

大规模城市场景建模与理解

陈宝权 万国伟
山东大学

关键词：场景重建 场景理解 自动扫描 智能建模

引言

我国城镇化进程正在不断向前推进，人口和资源不断向城市聚合，形成了北京、上海、广州、深圳等超大城市。城市化给人们的生活带来了便利，但也带来了人口膨胀、交通拥堵、环境恶化、事故频发等“城市病”，在很大程度上加重了城市的负担、制约了城市的发展且构成引发不稳定的因素。近年来提出的“智慧城市”理念为解决上述问题提供了思路。智慧城市运用物联网、云计算、大数据、城市三维场景等新一代信息技术，整合城市运行的各项关键信息，在城市服务、公共安全、工商业活动、环境保护等领域对出现的各种状况做出智能响应，实现城市的智慧化管理和运行，为人们创造更美好的生活。

城市三维场景数据为智慧城市提供完整的空间信息，包括建筑、桥梁、道路、树木等城市主要组成要素。智慧城市的应用涵盖各式各样的需求，其

对三维场景的需求可以归结为真实、完整、高精度。传统城市场景建模研究主要针对大尺度场景的三维化，采用卫星遥感观测、低空航拍、机载激光雷达扫描等技术获取三维数据，重建后的模型精度较低，越来越难以满足智慧城市的应用需求。本文的研究主要针对中尺度城市场景重建，采用车载激光雷达、全景相机作为数据采集设备，模型重建的精度更高，包含更多细节。

大规模复杂场景带来的挑战

传统三维数字化建模工作主要是针对单个物体扫描，扫描所得三维点云非常密集且完整。扫描所得数据满足采样理论所需的密度，可以由此重建数字化几何模型。我们的工作是针对室外多物体场景，采用车载激光点云扫描技术采集数据后建模，由于扫描车路线受限以及难以避开障碍物等因素，所得到的点云数据往往非常稀疏，且缺失严重。根据采样理论，用我们所得到的扫描数据重构三维几何模型面临极大的挑战，传统算法无法胜任。面对稀疏且缺失严重的点云数据，我们通过构建物体几何规则和场景先验知识，来指导缺失数据的重建。随之而来的问题是如何处理物体的多样性与场景的复杂性。

复杂场景的几何内容丰富多样，各类物体种类非常繁杂，场景也在不断变化更新。这些特点给三维模型的重建带来了以下技术问题：(1) 针对同类别的物体，如何构建共性特征和先验知识来化繁为简，提高同类物体处理的有效性；(2) 如何建立主动式扫



图1 稀疏点云三维重建

描机制来实现无监督的三维建模；(3) 针对几何特征的多样性，如何提取其内在规则，并通过规则来描述这些特征。

面对这些挑战，我们提出了相应的解决思路：(1) 基于共性特征与先验知识，实现建筑和植物两大类物体的高效重建；(2) 利用机器人平台，采用主动式扫描和增量式学习，实现自动扫描与智能建模；(3) 基于内在几何规则，实现场景层次化语义构建。

建筑和植物的三维重建

建筑和植物是城市中最常见的两类实体，其三维模型也是城市三维场景的主要构成要素。建筑属人工产物，包含大量重复结构，且分布具有一定的规则性。植物属自然产物，其种类繁多，姿态千差万别，结构特征很强，三维结构复杂。

建筑三维重建

我们利用建筑本身的先验知识，构建其共性特征，针对因激光扫描点云数据稀疏或缺失的模型拟合难题，实现了高精度的三维建模。我们还引入了图像数据，利用图像和点云的互补性特点，融合点云和图像，实现了高质量的建筑物墙面建模。

在早期的研究中，激光点云的获取多采用静态激光扫描仪，将扫描仪固定在建筑物周边的数个位置，得到独立的点云数据，再将多个点云数据注册到统一坐标系下。由于室外环境所限，要得到建筑物每个面的点云数据是不现实的。我们在文献 [1] 中提出了一种针对由平面组成的建筑物的三维重建方法，首先从稀疏点云中识别平面区域，然后计算平面之间的交线以及角点，最终得到一个完整的多边形。其核心算法是对稀疏且缺失的点云数据进行聚类得到平面集合，并在平面上提取边界线、多边形。图 1 给出了稀疏缺失的三维点云、聚类后的点云以及重建后的三维模型。

