容的灵感可以来自于很多已有的数据,例如文本可以描述一个场景的构成、图像可以直接指导某个场景的主题内容、点云可以直接辅助场景重建的过程,本综述将这些多通道、丰富输入的场景构建方法作为一个独立章节进行介绍。最后,介绍三维场景的交互布局方法,其输入虽然同样包含三维内容,但输入的主体是用户的交互,如用户的鼠标指针。交互的场景构建方法专注于让用户可以控制场景自动构建的过程,即用户决定场景构建的整体方法,并由计算机辅助完善场景构建的过程。例如,在房间中放入一个靠窗的双人床,由用户指定双人床的大致位置,其余关于双人床的选取与其他物体的搭配,由计算机提示。常见的交互模式包括面向添加小物体的场景增强方法、辅助场景构建优化的场景构建方法和面向实时场景物体提示的场景构建方法。



图 4 三维场景自动构建
Fig. 4 Automatically synthesis of 3D Scene

需要注意的是,三维场景的构建通常包括物体的选取与布局两个方面。一些工作同时包括两个方面,另一些则更侧重于选取或布局。本文将不区分这些工作,而是专注于其主要的技术与应用贡献。对于三维场景构建,Zhang 等人(2019)综述在技术层面将已有的工作分类,总结了近 10 年三维场景的

构建工作。如图 5 所示。相比于该综述,本文首先将交互式的场景构建作为单独一节讨论,而不是将交互归为自动构建的范畴。同时,本文也着重讨论了不同交互构建的形式。其次,本文并不在严格意义上从技术的角度将各个工作区分开,因为对于输入、技术路线等完全不同的两个场景构建方法,其实

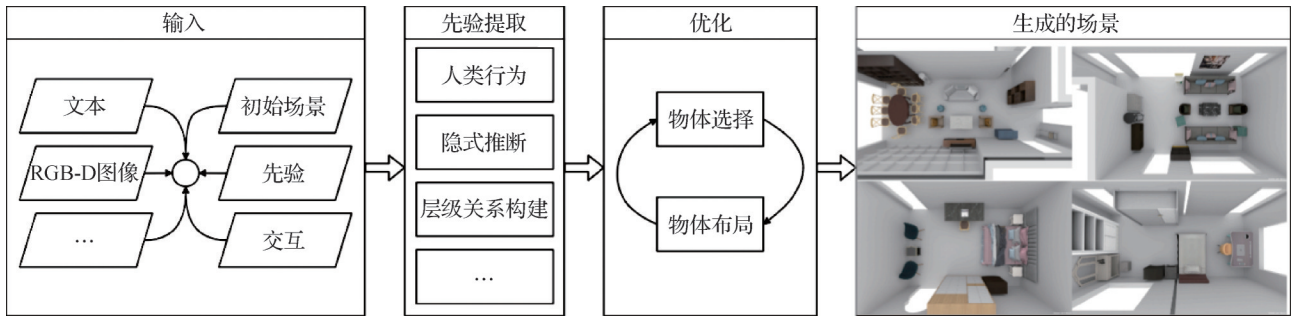


图5 三维场景构建综述(Zhang 等,2019)

Fig. 5 A review of 3D scene synthesis (Zhang et al. , 2019)

很难将某一个工作的某个技术部分直接挪用到另一个工作。故本文将对每个工作,从整体的角度,讨论其技术路线的优劣。最后,本文进一步综述了2019年后的更多场景构建工作。

2 相关主题

数字室内三维场景构建也可以从其他场景的构建中汲取灵感。本文列举了城市布局、户型构建以及自然环境构建领域中的优秀成果,并分析其中的研究方法和结论如何应用到数字室内场景构建之中。

2.1 城市结构布局

在城市布局这个问题上,分别从楼宇关系和街道构建等角度提出了解决问题的思路。Yang等人(2013)考虑城市的一片区域中如何自动地规划楼宇的问题。对于较大的一片空地,可以构建流线将其分割成小块;分割之后,又可以将成熟的楼宇模板适配进去。其自动构建的效果无论是从感性层面观察还是利用其中的量化指标来计算,都已达到相当不错的水平。

街道问题则更加复杂。因此,AlHalawani等人(2014)利用监督学习的方式对街道特征进行提取和理解。他们的工作为后续街道构建奠定了基础。Peng等人(2016)考虑在一片区域中规划街道的走向和位置。他们首先将待规划的区域离散化成网格,考虑从这些网格边沿线中挑选出哪些作为道路可以使得“道路通畅”、“全域可达”等目的实现,在此基础上由依据用户设定的“街道规划”的需求进行了量化评估。由此,“街道规划”转化为一个整数规划问题进行求解。Peng等人(2016)将粗规划作为对区域的划分,并在子区域上进一步运行这个街道规

划算法,可以更好地填充整个区域。

总的来说,以上工作在布局城市时的“待布局对象”是楼宇或街道,这些元素数量庞大,而且大都性质相似。反映到室内三维场景布局之中,则对一些需要布局大量同质对象的情景(例如大型办公室)更有借鉴意义;相应地,常见的室内场景构建一般需要面对更多类型的待布局对象,反映到城市布局上也可以体现为不同功能区布局。

2.2 房屋户型构建

区别于关注家具种类和摆放的室内场景构建,户型构建的核心在于根据用户规定的整套外型、房间种类及个数、房屋关系等要求将整套房子的房间功能和布局进行自动规划。

具体到解决问题的过程,不同的学者进行了不同的尝试。美国斯坦福大学的 Merrell等人(2010)提出在现实的户型数据集中利用贝叶斯网络进行训练,再在目标房屋外型结构上利用随机优化来构建出完整的户型的方法。Wu等人(2018)将“房屋不重叠”以及上述各种要求转化为不等式,从而将整个问题转化为一个“混合整数的二次规划问题”,并配合以一些高级先验的控制来优化构建的结果。Hu等人(2020)则是提取出每个户型的拓扑结构(房间代表点的颜色代表房间类型,点之间的边代表相邻),并依据用户需求在数据库中检索出符合要求的户型;最后,训练了深度神经网络来将这个户型适配到所需要的整套房屋外型之中。不过这种方法需要户型数据集的支持,而 Nauata等人(2020)则提出可以借助生成对抗网络来构建户型:将户型的拓扑结构视为“生成者”的生成种子,并将构建结果和真实的户型交给“辨认者”辨认,随着反复的迭代训练,生成者也能依据拓扑结构构建出令用户满意的户型图。

房屋户型构建和室内场景合成这两个问题比较接近,解决问题的方法也更相似。不过落实到实际应用之中,公寓楼里的住户似乎无法“改变户型”,因此室内布局调整往往比户型调整更为便捷,其应用场景也更广一些。

2.3 自然环境生成

高质量的动画电影或游戏往往需要生成出更为逼真的自然环境作为背景。而在自动布局这些自然环境时,人们往往会参考环境领域的相关研究成果,可见“先验知识”在场景构建过程中起着更加重要的作用。

现有的海洋场景大都是基于某些简单规则构建出来的,往往过于单调而且缺乏真实感。为了构建出一个更生动逼真的海洋环境,Zhu等人(2022)提出了一种更全面的方法。首先从真实图像中提取出各种海洋生物的位置分布信息,将不同的种群放在同一个“正八面体”中,之后在这些“正八面体”中挑选一些构成立方形状的“海洋方块”。应用Wang(1961)立方体采样技术能使得这些方块更真实地填充到海洋背景中。最后,再增加一些种类或方向的随机性,就能让鱼群得到更好的动态效果了。

Deussen等人(1998)用计算机模拟过一片地形图上植被会如何生长和分布。在用户定义的地形数据上先随机“种下”一些植物,根据这些植物的生长特性、韧性等属性配合土地的湿度、海拔等环境数据,再加入一些随机因素,模拟出了这些植物的生长过程。其中可能会有一些植物需要相互竞争,更成熟的植物或韧性更强的植物就可以存活下来。只要这些属性参数设置得合理,这个算法就能模拟出一个更为真实的植被分布图景。论文还讨论了利用简化模型等方法实现渲染加速的问题。

法国里昂大学的Génevaux等人(2013)考虑如何综合区域中大致的地势和河流信息,自动生成出更为精细的地形地貌。首先利用设计师提供的地形草图和河流草图生成出一些支流,从而形成较为发达的水系。再利用这些河流和山脉将整片区域分成不同的流域,然后在各个流域上分别生成出精细的地形结构。整个过程参考了很多水文学的知识,才得以获得令人满意的结果。

不难发现,各个领域的重要的先验知识在自然场景的构建中起到的作用超过了“深度学习”等新方法。如果后续能对“室内布局”进行一些高级抽象层

面的建模,或许可以避免持续使用一些可解释性较差的研究方法。

3 室内数字三维场景自动构建

三维场景自动布局的目的是利用房间轮廓和功能、部分家具的位置等信息来“自动”构建出一套合理并且美观的室内布局方案。这项工作的“自动”性体现在,其输入并不是一个最终的布局方案(如丰富输入的室内三维场景构建中讨论的问题),也不能期望通过不断的人为输入与调整来获得一个合理的布局(如室内三维场景交互布局中的情景),而是通过计算来自动布局和调节。

如何让计算模型了解到哪些室内场景是合理且美观的呢?一方面可以先验地将某些布局规则转化成计算规则交给计算模型,如Weiss等人(2019)和Yeh等人(2012)的工作。不过毕竟很难总结出所有的布局规则,因此随着数据集的不断增大和机器学习算法的不断发展,越来越多的工作选择了利用机器学习算法从现有的室内结构中学习隐藏的规律,再引入到生成模型中指导新场景的构建。因此,本文对这部分的各项研究成果从两个角度进行讨论:1)机器学习的方法和内容;2)场景构建的主要方法。根据获得先验主要技术方法将相关工作分为基于人为设定的规则的非机器学习、基于图模型、基于神经网络以及其他的机器学习方法共4类分别介绍。需要注意的是,三维场景自动布局是三维场景自动构建的一个子问题。场景构建通常包括物体选择和物体布局两个阶段。在物体选择阶段,需要根据房间功能选择一批合适的物体。而在布局阶段,通常默认物体已经由选择阶段选定了,重点考虑如何将其合理地布置到房间中(Zhang等,2019)。另外,场景构建过程中的风格设计通常在物体布局的过程中完成,这些风格信息大多通过布局规则给出或者从现有的场景数据集当中学习得到;也有一些工作直接聚焦于场景的风格设计,如Liu等人(2015)提出了一种算法以提取不同物体间的风格一致性,并将这一信息运用到了场景构建过程中。

