
同濟大學

毕业设计(论文)开题报告

(适用于工科类、理科类专业)

课题名称 基于无监督深度学习的土木工程场景三维重建

副标题

学院 软件学院

专业 软件工程

学生姓名 张喆 学号 1754060

2021 年 3 月 15 日

一、毕业设计（论文）课题背景（含文献综述）

1.1 土木工程场景三维重建的意义

建筑信息模型（Building Information Modeling, BIM）是土木工程领域最常用的工程数据模型之一，在开放的 IFC 工业标准下对设施的物理和功能特性及其相关的项目全寿命周期信息进行可计算、可运算的形式表现。BIM 以三维数字技术为基础，因其强大的信息整合能力和良好的数据信息格式等特点，广泛的应用在土木工程领域全生命周期中，尤其对决策支持提供了极大的便利，有效地支撑土木行业各个应用系统之间的数据交换和建筑物全过程的数据管理。

然而由于设计建筑时的图纸和模型与真实施工时的建筑有不少调整、真实施工时工期紧无法完全按照原始 BIM 图纸进行建造等多方面因素影响，在建筑建成后，真实的建筑数字模型与设计阶段的 BIM 模型存在出入，尤其是在细节处的误差难以接受，且这些出入很难在设计阶段原始的 BIM 模型上进行修改。如果决策工程师在设计阶段原始的 BIM 模型基础上进行土木领域的裂缝检测、结构薄弱点预测等关键问题实践，可能无法完全且精确地映射到真实建成的建筑物上，导致仿真结果的不理想；另一方面传统的 BIM 建模方式是通过专业建筑设计工程师进行正向建模，BIM 模型的准确性和精度不仅依赖于建模工程师的经验等主管能力，更十分受限于建筑图纸的规范和完整形，对于图纸残缺的建筑，正向建模难以保证建模的准确性和完整性，而对于大量古建筑，可能存在形状不规则和无法直接接触等诸多问题，导致常规的 BIM 建模方式难以适用。

因此，对已建成的建筑进行 BIM 逆向建模是非常必要的。而在土木工程领域将三维模型转换为 BIM 模型已经较为成熟，因此针对土木工程场景的三维重建得到三维模型尤为重要。

1.2 基于视频进行三维重建的意义

传统的三维建模采用的是三维激光扫描技术，近几年也开始出现基于三维激光扫描的 BIM 逆向建模技术。三维激光扫描技术可以克服上述问题，无论是施工过程中在原始建筑模型上进行了大量变动，还是无法直接获取信息的古建筑等，都可以使用三维激光扫描技术获取精确的位置信息。基于三维激光扫描的 BIM 逆向建模技术将信息模型和三维激光扫描得到的信息进行无缝对接，将扫描到的点云数据与 BIM 模型相结合，进一步优化建筑信息化和管理。但使用三维激光扫描技术面临成本高、专业技术人员少等诸多棘手的问题，在大规模土木工程场景的三维重建背景下难以适用。

三维重建最重要的的是要进行立体匹配，即深度估计，三维激光扫描技术是通过激光直接获取到光飞行的时间，进而通过光速得到距离，即深度信息；而基于多视点立体匹配（Multi-view stereo, MVS）和深度估计的算法在近些年里成为三维重建的热门研究方向。MVS 考虑的是如何通过普通 RGB 相机，通过计算机视觉相关算法获取深度信息，而通过普通相机拍摄的照片进行深度估计和三维重建，可以大大降低三维激光扫描技术所需的成本和专业技术人员。基于多视点的立体匹配采用多台相机同时对待重建场景进行拍摄，通过多台相机的视差估计出深度信息，而为进一步降低成本和复杂性，可以直接使用单台相机拍摄视频，后期通过帧提取的方法模拟多视点，如果环绕单一物体或场景进行拍摄，理论上视频抽帧后有多少帧就相当于多少视点，而视点越多理论上立体匹配和深度估计的结果也越准确。