定点激光扫描技术获取三维点云的操作繁琐，且须在后期对扫描点云进行注册，这样势必会降低整个三维重建流程的效率。车载激光点云扫描技术的出现解决了上述问题，通过惯性导航技术以及后期的基站校准技术可轻松获取统一坐标系下的三维点云。但因扫描车扫描路径所限，无法避开遮挡物，导致获取的点云数据较定点扫描技术更稀疏且缺失更严重。针对这种低质量的点云数据，我们提出了一种交互式快速建筑物建模方法 SmartBoxes^[2]，首先用户以二维动态框选形式构建初始 SmartBox 图元（立方体以及多边形组成的三维模型），然后用组合、拖放、填充等交互操作将初始图元复制到建筑物相似区域。算法通过数据项（图元与点云匹配）和上下文项（相邻图元大小、边对齐、间距分布）的双重约束优化图元的最佳位置及大小。用户在交互过程中隐含地使用了建筑物所具有的平面结构性、重复性以及相似性特征等先验知识，解决了因点云稀疏以及缺失而带来的问题；构建初始 SmartBox 图元采用二维操作方式，简化了用户在三维界面下操作的繁琐，消除了歧义性。图 2(a) 描绘了一个从基本图元（立方体、矩形）构建 SmartBox 图元的过程，通过简单的

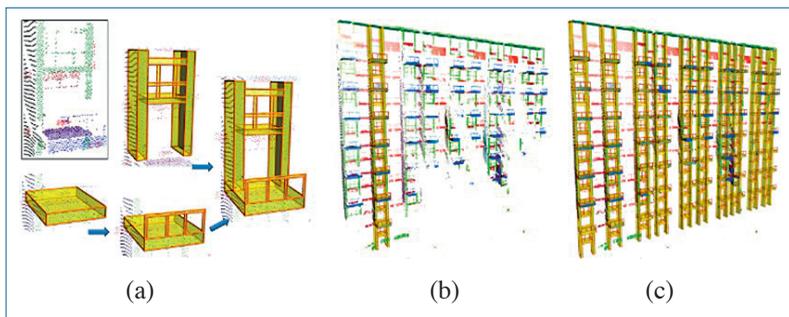


图2 交互式快速建筑物建模方法SmartBoxes

交互操作可从低质量点云中迅速重建出阳台、窗户等，并组合得到 SmartBox 图元；(b) 图描绘了通过拖放操作在竖直方向得到一个组合 SmartBox 图元；(c) 图描绘了将组合 SmartBox 图元作为一个整体适配到整个墙面的过程。我们根据用户对建筑物的先验知识以及建筑物本身的相似性、重复性特征，在点云稀疏甚至缺失区域重建了正确的几何结构，最终构建了一个完

整的建筑物墙面模型。

车载雷达扫描技术可以快速得到较大范围建筑物的三维点云数据,但点云具有嘈杂、稀疏以及大片缺失等特点。二维图像虽缺乏三维信息,但具有高分辨率、无噪声以及容易获得较完整的数据(例如整个建筑物墙面)等特点,可作为三维点云的补充。综合二维图像和三维点云的优点,我们提出了一种融合二维图像和三维点云的分层建筑物墙面重建方法^[3]。算法首先将二维图像注册到三维点云的坐标系下,对三维点云进行墙面深度层分割,并将三维点云投影到图像上;然后给图像上的每个像素赋予一个深度层,该问题可转化为一个多标记赋值问题;得到图像的深度层分割后,在每一个图像深度层上检测重复性结构,并利用重复性信息增强三维点云;最后建立相邻深度层的三维连接几何形状结构,得到最终的三维模型,并将图像的纹理信息映射到模型上。图3给出了三维点云与二维图像、注册后的点云和图像、重建后的建筑物三维模型以及粘贴纹理之后的模型。



图3 融合点云和图像建筑物三维重建

植物三维重建

针对植物的三维建模研究已有几十年的历史,过程式建模方法可以利用事先设计好的规则或语法描述虚拟树木的三维模型,但对于给定的模型或真实树木目前还很难得到简洁的规则或语法描述;而基于实际采集数据(主要是图像和点云)的建模方法,一般可得到低层次而且大数据量的模型描述,如三角网格模型。