3.1 基于数据驱动的学习

近10年来,比较有开创性的三维场景布局工作为Yu等人(2011)提出的Make-it-Home。其主要思路如图6所示。作者首先从数据集中学习出物体之

间的空间关系、从属关系和共生关系。Yu 等人(2011)将这些提取出来的先验信息引入一项评价场景的函数中。布局场景时,会先采集目标房间的墙体、门窗

等信息,并将家具随机放入房间内;之后会基于模拟退火对函数进行优化,从而实现对布局的优化;模拟退火的过程持续进行,直至得到一个满意的布局。

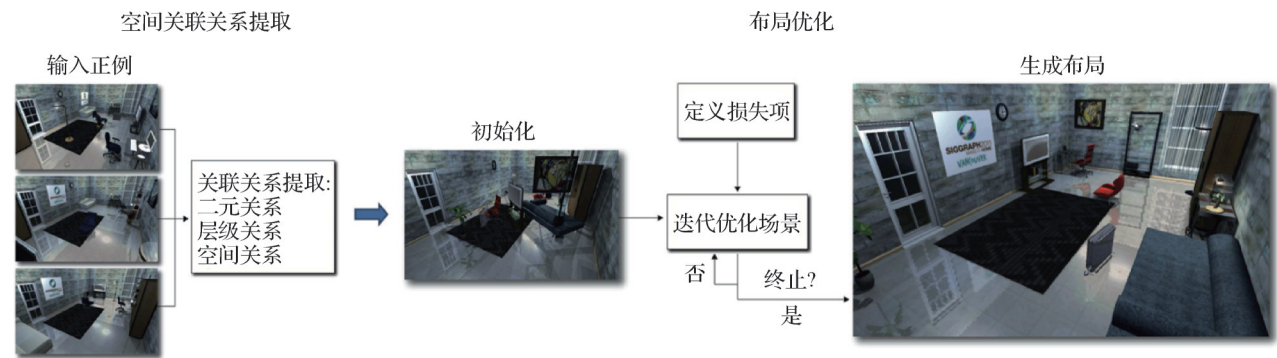


图6 Make-it-Home(Yu 等,2011)
Fig. 6 Make-it-Home(Yu et al. ,2011)

另一个代表性的工作是 Zhang 等人(2022)提出的混合式的三维场景布局。其特点在于综合了数据驱动与非数据驱动的各自优势:对于每个功能区内部,基于数据驱动模型来学习其布局的先验,对于功能区之间,基于优化或几何算法来对功能区在房间中布局。故混合模型既能发挥先验驱动出的多样性,又能得益于更简单直接的优化方法来更快速地得到布局结果。如图4所示,区别于以往的研究常常基于共现的频率(co-occurrence)来决定物体间的关联的强弱,这里采用了 test for CSR (test for complete spatial randomness)来决定物体间空间关联关系的强弱。其核心在于,在关联中引入了空间的概念,因此可以将一些由共现错误引入的关联解耦。如空调与衣柜通常不具有空间关联性,但由于空调的特殊性,它与任何物体都具有很高的共现。这篇文章还提出了基于离散采样的先验提取方式。由于三维场景中,物体与物体间的空间布局关系及其复杂、难以用特定的分布表达,取而代之,其并不假设具体的分布,而是期望能得到真实分布的一批离散采样。最终不仅提升了布局的速度,也提高了布局结果的合理性。

图7展示了另一项混合式的三维场景布局的工作(Zhang 等,2021b),其特色在于:1)提取3个或更多的物体间的多元关联关系;2)首次使用了基于几何算法的布局,相比于优化与神经网络,几何算法的速度更快、更易解释也更可控。这项工作对于中小型房间已经可以达到实时布局。

这项工作使用了 3D-Front 数据集(Fu 等,2021)。它是一个有着 34 个类别的 9 992 个三维物体的模型和超过 7 万个房间实例的大型数据集。这是中国团队自主研发的高质量室内场景数据集,其户型和家具摆放方式都是直接来自专业的设计,装潢和家居纹理等内容的设计也能保持一个相对统一的效果。另外,3D-Front 还在很大程度上缓解了国外数据集的各种版权问题,为国内的科研人员提供了很大的便利。

Xu 等人(2015)在进行室内场景构建时,利用场景中物体所处的环境,可以更好地提取物体的空间朝向。他们还更合理地划分了物体之间的“功能组”结构以避免出现风格明显不适配的场景效果。不过,其最核心的贡献在于,将地面和墙体空间以网格形式进行划分,在数据集上统计了各类“功能组”在各个网格上的频次之后,利用朴素贝叶斯算法学习出最终的场景构建方法。

Fu 等人(2020)受到 Fisher 等人(2015)思路的启发,将室内场景中人的活动引入到对场景的评价和构建中来。实际场景中,用户在房间里不仅会评估他们和物体的交互能力,还会在意整个房间的舒适程度。在考虑前者时,Fu 等人(2020)首先对家具进行设置,标识出人可以在上面“坐着”、“躺着”或“站着”,并考虑人在场景中的“拿起”物体和“看见”画面(如看电视)的活动;而对于后者,则通过针对大量室内场景的学习,评估了整个房间的拥挤程度、光照和通风情况。将这些与人密切相关的要素考虑进去之

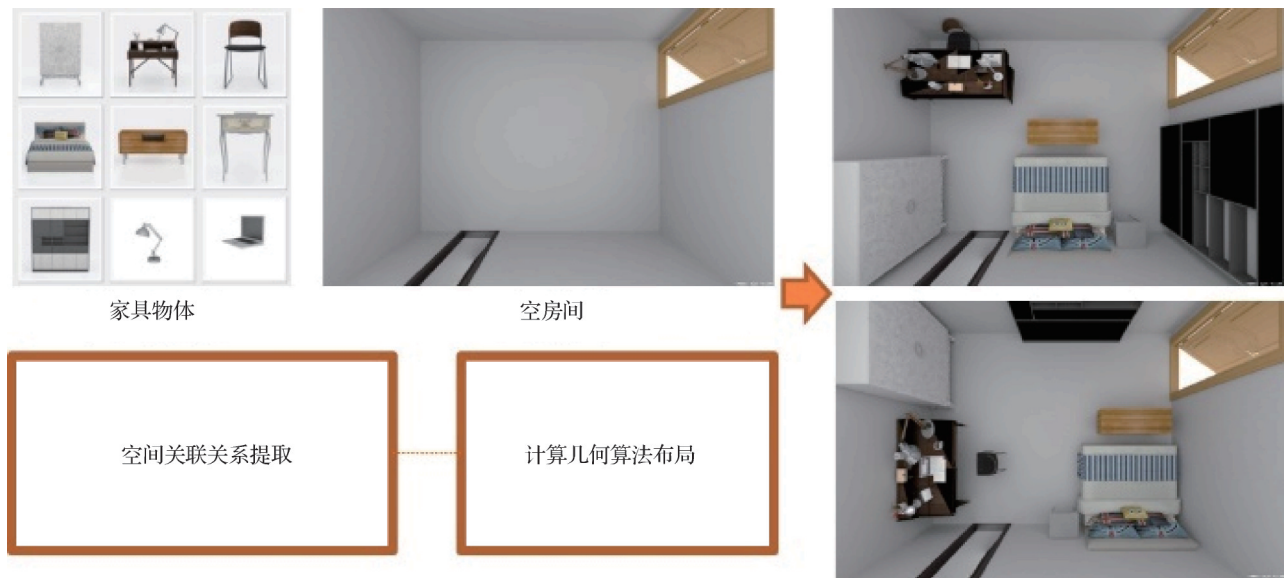


图7 基于物体间多元关系与计算几何算法的三维场景布局构建(Zhang 等,2021b)

Fig. 7 Generation of 3D scene layout based on objects' hyper-relation and computational geometry algorithms (Zhang et al. , 2021b)

后,效果得到了广泛提升。

3.2 基于人为设定的先验规则

尽管凭借人们的直觉,构建场景的先验要从数据中学习,然而三维的数据通常无法大量获取且存在各种各样的版权问题。基于人为设定的先验规则的一大好处即不需要数据集的支持。而非数据驱动的工作从根本上脱离了数据驱动的先验学习,转而人为设定一系列的先验。例如可以人为定义物体间的约束并优化。Weiss 等人(2019)定义一系列的物体间关联关系,如“相对距离约束”、“中心点约束”、“对称性约束”等。由于先验相对简单,这个方法可以快速地生成结果,但对于每个场景而言,都需要人工预设置一批先验,故此方法更适合于同一场景复用或大规模场景构建。

Yeh 等人(2012)对马尔可夫链蒙特卡洛方法(Markov chain Monte Carlo method, MCMC)优化过程提出了改进,提出“本地退火的可逆 MCMC 算法”。这种优化算法在优化的过程中同时调节场景中物体的数量,即物体的选择可参与优化过程。这使得开放世界下的三维场景自动构建成为可能。

3.3 基于图结构

近年的人机交互工作中处处可见图拓扑结构的身影,例如机器学习中的概率图模型、用于表征物体关系搭建的图关系网等。再加上图神经网络技术的发展,将室内三维场景构建转化为图结构并应用机器学习算法似乎成为一种重要的思路。本节介绍的

几项数字室内三维场景自动构建算法的工作,也取得了突出的效果。

Liang 等人(2017)不但尝试对已有家具进行布局,还考虑根据房间类型自动引入新家具。作者将房间比为文档、话题比为类别、家具比为单词,构造出一个话题模型用于挑选新家具,并使用坍塌的吉布斯采样方法求解这个话题模型。而在室内场景布局过程中,他们又将“摆放对齐”、“空间匀称”和“家具的功能”等因素纳入考量,构造出贝叶斯网络并使用基于 MCMC 的优化方法进行求解。其场景的构建效果如图 8 所示。相对于其他的场景构建算法,该工作将房间类型纳入考量,这样一来就不需要再人为引入某项家具,从而真正实现室内场景的自动构建。

Fu 等人(2017)的成果如图 9 所示。他们在构造物体间的关联关系图时,首先人为地将家具分成两类:1)用户会停留时间较长的一类功能性物体;2)用户会与之交互并停留时间较短的一类物体。之后再分别统计两类物体之间的共生关系。这样就在物体选取上引入了功能性先验,也是针对家具物体功能性较为创新性的分类。在构建场景时,考虑室内的门窗关系、面积大小、已有家具等因素,向已经建成的“物体间关联关系图库”中查找应该添入的家具,并在数据库中查询相似的房间场景用于指导新房的构建。该研究可解释性更强,操作较为简单,普遍性更好。



图8 室内场景生成效果(Liang 等,2017)

Fig. 8 The results of indoor scene synthesis (Liang et al. , 2017)
((a)toilet;(b)meeting room;(c)bedroom;(d)a mixture of dining room and kitchen)