因此本文提出一种基于视频（多视点图像序列）进行深度估计并进行三维重建，进而逆

向生成 BIM 模型的方法。

1.3 三维重建的常用方法

三维重建的常用方法主要分为两大类：基于传统多视图几何的三维重建方法和基于深度学习的三维重建方法。

基于传统多视图几何的三维重建方法按照传感器是否主动向物体照射光源可以分为主动式和被动式两种方法。

主动式主要有三种不同的方法：

1. 结构光法：依靠投影仪将编码的结构光投射到被摄物体上，然后由摄像头进行拍摄，由于被摄物体的不同部分相对于计算机的距离和方向不同，结构光编码图案的大小和形状也会发生改变。用摄像头捕捉到这种变化的图像，然后将其通过算法换算成深度信息，进而实现三维重建的目的。但该方法容易受环境光干扰，尤其是在土木工程场景下的室外光照变化剧烈，且大型建筑间存在遮挡，很难拍摄到投影出的光斑信息，而且随着检测距离的增加精度也会剧烈下降，在场景较大的情况下难以适用。
2. TOF 激光飞行时间法：通过向待重建物体发射连续的光脉冲，然后依据传感器接收到返回光的时间或相位差计算距目标的距离。该方法虽然可以测量距离较远的物体，且受环境光干扰较小，但该方法需要极为精确的时间测量模块，深度信息的精度直接取决于时间测量模块硬件的精确度，成本较高且鲁棒性较差。
3. 三角测量法：依据三角测距原理，激光器以一定的角度射出激光，照射到物体后光线反射，相机三角形和物体三角形可以通过相似三角形的相关性质进行计算，进而求出深度信息。该方法不需要较为精密的传感器，成本较低，但三角测量法的测量误差与距离有关，且只在距离较近的时候精度较高，随着测量距离的增加，误差也会越来越大，在本课题的土木工程场景下很难适用。

被动式主要有两种不同的方法：

1. 单目视觉方法：依靠一段时间内获得的连续图像的视差重建三维环境，主要有运动恢复法、渐进式方法等，常见的算法有 SfM、REMODE、SVO 等
2. 双目/多目视觉方法：利用在左右相机得到的两幅矫正图像中找到左右图片的匹配点，然后根据几何原理恢复出深度信息，常见的算法有 SGM、SGBM 等

基于传统多视图几何的三维重建算法中最著名的是 COLMAP 算法，它集成了从运动中恢复结构（Structure from Motion, SfM）和 MVS，是从一组不同视角下拍摄的无序或有序图像集合中同时恢复场景三维结构和相机位姿的技术，为图像集合的重建提供了广泛的特征。主要包括多视点数据集采集、稀疏重建、深度图估计和稠密重建四个步骤。该传统方法可用于快速三维重建、代码封装性良好，但深度图估计速度过慢，且在深度图估计精度上一般。在本课题中使用 COLMAP 进行简单的比对实验，观察结果、记录数据，并将其与本课题重点研究的基于无监督深度学习的方法进行比较。

近几年深度学习和人工智能的兴起，使得很多复杂的传统问题有了新的求解途径，而基于深度学习的三维重建方法无疑是深度学习在计算机视觉领域中的一次十分成功的应用。基于深度学习的三维重建算法从单张图像中用神经网络提取出若干基函数表示场景的深度，可以极大地简化传统几何方法中的优化问题。从抽象角度来讲，人类和动物的大脑同样是基于非严格的几何特征来观测物体的深度信息并进行三维重建的，同时对于单张图像，虽然已经

完全丢失深度信息，但人脑仍可以凭借经验进行大致深度的估计，从感官层面深度学习更适用于此非严格数学逻辑的问题。

基于深度学习的方法大致分为有监督深度学习和无监督深度学习两种，利用大量数据进行训练，克服传统方法虽然有较高的深度估计精度，但由于存在缺少纹理或光照剧烈变化场景中的错误匹配和完整度不足等缺点。

有监督深度学习中最经典的是香港科技大学权龙教授团队提出的 MVSNet 算法，该方法以一个参考图像和多张原始图像为输入，首先提取图像的深度特征，然后通过可微分投影变换构建 3D 代价体，再通过代价累积和正则化输出 3D 的概率体，最后沿深度方向求取深度期望，对原始深度图进行优化。但基于深度学习中的有监督学习方法大都受限于数据集精度及规模，而对于土木工程施工中大型场景三维信息的基准值非常难以进行标准化测量，同时考虑到精确度、成本、泛化性等多方面因素，该方法的实际可行性较低。

相比之下，基于无监督学习的方法具有以下优点：

1. 无需预先采集构建一个大规模数据集，节省人力物力，且土木工程场景很难构建高精度的大规模数据集
2. 泛化性好，可以较为方便的从单中心物体三维重建扩展到建筑室内的三维重建
3. 可以有针对性的对土木工程施工场景进行约束和设计