我们在文献[4]中提出了一种基于激光点云的自动树木骨架重建方法。算法的核心思想是通过一系列全局优化方法在稀疏的、不完整的、嘈杂的点

云中适配树木的骨架结构。该算法无须对点云进行分割即可重建互相交叠的树枝结构。图4(a)给出了树木的二维图像;(b)图给出了激光扫描点云;(c)图给出了重建后的树木骨架模型。

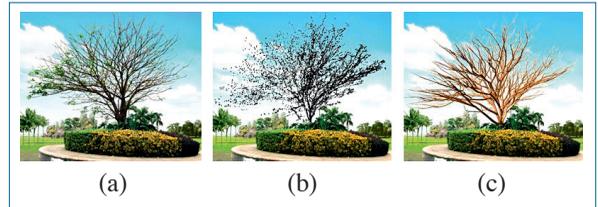


图4 基于激光点云的自动树木骨架重建

上述方法须进行一系列的全局优化才能得到最终的骨架模型,重建一棵树需要十多分钟的时间。要实现大规模场景中对众多树木快速建模,需要对场景树木点云数据进行完全自动化识别和建模,提高每棵树的重建速度、减少树木模型占用的存储空间。树木虽然具有极其复杂的几何外形结构、枝叶繁多的细节特征,但是在有限的视觉范围内下,人对树木的认知主要在于树木的主枝干结构和树冠的整体形状。其中树冠形状可以看作由一个个树枝构成的外轮廓(我们称之为“Lobe”)组成,而每个Lobe内部不同小枝的枝叶结构具有很强的相似性和重复性,而且这种相似结构具有不断细分的层次。基于这一观察,我们提出了一种基于Lobe表示的树木快速三维建模方法^[5]。首先进行树种识别,对给定树木点云数据,提取各种几何、统计特征,通过JointBoost分类器识别树种种类;然后分别对不同的树种预先建立枝叶模板库,包含不同树种的信息,例如树种名称、树种的一些几何特征以及用于几何重建的枝叶片段;最后是模型重建:对于主枝干,采用文献[4]提出的方法生成几何模型;而对于Lobe内的几何模型,借鉴基于块的纹理合成思想,在空间约束情况下,将枝叶模板库中的枝叶片段连接合成一个可以充分填充Lobe的树枝模型,如图5所示。

在植物学领域,研究植物生长过程具有重要意义。我们提出了一种基于四维点云的植物生长分析方法^[6]。采用三维扫描技术对生长植物按设定的时间间隔进行持续扫描以获取四维点云。这种四维数

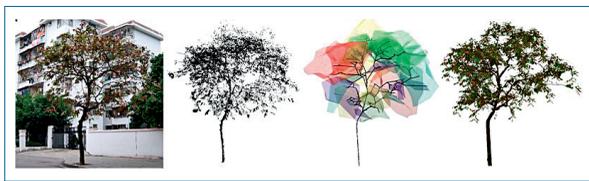


图5 基于Lobe表示的树木三维建模

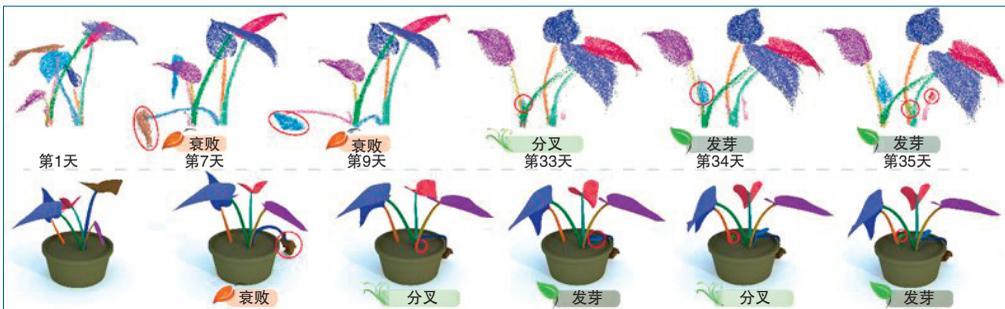


图6 基于四维点云的植物生长分析

据完整记录了植物生长过程中精细的形态变化，可以为植物学研究提供以前所未有的精度和采样密度采集的测量数据。四维扫描有望成为植物学研究中全新的测量手段。我们提出了一种新颖的计算框架可以从植物四维点云中自动地识别植物的出芽和分支等生长事件，通过前向-后向分析，实时跟踪植物分支，以确保使用局部自适应阈值检测到正确的时空事件。该成果可用于植物生长状态监控、对比试验组差异测量、植物四维标本库建立、植物生长动画合成和真实感植物动态模拟等。图6展示了滴水观音5周的生长情况，上半部分给出了四维点云，下半部分展示了滴水观音发芽、分叉、衰败的过程。