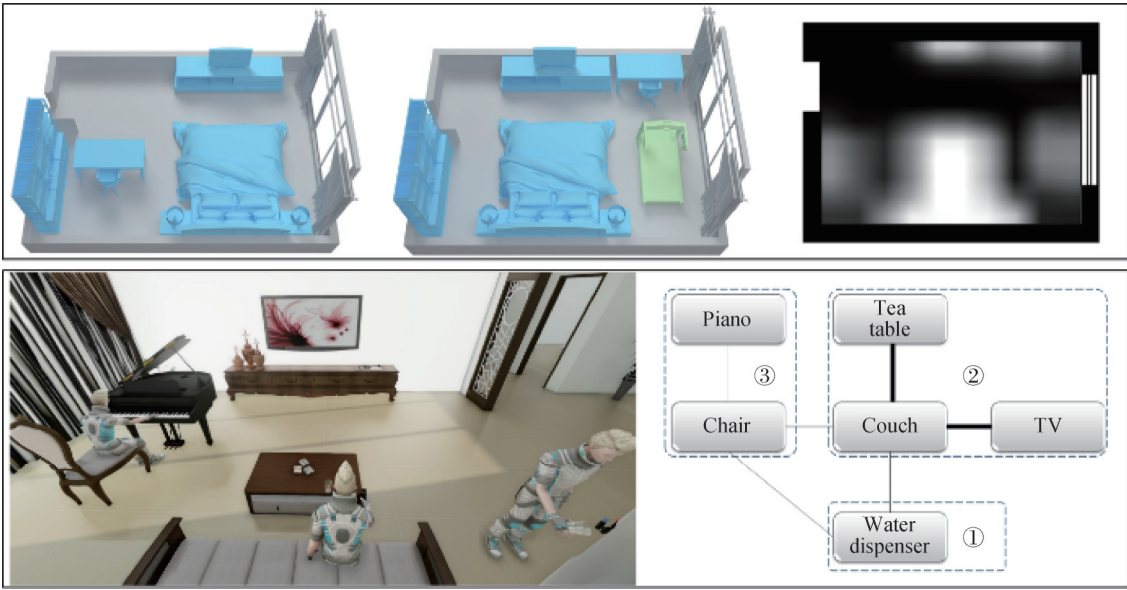


图9 基于功能约束与“俯视图先验”的三维场景构建(Fu 等,2017)

Fig. 9 Synthesize the 3D scene based on functional constraints and the “prior of vertical view” (Fu et al. , 2017)

Qi 等人(2018)工作的贡献为引入了空间维度的与或图S-AOG(and/or graph)机制。其构造了一个完整的图,图中父节点会包含子节点的主题。每次场景构建时,在优化布局前先行自动选择房间类型与对应的一批物体,即不需要人为选取物体。例如

构建某一房间时会根据“或”来选择到某一“卧室”节点,选择了“卧室”父节点后,其下的若干卧室物体子节点会被以子集的形式继续选中。选择到的一系列物体会被基于MCMC框架的方法继续优化,最终得到布局。

3.4 基于神经网络

自从神经网络在“识别”、“分类”、“判别”等任务上展现出极强的能力以来,人们就不断构思着将这一复杂的结构迁移到“生成”任务之中。实际上,也可以将“室内场景自动构建”理解为“生成”一个室内场景,从而将问题建模成一个“生成”任务。在深度学习生成任务取得各个里程碑式进展的同时,数字室内三维场景也都应用着这些进展来不断提升自己的效果。尽管基于神经网络的工作同样也是数据驱动,但考虑到基于神经网络的工作越来越多而且这些工作本身也不断细化并形成了自己的思路,如基于卷积神经网络或基于对抗生成网络,所以本文将

这些工作提取出来作为一个独立的部分进行专门介绍。

其中,比较有代表性的工作是由Wang等人(2018,2019)提出的基于卷积神经网络与场景俯视图学习先验的一系列工作。其核心流程较为类似,以目前较新的PlanIT为例进行介绍。其训练了若干神经网络,分别负责物体类别的选择、物体的放置、物体的朝向以及迭代是否终止。向场景中添加某个物体时,首先询问第一个网络要选择哪一个类别,后续的网络继续回答物体的位置、朝向。最终由一个网路判断迭代是否终止。训练网络与推理的输入都是预渲染好的俯视图,如图10所示。

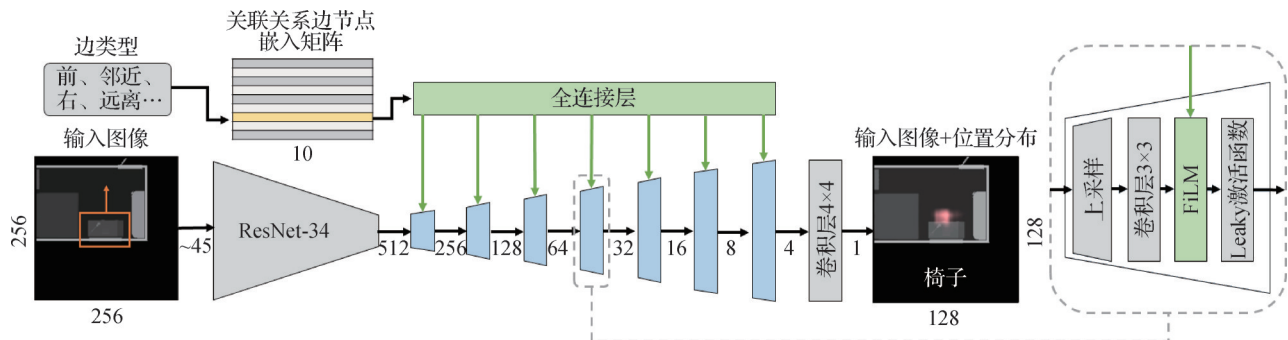


图10 基于神经网络的物体布局推断(Wang等,2019)

Fig. 10 Inference of object layout based on neural network (Wang et al., 2019)

基于卷积神经网络的方法需要迭代式地向场景中不断添加物体,其结果容易受到顺序的影响。Li等人(2019)、Purkait等人(2020)等研究者受到变分自解码器(variational auto-encoder, VAE)的启发,通过对场景构建语法树,再训练VAE用于将这些语法树和高维隐空间的随机数构建映射,从而实现了室内场景的自动构建。VAE后续还被用于场景的直接生成(Yang等,2021a)和对家具的自动构建(Ye等,2022),都取得了不错的效果。生成对抗网络(generative adversarial network, GAN)也作为引导训练神经网络实现生成的有利工具横空出世。Zhang等人(2020)将场景及其俯视图分别送入生成对抗网络的判别器之中,再将损失返回到生成器之中训练其生成三维场景。Yang等人(2021b)则将俯视图扩展为某视角下的深度纹理以及语义标注,并将三维场景的生成器改为生成体素之后再语义分割的形式,可以在很大程度上提升模型的泛化性能。Nie等人(2023)则将俯视图扩展为各个视角下的场景照片,并将隐空间限

定为“单位球面”,进一步提升了算法的构建效果。

随着Transformers模型的提出(Vaswani等人,2017),其很快用于生成文本等顺序性的信息。Wang等人(2021)和Paschalidou等人(2021)也将场景中的物体编码成序列的形式,之后再借助Transformers模型来进行迭代式的生成。基于扩散模型(diffusion model)的室内场景构建算法也不断提出(Tang等,2023)。实际上,扩散模型所常应用于的空间比数字室内三维场景的空间要大很多,因此在未来,这项模型可以用于进一步提升数字室内场景的多样性,使其不必局限于物体的素材数据库,而形成更丰富多样的模型和场景。

这几项基于神经网络的工作选择了SUNCG数据集。SUNCG是由美国普林斯顿大学构建的一套大型室内三维场景数据集(Song等,2017),其中包含了175个类别的2266种物体模型,包含了超过52万个房间实例。这个数据集凭借其密集的体素标记和优秀的模型质量,已经广泛应用于各类数字室内三

维场景构建的算法研究之中。总的来说,现有的三维物件素材种类多样、数量丰富,可以满足三维场景构建的需求。

与之相对,有些工作(He 等,2021)引入了更多约束。如图 11 所示,在构建场景的过程中,方法不仅仅要保证物体间的关系满足提取的先验,还必须

要服从实际场景带来的约束。例如,实际场景中不存在一张桌子,如果虚拟场景对应的位置不存在任何物体,则用户可能会在戴上 VR(virtual reality)头盔时撞到实际场景的物体。故虚拟场景也要在对应地点生成有意义的“障碍物”或放置家具,从而确保用户在交互期间的安全性。

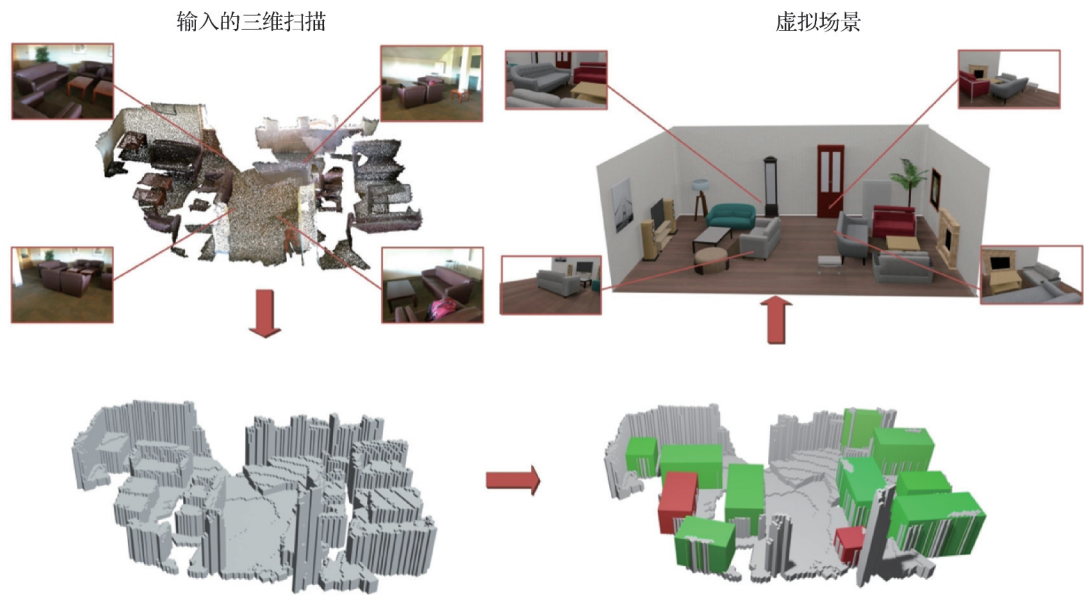


图 11 基于实际场景约束的三维场景自动构建(He 等,2017)

Fig. 11 Automatically construction of 3D scene based on physical scene constraints (He et al. , 2017)

3.5 小结

近些年的室内三维场景构建研究更多聚焦于构建的效果。在室内场景数据库中,人们对不同房间类型的场景数据分别训练并分别评估其构建效果,而没有将房间类型纳入模型中从而试图提出一种泛化能力更强的室内场景构建算法。较早先的工作(Yu 等,2011;Fu 等,2017)则会在构建前提供一些室内家具的信息,从而很好地规避了不同房间类型之间的差异。如何能在“房间类型”的层面上进一步提升生成模型的泛化能力,是一个值得进一步讨论的问题。另一个问题在于,这些工作往往忽略了房间、门、窗等“硬装修”的约束需求,而将更多的精力放在房间中的物体之上。然而,在实际应用上,通常数字家装都是基于已有的户型、门洞和窗洞构建场景,而几乎不存在基于物体间布局来重新设计户型的情景,故直接构建完全不考虑硬装的家具布局并不满足常见的商业需求。