综上，本课题选择基于无监督的深度学习方法进行实践，并辅以基于传统方法的 COLMAP 算法进行实验比较。

1.4 本课题基于无监督深度学习方法的基本情况

本课题采用无监督深度学习的方法进行土木工程场景的三维重建，主要需要进行源数据的采集和预处理、使用传统方法进行实验和比对、训练三维重建无监督深度学习模型、测试该模型的各项指标和重建结果等工作。

有关源数据的采集和预处理方面，本课题所想研究实现的三维重建方法是针对多视点问题的，而获取多视点数据最常用的方法是采用多台内参一样的相机同时拍摄待重建物体，但显然这种方法所需的成本比较大、运用起来也较为复杂，因此本课题中采用视频抽帧的方式进行多视点的模拟。传统视频抽帧图像序列仅通过计算机视觉算法进行相机标定和位姿估计，场景变化较大、光线和特征缺失等情况都会对使用造成影响。本课题希望融合运动惯性测量单元（Inertial Measurement Unit, IMU）和稀疏视觉特征进行相机内参标定和位姿估计，惯性补偿技术可以跟踪由物体运动惯性引起的视频序列在时域上的相关性，可使得重建效果更平滑，同时提取“关键帧”进行多视点三维重建。

基于传统三维重建方法的 COLMAP 算法使用上述源数据作为输入，通过稀疏重建、深度图估计和稠密重建三个核心步骤进行深度估计，并导出实验结果数据用于后续与无监督深度学习算法效果进行比较。

基于无监督深度学习的方法通过利用多个视图之间的光度一致性和像素梯度差异作为监督信号，进行深度图预测，通过在土木工程场景下特殊的约束和数据集的大规模训练不断降低表征代价，以训练出适用于土木工程场景或更具泛化性的模型。

最后使用无监督深度学习得到的模型，对融合了惯性数据的 RGB 视频序列视频进行处理，得到视频中各场景的深度信息并生成点云，再将点云进行重建生成重建后的模型，然后通过土木工程领域的软件或算法进行 BIM 模型的逆向建模。

综上，本课题将从融合 IMU 数据的 RGB 视出发，通过稀疏视觉特征和惯性测量单元进行相机内参标定和位姿估计，抽帧进行传统基于视觉的 COLMAP 和无监督深度学习多视点三维重建，最终将三维模型转换为 BIM 模型。

二、毕业设计（论文）方案介绍（主要内容）

1) 使用 RGB 相机和 imu 惯性测量单元融合采集视频，并进行抽帧和相机标定、位姿估计

考虑使用 intel RealSense L515 深度相机作为数据采集设备，关闭深度雷达数据流而只使用其 RGB+imu 数据流。参考 VINS-Mono，将关键帧提取分为如下四个主要步骤

- a. 提取 Harris 角点，使用 KLT 金字塔光流跟踪相邻帧
- b. 矫正二维特征点不失真，然后通过外点剔除后投影到单位球面上
- c. 进行 F 矩阵测试，通过 RANSAC 去除异常点

d. 满足以下两个条件时选取关键帧 ①当前帧相对最近关键帧的特征平均时差大于阈值(因为时差可以根据平移和旋转共同的到，而纯旋转则导致不能三角化成功，因此此步需要 IMU 预积分进行补偿) ②当前帧跟踪到的特征点数量小于阈值

2) 基于纯视觉的多视点三维重建 COLMAP 算法

主要包括：稀疏重建、深度图估计和稠密重建三个步骤

a. 稀疏重建：主要利用 SfM，通过增量方法可以得到场景中相机的姿态和表示场景结构的稀疏点云。主要通过特征匹配、几何纠正、三角测量恢复稀疏点云结构，并通过已有点云重新估计相对姿态，再进行全局的 BA 优化修正结构数据

b. 深度图估计：对于一个参考图像，筛选用于和参考图像配对计算视差的原始图，再通过代价匹配、代价聚合后，计算深度值，最后进行深度过滤，优化初始深度图

c. 稠密重建：利用深度图配准原理恢复场景稠密点云

3) 基于无监督深度学习的三维重建算法

相较于有监督 MVS 算法，不依赖于真实的 3D 数据，利用多个视图之间的光度一致性并加入了像素梯度差异作为监督信号，来预测深度图，核心包括两个问题：

a. 深度图预测：将同场景的 N 幅图像作为输入，通过 CNN 生成图像特征，利用可微单应变换，通过在深度值范围内对图像特征进行映射来构造视差体，然后通过 3D U-Net 来细化成本量，最后输出缩小分辨率的深度图

b. 光照一致性：利用投影光度一致性误差来产生自监督损失

4) 点云重建算法

使用 MeshLab、PCL 等常用工具进行点云重建，常用算法有柏松重建、Delaunay 三角剖分

其中 PCL 库中的柏松重建方法主要要经历如下步骤

- a. 使用 NormalEstimation 模块法向量估计
- b. 点云数据与法向数据拼接
- c. 使用 GreedyProjectionTriangulation 模块进行曲面重建