自动扫描与智能建模

在三维建模过程中，数据获取是一个耗时且繁琐的过程，使用机器人平台实现场景自动扫描是解决该问题的可行方法。然而，快速自动扫描场景，并重建大量形态各异、风格相似的场景模型是一项极具挑战的工作。

机器人自动扫描

针对单个物体，我们提出了一种基于质量驱动的泊松引导的全自动高精度三维扫描系统^[7]。不同于以往的扫描规划技术，该系统的目的不是将扫描次数最小化以覆盖扫描对象的表面，而是确保模型的高质量扫描。为了实现这一目标，可通过计算下一个最佳

视角，将扫描仪移动到对应位置获取被扫描对象的几何细节，直到扫描数据达到预期的完整性和精度。对扫描数据的泊松场以及几何关系进行分析，并在此基础

上产生置信图以生成观察向量场，进而计算下一个最佳视角来指导机器人将手臂移动到合适位置进行下一步扫描。这项工作完全由机器人自主完成。图7展示了机器人自动扫描物体的过程，随着扫描仪从多个最佳视角扫描物体，物体的表面数据逐渐增多，最终形成完整的高质量三维物体模型。



图7 基于机器人平台的自动扫描

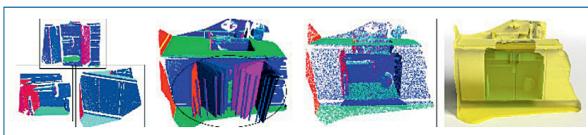


图8 人工主动式扫描过程及建模

使用三维扫描仪获取三维物体模型最大的瓶颈在于，复杂形状的物体往往存在自遮挡和内部隐藏区域。我们提出了一种新颖的主动式扫描方法^[8]，允许用户在扫描场景的同时与场景进行交互，以主动渐进方式将遮挡或隐藏区域暴露出来。我们还提出

了一种扫描注册算法,计算物体的运动轨迹,有效区分由用户操作物体所产生的运动、相机移动产生的运动以及物体自身形变产生的运动区,进而将物体所有静态部件注册并重建完整的三维模型。图8展示了主动式扫描的过程,从左到右分别为扫描过程中的三个独立帧数据、用户交互过程中多帧数据的叠加、注册好的完整的点云数据以及重建的完整三维模型。

室内场景的精细化扫描涉及场景物体之间的遮挡以及物体的自遮挡问题,对于人类而言是一项繁杂的任务。我们提出了利用机器人进行室内场景自动扫描,结合场景分析达到对场景的细致扫描。我们提出了一种以物体为中心的場景分析方法使重建场景更加精细化^[9],驱动机器人进行迭代式场景分析与主动式验证,减少物体间遮挡及物体自遮挡的影响,进而获取更加精细完整的三维场景。该算法结合机器人在线获取的知识以及基于图像的分割算法,实现对场景的在线分析;基于初始物体分割,评估场景物体置信度,驱动机器人主动对低置信度场景交互进行验证并提高分割结果的正确性,渐进式构建精细化室内场景。图9展示了机器人对复杂场景的自动扫描重建及主动式物体分析,从左至右分别为装配 Kinect¹ 摄像头的机器人扫描并与场景交互、重建的完整精细室内场景模型(场景中的物体已被自动分割)、房间部分区域的放大视图。



图9 机器人场景自动扫描

智能建模

在缺乏对目标风格的具体描述时,快速生成大量形态各异、风格相似的模型是极具挑战性的工作。我们提出了一种针对非规则三维建筑模型的保结构变形技术^[10],通过简单的交互操作,用

户可快速生成大量具有相同风格的三维建筑模型。对于输入三维模型,经过预处理操作形成带标记的包围盒层次结构,以此作为变形操作的分析基础。变形算法的核心思想是将原始的复杂结构分解为一组一维的结构序列,并逐条对一维结构序列执行变形操作。在对一维结构序列的变形过程中尽可能以可重复的元素来填充变形空间,实现对输入结构特点的保持。图10是一组具有相同风格建筑物模型的示例。