本文将室内数字三维场景自动构建一节的主要

内容列举在表 1 中。表中不但详细列出了相关工作所使用的具体技术,还对各项工作中利用机器学习所学习到的具体内容进行了整理,最后从室内三维场景自动布局的角度上探讨这些工作为了布局需要提供哪些输入。其中,基于纯规则的自动布局方法不需要进行学习,基于神经网络的布局方式则往往将不同“房间类别”的数据分别训练建模。纵观数据驱动、具体规则、图模型和神经网络等室内场景自动构建方法,虽然各有优劣和侧重,但在如今大量数据的支持之下已经能较好地实现家具布置、室内场景数据生成等任务。不得不说,场景的自动构建本身已经实现了 0 到 1 的突破,应当进一步向场景构建的其他问题或者其他领域进行推广。

基于数据驱动的自动构建算法中涉及到了许多室内三维场景的数据集,本文将其主要特点列举在表 2 中。其中的多通道指数数据集是如何提供对场景的扫描信息的,物体种类数目则是指数据集中提供了多少种家具的模型,几何关系则表示是否给出了

这些家具模型之间的几何位置关系。数据集 Structured3D 是一项记录了大量室内场景的 RGB-D 扫描结果及其家具标注的数据集,将在下一节中讨论。Fu 等人(2021)的工作中给出了更详细的通道、模型与布局信息的统计。

这些室内场景构建的研究工作是如何对算法的构建结果进行评价的呢? 作为一项人机交互方向的前沿领域,结果评价总体上还是采取了“用户实验”的形式,让用户主观地对算法的构建结果进行评价(Hu 等,2024)。由于这些场景最终是将用于于人交互的情景,因此用户实验才是最为直接的评价方法。不过用户实验的具体方式略有差异。Yu 等人(2011)采取了让设计师来为场景构建结果打分的方

式,这种方式在后续的研究工作中也非常常见(Fu 等,2017;Zhang 等,2021b)。为场景打分的问卷中既包含了代表整个场景的宏观信息(如美观程度、实用性和便捷性等),还包含了代表各个物体的微观层面的信息(如物件位置、尺寸、风格以及材质的多样性等),从而可以从全方位对室内三维场景进行评估。还有一些算法的研究(Yu 等,2016;Zhang 等,2021b)采取了将新算法的构建结果与其他构建结果同时呈现在用户面前,让他们选出更合理的布局方案的方式。近些年来,还有一些工作(Yeh 等,2012)将构建算法运行的中间结果直接交给用户来进行评价,大大提升了评价的可解释性,从而进一步提升了成果的价值。

表 1 室内三维场景自动构建总结
Table 1 Summary of indoor 3D scene automatic synthesise

方法	分类	核心技术	机器学习内容	室内布局输入
Yu 等人(2011)	数据驱动学习	基于模拟退火优化的方法	物体空间关联、从属、共现关系	物体随机摆放
Zhang 等人(2022)		空间关系先验	物体的空间相关性	物体列表
Zhang 等人(2021b)		多物体间多元关联	物体间的多元关联	物体列表
Weiss 等人(2019)	纯规则	基于 MCMC 优化	-	物体随机摆放
Yeh 等人(2012)		LARJ-MCMC 方法	-	场景功能
Fu 等人(2017)	图结构	物体关系图等	物体间共线关系等	场景中少量物体
Liang 等人(2017)		话题模型与贝叶斯网络	物体间及房间关系	房间功能
Qi 等人(2018)		S-AO 图,马氏随机场	物体间共现、承载	功能与外廓
Wang 等人(2018)	神经网络	多级卷积神经网络	房间布局总体信息	房间半成品
Ritchie 等人(2019)		多级卷积神经网络		
Wang 等人(2019)		物体-房间关系图		
Li 等人(2019)		变分自编解码器等		
Zhang 等人(2020)		生成对抗网络等		
Yang 等人(2021a)		贝叶斯公式,L-BFGS		

注:“-”表示该方法未涉及机器学习。

表 2 室内三维场景数据集比较
Table 2 Comparison of 3D indoor scene datasets

数据集	多通道	物体种类	几何关系
SUNCG	自由渲染	175	提供场景配置
3D-Front	自由渲染	34	提供场景配置
Structured3D	RGB-D	-	提供物体变换标注

注:“-”表示其中物体无种类标注。

4 丰富输入的室内数字三维场景构建

本节讨论的室内三维场景自动构建并不是一定标准下对场景构建过程的自动优化,而更倾向于一种对输入信息的更直接的表现。然而,如果这些输入信息只能体现三维场景的某些侧面,例如视觉图景(RGB 图像)或深度信息(RGB-D 扫描),并不能完

整体现整个场景,又应该怎么做呢?如何引入有关的先验来辅助算法完整构建出整个室内场景,便是这一部分工作的主要难点所在。这一节的内容也针对不同类型的输入讨论作者基于什么方法构建先验知识、构建出哪些先验知识,从而进一步实现了室内数字三维场景的构建。

4.1 基于RGB-D扫描结果的构建

相关领域最早的工作就是基于RGB-D的扫描图像的构建。Nan等人(2012)只有一幅表征深度的灰度图。他们构建场景的方法大致是:将整幅图像根据不同的深度划分成不同的区域,如果将某部分区域和其相邻部分合并所得的图形可以更像是一个室内场景中常见物体(如桌子椅子等),则将其合并;不断重复这个过程,整幅图像就得以被分为不同部分,各个部分分别对应一个物体;再根据这些物体的类别、尺寸、朝向等信息重构场景即可。Chen等人(2014)的工作在整体家具环境构建之外,还着重对

于场景中小物体进行了重建,解决了室内三维场景构建过程中的又一大难点,如图12所示。Fisher等人(2015)的工作在于让构建的场景同时满足输入的RGB-D扫描和用户行为,其场景构建需要满足两种约束:1)构建的场景在几何上要满足由RGB-D扫描导出的体素(实际上是离散化的若干长方体);2)构建的场景要让物体可以满足用户行为。实际上,正是将“人”的因素纳入考量使得其构建的效果大大提升。更多工作及其详细信息可以参考Chen等人(2015)整理的综述。

这一领域随着扩散模型的发展也有了不小的进步。借助文生图技术和人工智能扩图技术的不断进步,众多科研人员尝试通过先让这些模型生成的优质室内场景图像(Höllein等,2023)、环视图(Fang等,2023)或RGBD扫描结果(Lei等,2023),再将这些数据结构转换成数字三维场景的形式,从而构建三维场景,也都取得了显著的效果。

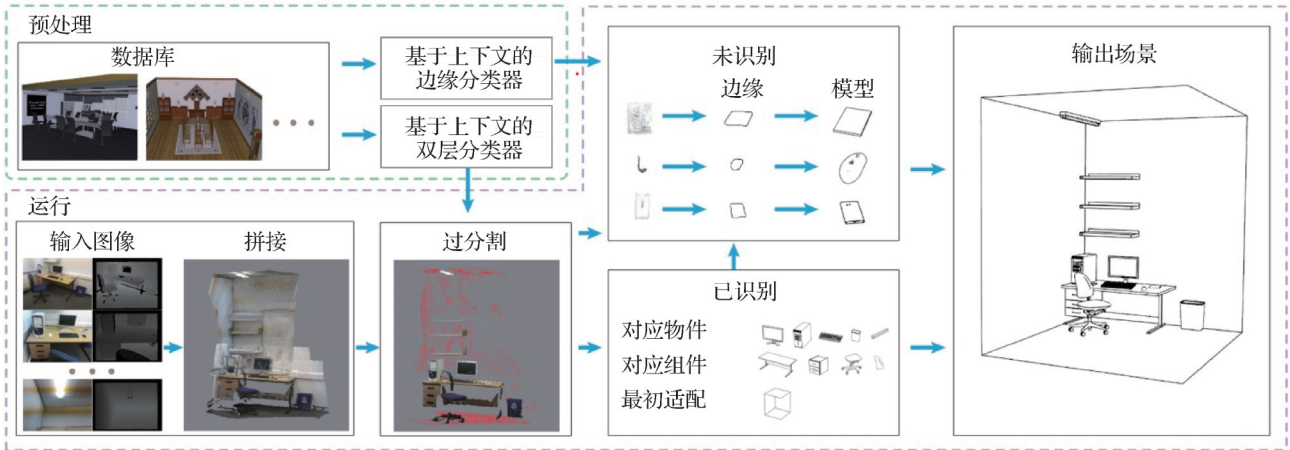


图12 利用RGB-D扫描结果构建数字三维场景(Chen等,2014)
Fig. 12 Generating digital 3D scene from RGB-D scan (Chen et al. , 2014)

4.2 基于RGB图像的构建

然而,人们平常更容易获取到的是一张照片,也就是RGB图像,并不总能利用深度信息。不过这样一来,输入数据的格式更进一步贴近了卷积神经网络的输入格式。因此,Zhu等人(2018)考虑先训练一个卷积的长短时记忆网络用于提取出照片中的物品的空间关系;再训练一个回归卷积神经网络用于从二维图像中构建三维场景。相较于其他的直接建模的工作,将上述空间关系纳入考量在很大程度上改善了构建的效果。

4.3 基于自然语言文本的构建

这一部分的工作试图将对于场景的自然语言描述转化为场景的三维模型。人们在描述场景时往往只会描述显著的内容而忽略次要的元素,那么构建算法该如何推断出这些隐藏的次要内容呢?Chang等人(2014,2017,2018)做了一系列工作。他们用结点表示物体、用边表示关系,将整个室内场景总结为一种称为“场景模板”的图结构。构建场景时,首先将输入的文本依据预设的关键词解析成场景模板,之后再根据场景库中的场景模板的特征,试图向场景模板图中补充一些可能被输入文本省略了的点和

边。补充完整之后的场景模板图可以指导构建出真实感更强的室内三维场景。Chang 等人(2018)后续的工作不但将上述“人为设定”的关键词和关系转化为一种到数据库中去自行学习的过程,从而提高了其泛化性能,还增加了基于自然语言的交互构建过程。Ma 等人(2018)发现文本在较小尺度的场景构建领域效果更好,因此对整个构建问题进行拆分构建再合并调整的处理方法。

4.4 基于草图的构建

Xu 等人(2013)提出了 Sketch2Scene 框架,将输入的手绘草图最终转换为数字三维场景。第 1 个思路在于可以协同检索即将放入场景中的物体,即只要某一个草图可以被检索到,其余由于手绘偏差、手绘视角带来的不确定性便可以弥补;第 2 个思路在于可以协同布局三维场景,即结合手绘的内容和类似于前文场景自动布局的先验(如高斯混合模型)来得到最终结果。效果如图 13 所示。