5) BIM 模型生成

通过导入点云数据模型 (rcp) 至 revit (rvt) 软件中分析项目的特征且查阅点云数据的完整性，制作可视化 BIM 模型

三、毕业设计（论文）的主要参考文献

-
- [1]Khot T, Agrawal S, Tulsiani S, et al. Learning unsupervised multi-view stereopsis via robust photometric consistency[J]. arXiv preprint arXiv:1905.02706, 2019.
- [2]Schonberger J L, Frahm J M. Structure-from-motion revisited[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4104-4113.
- [3]Schönberger J L, Zheng E, Frahm J M, et al. Pixelwise view selection for unstructured multi-view stereo[C]//European Conference on Computer Vision. Springer, Cham, 2016: 501-518.
- [4]Yao Y, Luo Z, Li S, et al. Mvsnet: Depth inference for unstructured multi-view stereo[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 767-783.
- [1] Ullman S. The interpretation of structure from motion[J]. Proceedings of the Royal Society of London. Series B. Biological Sciences, 1979,203(1153):405-426.DOI:10.1098/rspb.1979.0006.
- [2] Pizzoli M, Forster C, Scaramuzza D B I. REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time: IEEE International Conference on Robotics and Automation (ICRA), 2014[C]. IEEE.
- [3] Forster C, Pizzoli M, Scaramuzza D B I. SVO: Fast Semi-Direct Monocular Visual Odometry: IEEE International Conference on Robotics and Automation (ICRA), 2014[C]. IEEE.
- [4] Hirschmuller, H., Accurate and efficient stereo processing by semi-global matching and mutual information, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2005[C], IEEE.
- [5] Hirschmüller, H., Stereo Processing by Semi-Global Matching and Mutual Information. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007. 30(2): p. 328-341
- [6]Qin T, Shen S. Online temporal calibration for monocular visual-inertial systems[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 3662-3669.
- [7]Qin T, Li P, Shen S. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.
- [8] Zhengxin Mi, Yiming Luo, Wenbing Tao. SSRNet: Scalable 3D Surface Reconstruction Network. CVPR 2020.
- [9] Yiming Luo, Zhengxin Mi, Wenbing Tao. DeepDT: Learning Geometry From Delaunay Triangulation for Surface Reconstruction. AAAI 2021.
- [10]朱煜捷,张得煊,申玉忠,孙菁远.基于三维激光扫描的 BIM 逆向建模技术研究热点和趋势分析 [J/OL]. 土木建筑工程信息技术:1-7[2021-03-05].<http://kns.cnki.net/kcms/detail/11.5823.TU.20210224.1126.008.html>.
- [11]曹娟.三维激光扫描技术在 BIM 中的应用[J].信息技术与信息化,2020(12):87-89.
- [12]刘胜男,陶钧,史艾嘉.基于三维激光点云的既有建筑 BIM 建模方法研究与应用[J].安徽建筑,2020,27(12):142-143+147.
- [13]刘莎莎. 点云数据与 BIM 集成的建筑物施工进度监测技术方法[D].西南交通大学,2019.
- [14]郑华海,刘匀,李元齐.BIM 技术研究与应用现状[J].结构工程师,2015,31(04):233-241.
- [15]Asadi K, Han K. Real-time image-to-BIM registration using perspective alignment for automated construction monitoring[C]//Construction Research Congress. 2018, 2018: 388-397.
- [16]Banfi F, Stanga C, Brumana R. A Digital Workflow for Built Heritage: From SCAN-to-BIM

Process to the VR-Tour of the Basilica of Sant'Ambrogio in Milan[C]//Euro-Mediterranean Conference. Springer, Cham, 2018: 334-343.

[17]THOMSON C, BOEHM J. Automatic Geometry Generation from Point Clouds for BIM [J]. Remote Sensing, 2015, 7(9): 11753-11775.

四、审核意见

指导教师审核意见：（针对选题的价值及可行性作出具体评价）

字数不宜太少

指导教师签名_____

_____年_____月_____日

专业审核意见：

适当作出有针对性的评价，字数不限。

负责人签名_____

_____年_____月_____日