图10 一组相同风格的建筑物模型

场景理解与语义构建

场景分析旨在利用计算机技术模拟人类视觉系统,使得机器能够从图像、三维场景等可视信息中快速认知物体及其运动属性,从而理解场景所蕴含的三维特性及功能语义。场景理解与语义分析是计算机视觉、计算机图形学等领域的研究热点,其核心在于分析和理解场景的结构。然而,自然场景建筑物中存在大量的规则结构(比如窗户、阳台)和不规则结构(比如欧式风格大门),而且场景对象往往存在大量运动结构(比如可旋转的吊车),现有的数据获取方式难以对其完全表达。因此,自然场景所蕴含的丰富的语义信息给场景分析与理解带来了极大挑战。

针对蕴含丰富语义的复杂自然场景,我们提出了利用对称性、规则性及重复性等内在特征高效表达场景的思想,不仅能够提升计算的精度与鲁棒性,还能构建场景的语义层次化结构,为复杂场景的分析与理解提供了有效的解决思路及技术基础。

¹ 微软XBOX360体感周边外设。

格式塔 (Gestalt) 理论总结了人类认知规则、模式和语义的若干规律, 已有大量的研究针对独立的格式塔规则建立计算模型。我们在文献 [11] 中基于格式塔理论的主要规则建立了多规则耦合的计算模型, 并将该模型用于建筑物线画图的抽象。我们针对输入的建筑物矢量化模型用一幅图描述图画中元素的邻接关系, 然后利用多种格式塔准则检测出一系列备选分组并以此作为优化标签, 用 Graph-cut² 优化图画中重复结构元素的数目或者将复杂的结构用简单结构替换, 从而实现了对建筑物结构的概括和抽象 (如图 11 所示)。该算法可应用于风格各异的建筑物线画图, 并可拓展至其他类型的图画及三维模型。



图 11 基于格式塔的建筑物图像抽象

格式塔理论强调局部对象之间的对称性及规则性在整体认知中的作用, 我们在文献 [12] 中基于对称性驱动层次化分析, 针对不规则建筑物提出了一种高层语义场景结构理解方法。我们利用深度分层及平面分解的方式 (见图 12(b)、(c)), 基于对称性迭代优化建筑物的层级化结构, 直至达到对称最大化的目标; 最后基于局部对称的离散层级化结构, 评估整体对称性。然而, 该算法受限于包含轴对齐结构 (axis-aligned) 及递归二分结构 (recursive binary) 的场景。图 12 描述了墙面正面分析过程, 图 (a) 为非规则建筑输入, 图 (b) 为场景深度层次分析, 图 (c) 为非规则建筑平面分解, 图 (d) 为结

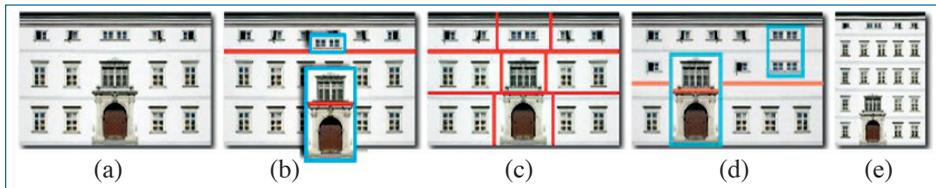


图 12 基于对称最大化的非规则建筑正面几何分析分层

构化的建筑物编辑结果, 图 (e) 为非规则结构重定位结果。

除了规则以及不规则结构, 场景中往往还存在大量运动结构 (比如场景对象的旋转性自由度) 及功能信息 (比如场景对象间的相互关联), 为场景的理解和分析提供了丰富的语义信息。我们在文献 [13] 中针对复杂三维室内场景, 通过自动分析检测场景对象的运动结构及重复性结构, 提出了一种运动信息树 (mobility-tree) 结构的高层功能性场景表达方式。运动分析是指场景对象的运动功能 (比如自由度) 及其子对象在室内场景中的作用。通过分析场景对象的空间布局、重复性以及与其他对象之间的关系, 计算出该场景对象的运动属性, 并保存于运动信息树结构中。场景中的重复性运动对象可组成“运动功能组”, 从而基于语义来加速对场景的高层交互式编辑。图 13 展示了基于运动信息树进行层次化分析对室内场景编辑的结果, 左图为输入场景, 右图为重布局后的场景, 其中书架上书籍的摆放位置、餐桌上梨子的布局以及办公桌抽屉的闭合均发生了合乎语义的变化。