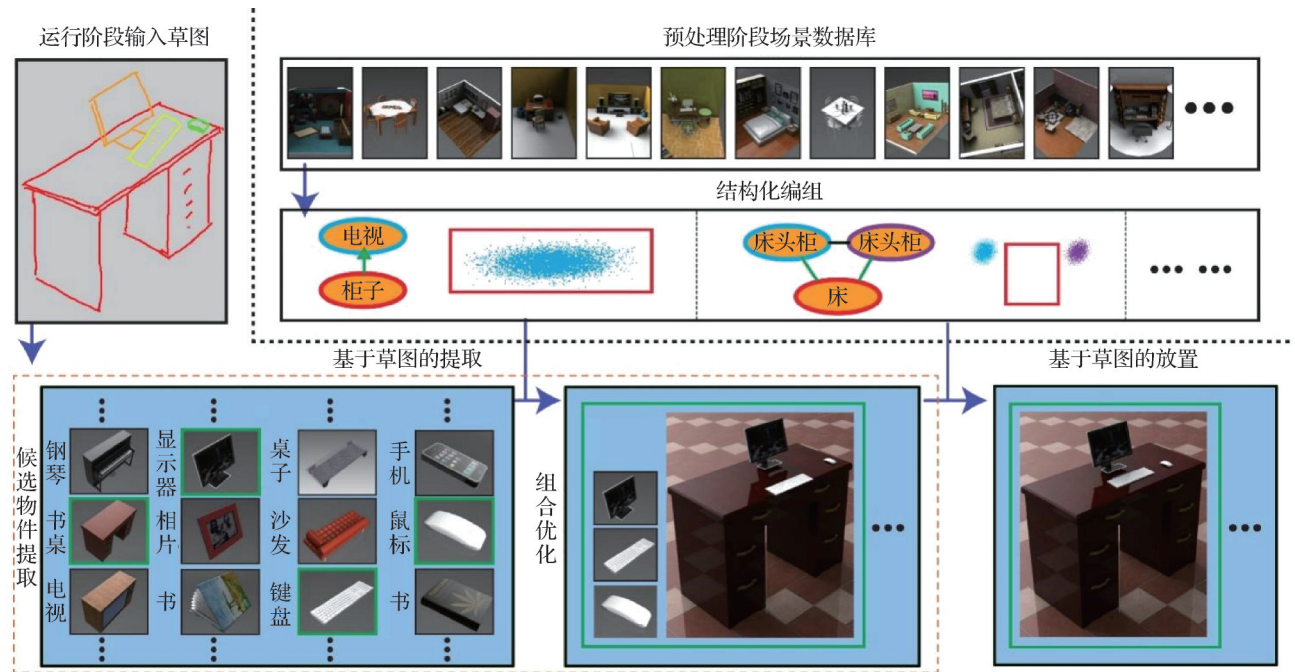


图 13 利用手稿草图作为输入自动构建室内三维场景(Xu 等,2013)

Fig. 13 Automatically constructing 3D indoor scene, using sketches as input (Xu et al. , 2013)

多通道输入的数字室内三维场景构建算法研究同样离不开数据集的支撑,其中Structured3D(Zheng 等人,2020)是一项非常重要的数据集。它是一款由酷家乐开发团队开发的数据集,其中保存了大量室内场景的RGB-D扫描结果,并对扫描图像中各个像素位置的家具种类进行了标注。该数据集利用先进的RGB-D扫描引擎,在专业的室内设计结果中扫描获得大量图像。这也使得它在物体识别和数字场景重构方面体现出更好的效果,广泛应用于最新的科研工作中。

不同于室内场景自动构建,基于丰富输入的场景构建问题在评价时则有更多的真实结果可以用于验证。正如本文前面所述,基于丰富输入的场景构建可以看做将文本、图像和RGBD扫描等非格式化

记录的三维室内场景转变为格式化的数字三维场景的过程,这样一来,就可以更为便捷地将相机、扫描器和文字描述等方式记录的三维场景转变为数字三维场景并进行进一步的展示或编辑。然而,反向的过程,即从数字三维场景向非规格化表示的转化过程,相对容易实现(例如场景的RGBD扫描可以通过对数字三维场景进行渲染并保留Z-缓冲的形式,文字描述也可以通过根据家具类别来自动组织文字的形式)。因此相关研究在掌握了数字三维场景数据集后,便可以直接生成各种非格式化的场景作为算法的输入,并将原数字三维场景中的信息作为“真实结果”,从而评价出场景构建算法,还原出这些原数字三维场景的能力。

5 室内数字三维场景交互构建

5.1 主动交互

本文介绍了室内三维场景的两种构建方式,它们的构建过程中都不需要人的实时干预。然而,自动构建的场景并不能保证用户最终的满意,在某些情况下,三维场景的构建需要满足用户的个性化需求,例如,用户要求最终的构建结果要将某个特定物体放在窗户下正中间。显然,没有人工干预的三维场景自动构建并不能保证构建结果能完全满足用户需求。另外,完全依靠人工手动构建的场景,虽然可以保证最终场景的合理性和满足用户的所有个性化需求,但其构建过程非常烦琐并且耗费大量人力。因此,需要创造一些权衡人工干预和自动构建两者的工作,让用户以交互的方式干预场景自动构建的过程。

较开创性的工作是 Yu 等人(2016)的 Clutterpalette(调色盘),如图 14 所示,将某一位置可能出现的小物体映射到“调色盘”上,供用户选择。其生成对象主要是室内场景中的小物体(如书、食品、衣物等),该工作旨在为已经生成好家具物体的室内场景增添更加丰富的细节物体。对于一个大体布局好家具物体的室内场景,用户通过点击场景中某个位置(如书桌),可自动获取该系统推荐的一批细节物体(如书本、电脑、台灯等)。在该位置上的细节物体按照推荐概率排序在“调色盘”上,让用户自行选取。Zhang 等人(2016, 2021c)在交互式添加小物体的应

用背景下,首先通过数据集中小物体之间的相对位置信息,让物体的放置(布局)自动化。然后通过交互界面,确定用户偏好,将用户偏好引入该自动化构建的结果中,并由此推广到所有场景,从而进一步减少了人力交互所用的成本。Savva 等人(2017)的工作主要针对完整场景。用户单击场景中的位置时,系统会弹出对于该位置推荐的物体菜单,用户可以通过该菜单选取自己满意的物体进行摆放。但物体的选取与摆放过程需要较多的人力交互。Yan 等人(2017)的工作同样针对完整场景,对于已经自动构建的场景,用户可以通过选择、拖动一个或者一组物体,改变被选择物体的位置。该场景会针对用户的操作相应地做出布局改变,从而使场景更加符合用户的审美。但该工作的优化基于 MCMC,相对于其他优化方式耗费较多时间。Merrell 等人(2011)的工作不是通过先验的学习来构建场景,而是一系列室内设计原则确定一个成本函数,经过 MCMC 优化得出布局结果。用户首先确定房间形状和需要放置的家具,再通过前述方法得到系统建议的布局,用户继而对该布局中的物体进行操作,系统继续优化,并在再次优化的过程中保留用户之前对物体操作的结果。在整个优化和交互的过程中,得出最终令用户满意的布局结果。但该工作的方法基于 MCMC,也存在优化耗时较长的问题。

MageAdd (Zhang 等, 2021a) 和 Scelle Director (Zhang 等, 2023b) 实现了实时交互工作,以前者为为例,如图 15 所示,在任意的时间,当用户使用光标在场景中移动和单击时,方法都会实时地自动选择合

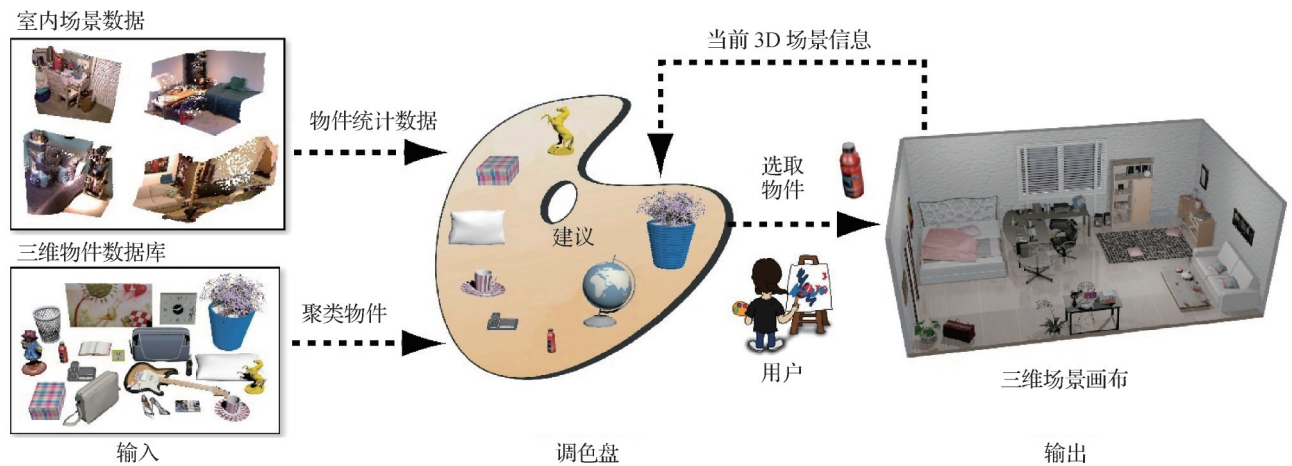


图 14 Clutterpalette(Yu 等, 2016)
Fig. 14 Clutterpalette (Yu et al., 2016)

适该位置的物体,并将其置入数字三维场景。该方法的特点在于实时维护的先验库,每当用户与场景发生交互,如添加或删除物体,先验库都会根据用户的行为相应更新。例如用户添加了双人床,系统会立即在一些情况下提示出床头柜。该方法能达到实时的原因在于,每次交互后都仅仅使用很少的一部分先验进行计算,其计算过程并不需要考虑完整的先验库,从而减少了计算时间。因此当用户用鼠标在场景中快速移动时,可以很快地探索出场景中所缺少的物体,以及这些物体的合理排放位置。

随着虚拟现实的发展,媒体的呈现形式已从传统的二维屏幕转变成成为三维的虚拟现实设备。由于虚拟现实的沉浸性和可交互性,未来将会有工作致力于通过 VR 中的主动交互构建三维场景,如手势、眼动、头动和触觉等。如何有效地将这些交互方式与场景构建自动算法结合起来,从而高效地构建三维场景,是未来亟需研究的方向。

此外,现有的绝大多数主动交互工作中均设计了独立的交互界面来完成场景构建过程。由于这些

界面由学术界所设计,其交互过程未必能很好地契合设计师用户的需求;另外,设计师目前所使用的商业软件中的各种设计未必能直接兼容各种交互构建算法,例如场景表达方式之间可能存在着差异,软件中缺少算法所需要的某些参数等。如何设计交互构建过程使之兼容现有的设计方法,进而易于设计师用户理解和使用,亦是值得深入探讨的研究课题。

三维室内场景交互构建的算法评价主要基于两个角度。首先,与自动构建时类似,实验者会将算法生成的场景制作成问卷让被试进行打分,并统计评分结果。这些场景的最终应用场景依然是用户,因此让用户依据他们的主观感受进行打分是最直接的评价方法。其次,由于交互构建不仅评价场景的最终效果,还会研究构建者在构建场景过程中的行为以及如何提升其工作效率,因此评价过程中也会请被试对算法提出的交互流程进行实际操作,并统计交互过程中的耗时、点击次数和推荐条目使用率等指标,从而显示出构建算法提升效率的程度。

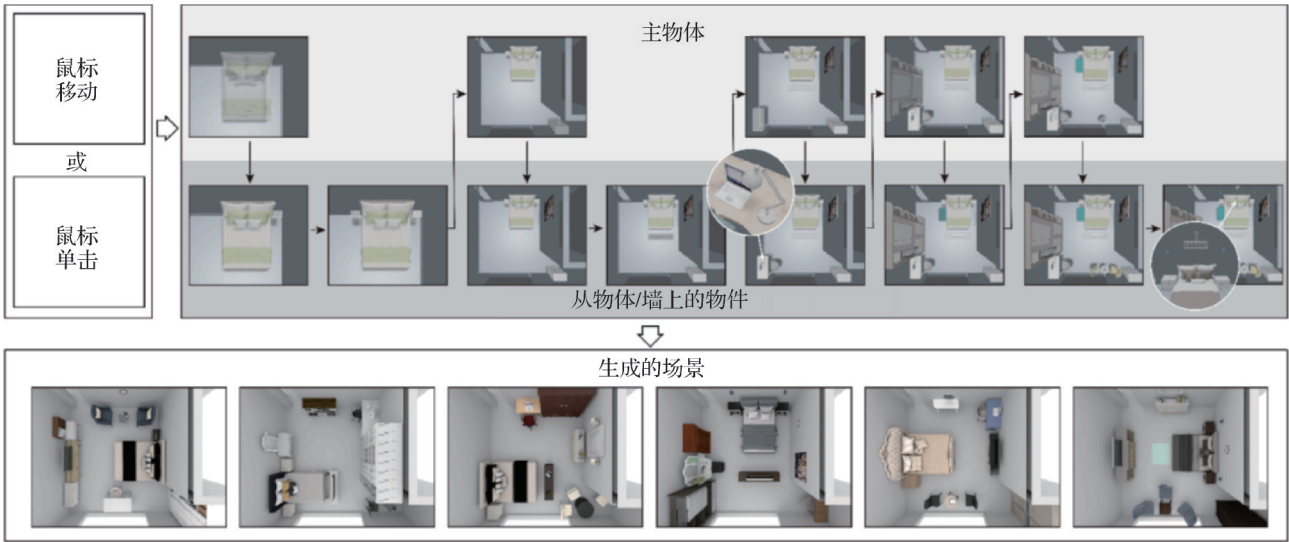


图 15 MageAdd:交互式的实时三维场景自动布局

Fig. 15 MageAdd: interactive real-time automatic layout of 3D scene

5.2 被动交互

基于主动交互的方法,用户通过与场景交互主动干预场景的构建过程,在交互过程中将个人偏好引入到场景当中,从而得出满意的构建结果;在被动交互的方法中,系统将从用户的行为轨迹、个人能力、工作习惯以及身高信息等方面学习用户的个人偏好,自动构建出符合用户偏好的场景,同时亦提高

用户在具体场景中的工作效率。

Liang 等人(2019)的工作记录了用户在 VR 环境中完成任务时的行为习惯,从中学习用户的个人偏好,用于调整工作场景的布局方式,继而提高用户在此布局环境下的工作效率。此方法首先记录用户在一个初始场景中工作时的行为模式,包括移动轨迹与负重情况;针对新构建的布局场景,模拟用户的工

作模式偏好来进行任务,找出能使用户在工作台之间行走以及搬运重物距离最短的布局方式。除了用户偏好约束,方法亦考虑了其他空间约束来调整地面设备的摆放位置,例如物体的可达性、其周围的自由空间大小等。而针对挂于墙面上的物体,方法考虑到物体可见性、用户与物体交互时的姿态等约束,利用REBA(rapid entire body assessment)评分来调整物体位置,避免用户做出容易疲倦或受伤的姿势。最后使用模拟退火算法找出最优解,每次迭代考虑的操作包括平移、旋转以及设备位置对调。类似地,Liu 等人(2021)通过学习用户在工作桌面上的行为习惯,为用户计算出其桌面上物品最优的个性化摆放方案,这里的优化约束还考虑到物体的堆叠、摆放的紧凑程度等。

除了对工作区进行构建,Zhang 等人(2021d)还根据员工的个人工作能力联合计算出最优的工作分配方案。此方法在分配任务时考虑了员工的个体差异,量化了员工对各项工作任务熟练度、行走速度以及对行走转向的容忍度。针对行动能力较弱的轮椅使用者,其对移动的容忍度将更低,因此在分配任务及设计工作区布局时,更偏向让其能在小范围工作区内完成工作。在搜寻工作区布局和工作分配方案的最优解时,此方法从工作效能、员工工作量以及布局约束3方面入手进行优化。其中,工作效能考虑了按序完成工作的用时、行走路线的拥堵程度以及障碍物情况;员工工作量平衡了每人的行走距离和转向多寡;而布局约束则从贴墙和对齐两方面考虑,促使一些电器在布局时能更贴近墙面,以及使布局更美观。最后在使用模拟退火算法寻找最优解时,交替地对工作区布局和工作分配计划进行优化。视优化目标的侧重点不同,可以得出多种不同意义上最优的布局与分工方案,例如工作效率最高的方案、员工行走距离最短的方案,以及行走路线最不拥堵的方案等。此方法能结合员工的个人能力和工作需求,订制出适合不同目标的工作区布局与分工方案。

表3总结了室内三维场景的交互布局的各项工作的交互方式和其中基于数据驱动学习的内容。除了上述针对现实场景中工作区构建的工作外,三维场景的被动交互合成也应用在VR场景当中。Dong 等人(2021)提出的方法能因现实场地的大小以及障碍物位置而调整VR场景中物体的布局,使得用户

在VR场景中漫游时能够没有碰撞地连续行走,不会因为触碰到现实边界或障碍物而被打断体验。此方法因应用户身处的物理空间情况改变了虚拟空间之中物体的布局。具体而言,它基于第一人称视角的约束计算了一个虚空间到物理空间的映射,考虑了物体的缩放比例、映射位置、相邻物体的位置关系、用户行走空间以及物体可见性。

表3 室内三维场景多通道构建小结
Table 3 Summary of indoor scene multi-channel synthesis

方法	交互方式	机器学习内容
Yan 等人(2017)	主动交互	物体推荐
Savva 等人(2017)	主动交互	物体推荐
Yu 等人(2016)	主动交互	小物体推荐
Zhang 等人(2021c)	主动交互	小物体推荐
Zhang 等人(2021a)	主动交互	物体实时推荐
Liu 等人(2021)	被动交互	场景自适应布局
Liang 等人(2019)	被动交互	场景自适应布局
Zhang 等人(2021d)	被动交互	场景自适应布局

6 结 语

本综述从3个方面着手调研并归纳了3种场景构建的方法:场景的自动布局方法、基于用户交互的场景辅助构建方法与基于多通道、丰富输入的场景自动构建方法。自动构建的核心在于让计算机直接基于当前三维内容给出场景构建结果,可以主要归纳为基于数据驱动学习、基于人为设定的具体规则、基于图结构和基于神经网络等多种技术方法;交互构建的核心在于让用户控制计算机辅助构建场景的过程,可以分为“主动交互”和“被动交互”两个主要问题;多通道构建的核心在于让构建的场景参考输入的图像、文本和点云等输入。

6.1 领域发展前景广阔

对于场景自动构建而言,神经网络技术和各种物体间关联关系提取方法已经让其构建效果可以在实际中应用。从学术的角度,场景构建的效果更多在于对特定布局的调整、布局策略的丰富与模型的简化。首先,特定布局的调整指在特殊的场景中,对物体的布局予以细微调整,使其进一步满足场景需求,如在商业场景中货架的摆放可以不满足美观性

但一定要满足客户动线的功能性(Zhang等,2023a)。其次,布局策略的丰富要求补充更多的物体间摆放策略,例如宴会椅可以和餐桌这种主物体摆放,也可以靠墙摆放以为中心区域留出空间。同时,模型的简化又要求可以用更泛化的模型表达更多的布局策略。综上所述,场景的自动构建本身已经实现了0到1的过程,之后应注重于子问题和相关问题的提出与解决。对于丰富输入的场景构建而言,已有工作已经探索了RGB、RGBD、文本和草图等输入,未来应探索更多潜在的输入形式,如音乐和符号等。对于场景交互构建而言,目前主动交互方法仍局限于鼠标、键盘的输入,基于诸如虚拟现实、增强现实和混合现实等交互场景构建方法还有待在未来探索。例如,基于手势直接操控三维场景。目前被动的交互方法也同样局限于观察场景、在场景中走动和与物体交互等,这本身限制了场景优化的导向,即场景一定会为了满足这些交互形式而做出调整。未来应探索更多的人类在场景中的其他行为,如跑步和长时间交互等。最后,本文主要调研场景构建相关工作,三维场景构建框架目前仅为输入的初始场景到构建后的场景,而场景构建后同样可以引出很多后处理、后期交互的工作,如风格一致性(Liu等,2015)、摄影学视角自动选取(Zhang等,2023c)等。

神经辐射场(neural radiance field,NeRF)和三维高斯泼溅模型(3D Gaussian splatting, 3DGS)都是近年热门的三维模型或场景的数据结构,其新颖的解决思路极大提升了三维目标的渲染效率。实际上,本文中所研究的室内三维场景主要指的是由一系列物体及其平移旋转缩放属性所构成的集合。各个物体分别拥有各自的模型网格,而非整个场景公用一个模型。这样的数据结构虽然在场景的多样性上会有所欠缺,但其更加便于场景的编辑和理解。在未来,如何将新兴的三维重建算法应用于三维场景的构建也是一个十分值得深入探究的问题。

6.2 应用场景不断拓宽

三维场景对于其他领域同样有很重要的价值,如前文所说,构建出的三维场景在工业应用与学术研究都有用处。工业应用中,很自然能联想到数字家装。首先得以应用的案例是户型迁移。在酷家乐的构建平台中,可以将已有的三维场景映射到全新的户型中。同时,电子游戏中对于三维场景的需求逐渐成为开发的瓶颈,尤其是主打室内游玩的电子

游戏。由于游戏行业的场景仍需要人为构建,故游玩的内容永远受到开发成本的限制。若这些内容可以由计算机自动提供,那么可以通过近似无限的游玩内容来加大游戏的可玩性。在设计师与户型交互的过程中,自动构建场景、自动布局场景、自动推荐物体以及自动推荐物体的位置等功能,可为设计师节省可观的工作量。图16展示了Zhang等人(2022)的开源三维场景构建平台,可以运行各种三维场景的布局与构建算法,支持物体的基础增删改查操作、多人多终端协同交互,其嵌入的快速场景构建算法已在构家网络科技有限公司应用。在学术研究中,三维场景的构建同样重要。在虚拟现实领域,Li等人(2020)通过不断构建当前的场景来得到轮椅使用者的训练样例。Handa等人(2016)更是通过引入不同的三维场景来源源不断地获取计算机视觉的数据集,即每一个三维场景都可以渲染出各个视角、各个