图 13 室内场景编辑中运动信息树层次化分析

结语

大规模城市场景的建模与理解是有效建立非结构化信息的索引与表达的基础, 在智慧城市中对管理决策及信息响应有着重要影响。目前城市场景建模与理解正朝着多维数据融合的方向发展, 已有大批研究课题或应用基于

² Graph-cut是一种十分有用和流行的能量优化算法。

车载或机载激光雷达、卫星定位、图像等多源信息构建城市场景。然而,应用的深入以及技术的发展对城市三维建模提出了更高要求,即场景规模日趋增大、颗粒度日益精细、更新频率逐渐提高。对于大规模场景,如何实现细粒度扫描,并且利用已有语义信息提高更新频率是需要解决的关键问题。我们认为,可利用传统车载激光扫描,结合机器人主动式探索场景动态信息,实现闭环式的三维场景采集处理。利用场景中既有的语义信息,通过人类智慧进行数据标注(包括城市场景的几何关系、材质类型、场景对象行为属性等),有助于快速重构城市三维场景以及动态更新。主动被动相结合的测量手段和众包式数据、信息采集机制为高精度、高动态性的大规模城市三维建模提供了可行的解决思路,是未来实现大规模城市场景建模与理解的重要方向,非常值得关注。■



陈宝权

CCF常务理事,长江学者特聘教授,山东大学计算机学院和软件学院院长。主要研究方向为基于车载移动激光扫描的大规模城市场景三维建模及海量数据可视化。baoquan@computer.org



万国伟

山东大学博士。主要研究方向为基于图像和车载扫描点云的场景三维重建。gwwan.nudt@gmail.com

参考文献

- [1] Chen J, Chen B. Architectural Modeling from Sparsely Scanned Range Data[J]. *International Journal of Computer Vision*, 2008, 78(2-3):223-236.
- [2] Nan L, Sharf A, Zhang H, et al. SmartBoxes for interactive urban reconstruction[J]. *ACM Transactions on Graphics*, 2010, 29(4): 93:1-93:10.
- [3] Li Y, Zheng Q, Sharf A, et al. 2D-3D fusion for layer decomposition of urban facades[C]// *Proceedings of International Conference on Computer Vision*. IEEE Computer Society, 2011:882-889.
- [4] Livny Y, Yan F, Olson M, et al. Automatic reconstruction of tree skeletal structures from point clouds[J]. *ACM Transactions on Graphics*, 2010, 29(6):81-95.
- [5] Cheng Z, Yan F, Baoquan A C. Texture-Lobes for Tree Modelling[J]. *ACM Transactions on Graphics*, 2011, 30(4):76-79.
- [6] Li Y, Fan X, Mitra N J, et al. Analyzing growing plants from 4D point cloud data[J]. *ACM Transactions on Graphics*, 2013, 32(6):1-10.
- [7] Wu S, Sun W, Long P, et al. Quality-driven poisson-guided autoscanning[J]. *ACM Transactions on Graphics*, 2014, 33(6):1-12.
- [8] Yan F, Sharf A, Lin W, et al. Proactive 3D scanning of inaccessible parts[J]. *ACM Transactions on Graphics*, 2014, 33(4):1-8.
- [9] Xu K, Huang H, Shi Y, et al. Autoscanning for coupled scene reconstruction and proactive object analysis[J]. *ACM Transactions on Graphics*, 2015, 34(6):1-14.
- [10] Lin J, Or D C, Hao Z, et al. Structure-Preserving Retargeting of Irregular 3D Architecture[J]. *ACM Transactions on Graphics*, 2012, 30(6): 183:1-183:10.
- [11] Nan L, Sharf A, Ke X, et al. Conjoining Gestalt Rules for Abstraction of Architectural Drawings[J]. *ACM Transactions on Graphics*, 2011, 30(6): 185:1-185:10.
- [12] Zhang H, Xu K, Jiang W, et al. Layered Analysis of Irregular Facades via Symmetry Maximization[J]. *ACM Transactions on Graphics*, 2013, 32(4):96-96.
- [13] Sharf A, Huang H, Liang C, et al. Mobility-Trees for Indoor Scenes Manipulation[J]. *Computer Graphics Forum*, 2013, 33(1):2-14.