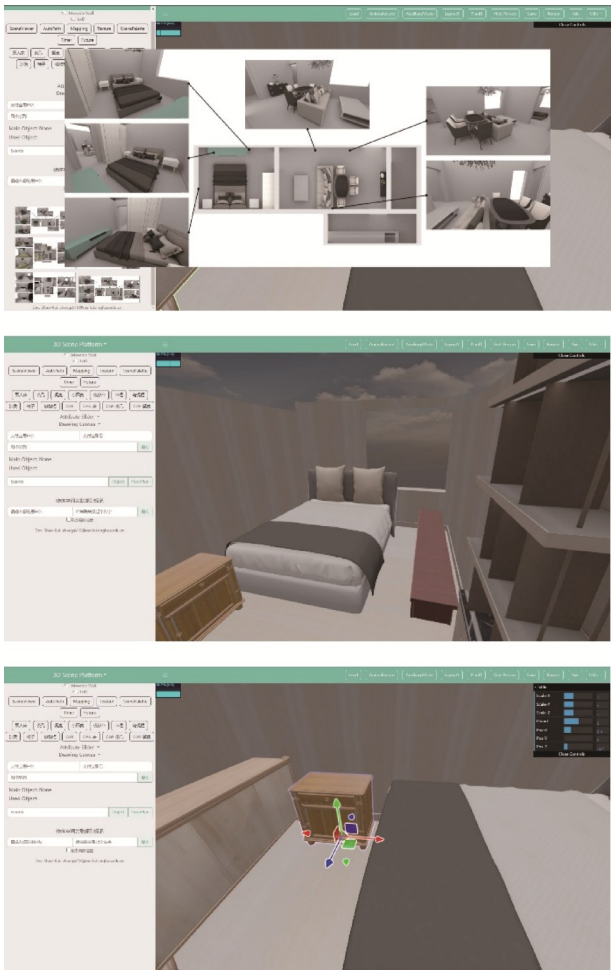


图 16 开源的三维场景构建平台(Zhang 等, 2022)
Fig. 16 Open-source platform for academical use developed (Zhang et al. , 2022)

通道图像,这样可以大幅度减少获取图像所需的成本。

6.3 机遇与挑战并存

这个领域的研究目前仍困难重重。首先,三维场景本身作为一种数据形式,结构复杂且不统一,到目前为止尚无一个学术界均认同的、统一的数据格式(Fu等,2021;Song等,2017)。门洞、窗洞的标注也一直需要研究人员自行得到,甚至一个门可能同时属于两个房间,但已有数据集都是在图3的结构上将门归属于某一个房间。同时,三维的几何本身存在版权问题,故难以像图像、文本数据集那样被大量获取并将其应用。这些都限制了三维场景构建领域的发展。因此,风格更丰富、模型数量更多、户型更多以及结构更为完善的三维场景数据集是该领域日后可以持续发展的基础。

更多围绕三维场景构建的研究同样在兴起,也会是之后值得探索的方向。前文探讨了基于交互的场景构建,若认为交互本身也是一种丰富输入,则目前场景构建的交互仍局限于鼠标、键盘的输入,基于诸如虚拟现实、增强现实、混合现实等交互场景构建方法还有待在未来探索。同时,三维场景构建框架目前仅为输入的初始场景到构建后的场景,而场景构建后可以引入更多的后处理,如风格一致性、摄影学视角自动选取等。

本文总结了数字室内三维场景构建的主要工作。分为数字三维场景自动构建、丰富输入的三维场景构建和数字三维场景交互构建3个方面,并分别讨论了各项工作的核心技术与应用场景。对于场景自动构建而言,神经网络技术和各种物体间关联关系提取方法等各种方法的兴起,已经让其构建效果可以在实际中投入使用。随着点云识别、自然语言处理等相关技术的不断发展,数字三维场景自动构建逐渐可以利用更丰富的输入类型来驱动。场景的交互构建也从主动交互和被动交互两个方面提升用户的场景构建体验。实际上三维场景在学术界和工业界已经有着十分广泛的应用。然而,复杂且不统一的数据结构、版权等問題也在制约着数字三维场景构建技术的快速发展。这项技术的未来发展,需要各国、各高校及各领域研究者的共同努力。

参考文献(References)

- AlHalawani S, Yang Y L, Wonka P and Mitra N J. 2014. What makes London work like London? Computer Graphics Forum, 33(5): 157-165 [DOI: 10.1111/cgf.12441]
- Chang A, Savva M and Manning C D. 2014. Learning spatial knowledge for text to 3D scene generation//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: ACL: 2028-2038 [DOI: 10.3115/v1/D14-1217]
- Chang A X, Eric M, Savva M and Manning C D. 2017. SceneSeer: 3D scene design with natural language [EB/OL]. [2024-02-03]. <https://arxiv.org/pdf/1703.00050.pdf>
- Ching F D K and Binggeli C. 2018. Interior Design Illustrated. 4th ed. Hoboken, USA: John Wiley and Sons
- Chen K, Lai Y K and Hu S M. 2015. 3D indoor scene modeling from RGB-D data: a survey. Computational Visual Media, 1(4): 267-278 [DOI: 10.1007/s41095-015-0029-x]
- Chen K, Lai Y K, Wu Y X, Martin R and Hu S M. 2014. Automatic semantic modeling of indoor scenes from low-quality RGB-D data using contextual information. ACM Transactions on Graphics, 33(6): #208 [DOI: 10.1145/2661229.2661239]
- Deussen O, Hanrahan P, Lintermann B, Měch R, Pharr M and Prusinkiewicz P. 1998. Realistic modeling and rendering of plant ecosystems//Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA: Association for Computing Machinery: 275-286 [DOI: 10.1145/280814.280898]
- Dong Z C, Wu W M, Xu Z H, Sun Q, Yuan G J, Liu L G and Fu X M. 2021. Tailored reality: perception-aware scene restructuring for adaptive VR navigation. ACM Transactions on Graphics, 40(5): #193 [DOI: 10.1145/3470847]
- Fang C, Hu X T, Luo K M and Tan P. 2023. Ctrl-Room: controllable text-to-3D room meshes generation with layout constraints [EB/OL]. [2024-02-06]. <https://arxiv.org/pdf/2310.03602.pdf>
- Fisher M, Savva M, Li Y Y, Hanrahan P and Nießner M. 2015. Activity-centric scene synthesis for functional 3D scene modeling. ACM Transactions on Graphics, 34(6): #179 [DOI: 10.1145/2816795.2818057]
- Foo L G, Rahmani H and Liu J. 2023. AI-generated content (AIGC) for various data modalities: a survey [EB/OL]. [2024-02-03]. <https://arxiv.org/pdf/2308.14177.pdf>
- Fu H, Cai B W, Gao L, Zhang L X, Wang J M, Li C, Zeng Q X, Sun C Y, Jia R F, Zhao B Q and Zhang H. 2021. 3D-FRONT: 3D furnished rooms with layOuts and semaNTics//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 10933-10942 [DOI: 10.1109/iccv48922.2021.01075]

- Fu Q, Chen X W, Wang X T, Wen S J, Zhou B and Fu H B. 2017. Adaptive synthesis of indoor scenes via activity-associated object relation graphs. *ACM Transactions on Graphics*, 36 (6) : #201 [DOI: 10.1145/3130800.3130805]
- Fu Q, Fu H B, Yan H, Zhou B, Chen X W and Li X M. 2020. Human-centric metrics for indoor scene assessment and synthesis. *Graphical Models*, 110: #101073 [DOI: 10.1016/j.gmod.2020.101073]
- Génevaux J D, Galin É, Guérin E, Peytavie A and Benes B. 2013. Terrain generation using procedural models based on hydrology. *ACM Transaction on Graphics*, 32 (4) : #143 [DOI: 10.1145/2461912.2461996]
- Handa A, Patraucean V, Badrinarayanan V, Stent S and Cipolla R. 2016. Understanding RealWorld indoor scenes with synthetic data// *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 4077-4085 [DOI: 10.1109/cvpr.2016.442]
- He Y, Liu Y T, Jin Y H, Zhang S H, Lai Y K and Hu S M. 2022. Context-consistent generation of indoor virtual environments based on geometry constraints. *IEEE Transactions on Visualization and Computer Graphics*, 28 (12) : 3986-3999 [DOI: 10.1109/tvcg.2021.3111729]
- Höllein L, Cao A, Owens A, Johnson J and Nießner M. 2023. Text2Room: extracting textured 3D meshes from 2D text-to-image models//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 7909-7920 [DOI: 10.1109/ICCV51070.2023.00727]
- Hu F, Ma Y B, Zhong W, Ye L, Yang X Y, Fang L and Zhang Q. 2024. A dataset and benchmark for 3D scene plausibility assessment. *IEEE Transactions on Multimedia*, 26: 6529-6541 [DOI: 10.1109/TMM.2024.3353456]
- Hu R Z, Huang Z Y, Tang Y H, van Kaick O, Zhang H and Huang H. 2020. Graph2Plan: Learning floorplan generation from layout graphs. *ACM Transactions on Graphics*, 39 (4) : #118 [DOI: 10.1145/3386569.3392391]
- Lei J B, Tang J P and Jia K. 2023. RGBD2: Generative scene synthesis via incremental view inpainting using RGBD diffusion models//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 8422-8434 [DOI: 10.1109/CVPR52729.2023.00814]
- Li M Y, Patil A G, Xu K, Chaudhuri S, Khan O, Shamir A, Tu C H, Chen B Q, Cohen-Or D and Zhang H. 2019. GRAINS: generative recursive autoencoders for INdoor scenes. *ACM Transactions on Graphics*, 38(2): #12 [DOI: 10.1145/3303766]
- Li W W, Talavera J, Samayoa A G, Lien J M and Yu L F. 2020. Automatic synthesis of virtual wheelchair training scenarios//*Proceedings of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces*. Atlanta, USA: IEEE: 539-547 [DOI: 10.1109/VR46266.2020.00075]
- Liang W, Liu J J, Lang Y N, Ning B and Yu L F. 2019. Functional workspace optimization via learning personal preferences from virtual experiences. *IEEE Transactions on Visualization and Computer Graphics*, 25(5): 1836-1845 [DOI: 10.1109/tvcg.2019.2898721]
- Liang Y, Zhang S H and Martin R R. 2017. Automatic data-driven room design generation//*Proceedings of the 3rd International Workshop on Next Generation Computer Animation Techniques*. Bournemouth, UK: Springer: 133-148 [DOI: 10.1007/978-3-319-69487-0_10]
- Liu J J, Liane W, Ning B and Mao T. 2021. Work surface arrangement optimization driven by human activity//*Proceedings of 2021 IEEE Virtual Reality and 3D User Interfaces*. Lisboa, Portugal: IEEE: 270-278 [DOI: 10.1109/vr50410.2021.00049]
- Liu T Q, Hertzmann A, Li W and Funkhouser T. 2015. Style compatibility for 3D furniture models. *ACM Transactions on Graphics*, 34(4): #85 [DOI: 10.1145/2766898]
- Ma R, Patil A G, Fisher M, Li M Y, Pirk S, Hua B S, Yeung S K, Tong X, Guibas L and Zhang H. 2018. Language-driven synthesis of 3D scenes from scene databases. *ACM Transactions on Graphics*, 37(6): #212 [DOI: 10.1145/3272127.3275035]
- Merrell P, Schkufza E and Koltun V. 2010. Computer-generated residential building layouts. *ACM Transactions on Graphics*, 29(6): #181 [DOI: 10.1145/1882261.1866203]
- Merrell P, Schkufza E, Li Z Y, Agrawala M and Koltun V. 2011. Interactive furniture layout using interior design guidelines. *ACM Transactions on Graphics*, 30 (4) : #87 [DOI: 10.1145/2010324.1964982]
- Nan L L, Xie K and Sharf A. 2012. A search-classify approach for cluttered indoor scene understanding. *ACM Transactions on Graphics*, 31(6): #137 [DOI: 10.1145/2366145.2366156]
- Nauata N, Chang K H, Cheng C Y, Mori G and Furukawa Y. 2020. House-GAN: relational generative adversarial networks for graph-constrained house layout generation//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 162-177 [DOI: 10.1007/978-3-030-58452-8_10]
- Nie Y Y, Dai A, Han X G and Nießner M. 2023. Learning 3D scene priors with 2D supervision//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 792-802 [DOI: 10.1109/CVPR52729.2023.00083]
- Paschalidou D, Kar A, Shugrina M, Kreis K, Geiger A and Fidler S. 2021. ATISS: autoregressive transformers for indoor scene synthesis//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. [s.l.]: [s.n.]: 12013-12026
- Peng C H, Yang Y L, Bao F, Fink D, Yan D M, Wonka P and Mitra N J. 2016. Computational network design from functional specifications. *ACM Transactions on Graphics*, 35 (4) : #131 [DOI: 10.1145/2897824.2925935]
- Purkait P, Zach C and Reid I. 2020. SG-VAE: scene grammar variational autoencoder to generate new indoor scenes//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK:

- Springer: 155-171 [DOI: 10.1007/978-3-030-58586-0_10]
- Qi S Y, Zhu Y X, Huang S Y, Jiang C F F and Zhu S C. 2018. Human-centric indoor scene synthesis using stochastic grammar//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5899-5908 [DOI: 10.1109/cvpr.2018.00618]
- Ritchie D, Wang K and Lin Y A. 2019. Fast and flexible indoor scene synthesis via deep convolutional generative models//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 6182-6190 [DOI: 10.1109/cvpr.2019.00634]
- Savva M, Chang A X and Agrawala M. 2017. SceneSuggest: context-driven 3D scene design [EB/OL]. [2024-02-03]. <https://arxiv.org/pdf/1703.00061.pdf>
- Song S R, Yu F, Zeng A, Chang A X, Savva M and Funkhouser T. 2017. Semantic scene completion from a single depth image//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1746-1754 [DOI: 10.1109/cvpr.2017.28]
- Tang J P, Nie Y Y, Markhasin L, Dai A, Thies J and Nießner M. 2023. DiffuScene: denoising diffusion models for generative indoor scene synthesis [EB/OL]. [2024-02-06]. <https://arxiv.org/pdf/2303.14207.pdf>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin, I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc: 6000-6010
- Wang H. 1961. Proving theorems by pattern recognition—II. Bell System Technical Journal: 40(1): 1-41 [DOI: 10.1002/j.1538-7305.1961.tb03975.x]
- Wang K, Lin Y A, Weissmann B, Savva M, Chang A X and Ritchie D. 2019. PlaniT: planning and instantiating indoor scenes with relation graph and spatial prior networks. ACM Transactions on Graphics, 38(4): #132 [DOI: 10.1145/3306346.3322941]
- Wang K, Savva M, Chang A X and Ritchie D. 2018. Deep convolutional priors for indoor scene synthesis. ACM Transactions on Graphics, 37(4): #70 [DOI: 10.1145/3197517.3201362]
- Wang X P, Yeshwanth C and Nießner M. 2021. SceneFormer: indoor scene generation with transformers//Proceedings of 2021 International Conference on 3D Vision. London, United Kingdom: IEEE: 106-115 [DOI: 10.1109/3DV53792.2021.00021]
- Weiss T, Litteneker A, Duncan N, Nakada M, Jiang C F F, Yu L F and Terzopoulos D. 2019. Fast and scalable position-based layout synthesis. IEEE Transactions on Visualization and Computer Graphics, 25(12): 3231-3243 [DOI: 10.1109/tvcg.2018.2866436]
- Wu W M, Fan L B, Liu L G and Wonka P. 2018. MIQP-based layout design for building interiors. Computer Graphics Forum, 37(2): 511-521 [DOI: 10.1111/cgf.13380]
- Xu K, Chen K, Fu H B, Sun W L and Hu S M. 2013. Sketch2Scene: sketch-based co-retrieval and co-placement of 3D models. ACM Transactions on Graphics, 32(4): #123 [DOI: 10.1145/2461912.2461968]
- Xu W Z, Wang B and Yan D M. 2015. Wall grid structure for interior scene synthesis. Computers and Graphics, 46: 231-243 [DOI: 10.1016/j.cag.2014.09.032]
- Yan M, Chen X J and Zhou J. 2017. An interactive system for efficient 3D furniture arrangement//Proceedings of the Computer Graphics International Conference. Yokohama, Japan: ACM: #29 [DOI: 10.1145/3095140.3095169]
- Yang H T, Zhang Z W, Yan S M, Huang H B, Ma C Y, Zheng Y, Bajaj C and Huang Q X. 2021a. Scene synthesis via uncertainty-driven attribute synchronization//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5630-5640 [DOI: 10.1109/iccv48922.2021.00558]
- Yang M J, Guo Y X, Zhou B and Tong X. 2021b. Indoor scene generation from a collection of semantic-segmented depth images//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 15203-15212 [DOI: 10.1109/ICCV48922.2021.01492]
- Yang Y L, Wang J, Vouga E and Wonka P. 2013. Urban pattern: layout design by hierarchical domain splitting. ACM Transactions on Graphics, 32(6): #181 [DOI: 10.1145/2508363.2508405]
- Ye S F, Wang Y X, Li J M, Park D, Liu C K, Xu H Z and Wu J J. 2022. Scene synthesis from human motion//Proceedings of 2022 SIGGRAPH Asia Conference Papers. Daegu, Korea(South): Association for Computing Machinery: #26 [DOI: 10.1145/3550469.3555426]
- Yeh Y T, Yang L F, Watson M, Goodman N D and Hanrahan P. 2012. Synthesizing open worlds with constraints using locally annealed reversible jump MCMC. ACM Transactions on Graphics, 31(4): #56 [DOI: 10.1145/2185520.2185552]
- Yu L F, Yeung S K, Tang C K, Terzopoulos D, Chan T F and Osher S. 2011. Make it home: automatic optimization of furniture arrangement. ACM Transaction on Graphics, 30(4): #86 [DOI: 10.1145/2010324.1964981]
- Yu L F, Yeung S K and Terzopoulos D. 2016. The clutterpalette: an interactive tool for detailing indoor scenes. IEEE Transactions on Visualization and Computer Graphics, 22(2): 1138-1148 [DOI: 10.1109/tvcg.2015.2417575]
- Zhang S H, Zhang S K, Liang Y and Hall P. 2019. A survey of 3D indoor scene synthesis. Journal of Computer Science and Technology, 34(3): 594-608 [DOI: 10.1007/s11390-019-1929-5]
- Zhang S H, Zhang S K, Xie W Y, Luo C Y, Yang Y L and Fu H B. 2022. Fast 3D indoor scene synthesis by learning spatial relation priors of objects. IEEE Transactions on Visualization and Computer Graphics, 28(9): 3082-3092 [DOI: 10.1109/tvcg.2021.3050143]
- Zhang S K, Li Y X, He Y, Yang Y L and Zhang S H. 2021a. MageAdd:

- real-time interaction simulation for scene synthesis//Proceedings of the 29th ACM International Conference on Multimedia. [s.l.] : ACM: 965-973 [DOI: 10.1145/3474085.3475194]
- Zhang S K, Liu J H, Li Y K, Xiong T Y, Ren K X, Fu H B and Zhang S H. 2023a. Automatic generation of commercial scenes//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 1137-1147 [DOI: 10.1145/3581783.3613456]
- Zhang S K, Tam H, Li Y K, Ren K X, Fu H B and Zhang S H. 2023b. SceneDirector: interactive scene synthesis by simultaneously editing multiple objects in real-time. IEEE Transactions on Visualization and Computer Graphics: #3268115 [DOI: 10.1109/TVCG.2023.3268115]
- Zhang S K, Tam H, Li Y X, Mu T J and Zhang S H. 2023c. SceneViewer: automating residential photography in virtual environments. IEEE Transactions on Visualization and Computer Graphics, 29(12): 5523-5537 [DOI: 10.1109/tvcg.2022.3214836]
- Zhang S K, Xie W Y and Zhang S H. 2021b. Geometry-based layout generation with hyper-relations AMONG objects. Graphical Models, 116: #101104 [DOI: 10.1016/j.gmod.2021.101104]
- Zhang S Y, Han Z Z, Lai Y K, Zwicker M and Zhang H. 2021c. Active arrangement of small objects in 3D indoor scenes. IEEE Transactions on Visualization and Computer Graphics, 27(4): 2250-2264 [DOI: 10.1109/tvcg.2019.2949295]
- Zhang S Y, Han Z Z and Zhang H. 2016. User guided 3D scene enrichment//Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry. Zhuhai China: ACM: 353-362 [DOI: 10.1145/3013971.3014002]
- Zhang Y Q, Huang H K, Plaku E and Yu L F. 2021d. Joint computational design of workspaces and workplans. ACM Transactions on Graphics, 40(6): #228 [DOI: 10.1145/3478513.3480500]
- Zhang Z W, Yang Z P, Ma C Y, Luo L J, Huth A, Vouga E and Huang Q X. 2020. Deep generative modeling for scene synthesis via hybrid representations. ACM Transactions on Graphics, 39(2): #17 [DOI: 10.1145/3381866]
- Zheng J, Zhang J, Li J, Tang R, Gao S and Zhou Z. 2020. Structured3D: a large photo-realistic dataset for structured 3D modeling//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 519-535 [DOI: 10.1007/978-3-030-58545-7_30]
- Zhu F, Liu L, Xie J, Shen F M, Shao L and Fang Y. 2018. Learning to synthesize 3D indoor scenes from monocular images//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 501-509 [DOI: 10.1145/3240508.3240700]
- Zhu S Y, Wang X J, Wang M, Wang Y C, Wei Z Q, Yin B and Jin X G. 2022. Example-based large-scale marine scene authoring using Wang Cubes. Visual Informatics, 6(3): 23-34 [DOI: 10.1016/j.visinf.2022.05.004]

作者简介

岳亮,男,硕士研究生,主要研究方向为计算机图形学、三维场景的构建与交互。E-mail: yuel22@mails.tsinghua.edu.cn

张少魁,通信作者,男,博士后,主要研究方向为计算机图形学、三维场景的构建与交互。E-mail: shaokui@tsinghua.edu.cn

谈皓,男,硕士研究生,主要研究方向为计算机图形学。

E-mail: th21@mails.tsinghua.edu.cn

黄俊凯,男,本科生,主要研究方向为计算机图形学。

E-mail: huangjk21@mails.tsinghua.edu.cn