

基于巡检视频的 COLMAP 图像三维重建效率与精度评估

林佳瑞^{1,2}, 周绍杰^{1,2}, 王宇平³, 曾申翔¹, 衣子俊¹, 潘鹏^{1,2,*}

(1. 清华大学土木工程系, 北京 100084; 2. 住房和城乡建设部数字建造与孪生重点实验室, 北京 100084; 3. 国网北京市电力公司, 北京 100031)

摘要: 室内巡检是贯穿建筑全生命周期的关键环节, 而图像三维重建为其数字化提供了低成本、高灵活性的解决方案。多视图几何重建算法因其高效稳定的优势被广泛应用, 但在室内环境下的重建效率和精度仍不明确。因此, 本文采用开源工具 COLMAP, 选取 7 个典型场景并基于巡检视频开展三维重建实验, 以揭示图像数量、重叠程度等因素对重建效率和精度的影响规律。结果表明: 1) 重建时间与图像数量的 1.15 次幂呈正比; 2) 重建效率随图像重叠程度先增后减, 存在最佳的重叠程度; 3) 消费级相机的重建几何精度为厘米级, 同一场景误差受重建密度影响小, 不同场景的平均绝对(相对)误差小于 3 cm (4%)。研究揭示了室内场景图像三维重建效率与精度的一般规律, 为更高效的室内巡检影像采集与应用指明了方向。

关键词: 室内巡检; 巡检机器人; 三维重建; COLMAP; 重建精度; 重建效率

中图分类号: TU18

Evaluation of Efficiency and Accuracy for COLMAP-based 3D Reconstruction through Inspection Videos

Lin Jiarui^{1,2}, Zhou Shaojie^{1,2,#}, Wang Yuping^{3,#}, Zeng Shenxiang¹, Yi Zijun¹, Pan Peng^{1,2,*}

(1. Department of Civil Engineering, Tsinghua University, Beijing 100084, China;

2. Key Laboratory of Digital Construction and Twinning, Ministry of Housing and Urban-Rural Development, Beijing 100084, China;

3. State Grid Beijing Electric Power Company, Beijing 100031, China)

Abstract: Indoor inspection is a crucial process throughout the entire lifecycle of a building, while image-based 3D reconstruction provides a cost-effective and highly flexible solution for its digitization. Multi-view geometric reconstruction algorithm is widely adopted due to its productivity and stability, yet its reconstruction efficiency and accuracy in indoor environments remain unclear. Therefore, this study employs the open-source tool COLMAP, conducting 3D reconstruction experiments on seven typical scenarios using inspection videos to reveal the impact of factors such as image quantity and overlap ratio on reconstruction efficiency and accuracy. Results indicate that: 1) Relationship between reconstruction time and the number of images follows a power law with an exponent of 1.15; 2) Reconstruction efficiency first increases and then decreases as the image overlap ratio grows, suggesting an optimal overlap ratio; 3) Geometric accuracy of reconstruction using consumer-grade cameras reaches centimeter-level precision, with

收稿日期: XXXX-XX-XX; **网络出版日期:** XXXX-XX-XX.

作者简介: 林佳瑞 (1987—), 男, 博士, 副研究员, lin611@tsinghua.edu.cn; 潘鹏 (通信作者), 男, 博士, 教授, 博士生导师, panpeng@tsinghua.edu.cn.

基金项目: 国家电网公司科学技术项目 (5200-202311481A-3-2-ZN).

scene-specific errors minimally affected by reconstruction density, and the mean absolute (relative) error across different scenes is below 3 cm (4%). The study reveals the general rules of efficiency and accuracy for 3D reconstruction of indoor scenes, offering guidance for optimizing image acquisition and application in indoor inspection workflows.

Key words: indoor inspection; inspection robotics; 3D reconstruction; COLMAP; reconstruction accuracy; reconstruction efficiency

在建筑工程全生命周期管理中，室内巡检作为质量安全控制的关键环节，其数字化升级对预防结构隐患、提升运维效率具有重要价值。研究报告^[1]表明，中国智能巡检机器人行业呈现出蓬勃发展的趋势，市场规模快速增长，行业前景广阔。近年来，室内巡检机器人应用也出现了显著的增长，并且提升了巡检的效率、安全性和便利性^[2]，是建筑业转型的重要举措^[3]。三维环境感知是计算机理解物理空间的基础，为建筑语义理解^[4]和进度跟踪等巡检任务^[5]提供技术支撑。其中，三维重建的效率和精度对路径导航^[6]、远程遥控^[7]、物体识别^[8]等机器人巡检任务具有重要意义。常见的三维重建手段包括依赖结构光、激光雷达等设备的主动式重建和依赖多视角立体视觉的被动式重建^[9]。尽管前者往往具有较高的精度，但其硬件成本高、部署灵活性差，难以满足轻量化巡检需求^[10]。相比之下，图像三维重建技术只需要消费级相机即可实现场景的重建^[11]，在成本与灵活性上具有显著优势，为基于图像的室内巡检带来了新的契机。

图像三维重建旨在通过二维图像恢复场景的三维几何结构。传统几何重建的方法基于运动结构恢复（SfM）和多视图立体几何（MVS）进行相机位姿的解算和稠密重建，依赖多视图几何约束，算法可解释性高，具有几何一致性、高效性和鲁棒性，在植物学^[12]、遥感测绘^[13]、市政工程^[14]等领域都有广泛的验证。而COLMAP^[15]作为常用的开源多视图重建工具，集成了上述算法，支持从图像到完整三维模型“一键式”生成，在文化遗产数字化^[16]、无人机测绘^[17]等领域都有所应用。但在建筑领域，传统重建算法更多关注室内场景纹理丢失等真实感问题^[18]，但鲜有三维重建效率和精度的相关研究，致使难以评估相关技术在室内巡检场景中的适用性及有效性。

因此，本文以 COLMAP 开源框架为重建工具，深入研究和评估了室内场景中图像三维重建算法的效率与精度。研究选取了学校、住宅建筑中常见的室内场景，基于智能手机（消费级相机）按指定方式采集巡检视频数据，抽帧采样以构建不同参数的图像数据集。在此基础上提出有效图像比例、重建时间函数、重建速率、图像利用率、重建误差等指标，从时间、效率和精度等角度揭示了图像三维重建的一般规律，并进一步探讨了基于 COLMAP 图像三维重建在室内巡检任务中的应用前景。

批注 [Author1]: 引言中背景介绍较多，三维重建内容较少。建议作者增加基于 COLMAP 的相关工作研究介绍，以更好的说明研究基于 COLMAP 三维重建的重要性。
1. 强调三维重建在室内巡检的重要性；2. 与传统方法显著差异。

1 传统图像三维重建方法与影响因素

1.1 COLMAP 三维重建原理

如图 1 所示，作为一款功能强大的图像三维重建开源工具，COLMAP 集成了从 SfM 到 MVS 的完整重建流程，包括特征提取与匹配、稀疏重建和稠密重建等关键步骤。

1、特征提取与匹配：COLMAP 首先对输入的 n 张图像进行特征提取，常用的算法如 SIFT (Scale-Invariant Feature Transform) 等，能够检测出图像中存在颜色变化的轮廓和角点像素作为关键点，也称为特征点，并生成对应的特征描述符。随后，通过比较特征描述符间的相似程度可以匹配不同图像中的相同特征点。

2、稀疏重建：COLMAP 将从成功匹配的图像开始，根据不同图像之间相同特征点的约束条件构建方程组，基于三角测量 (Triangulation) 等原理计算出特征点的深度信息和对应的相机参数，所有特征点在三维空间的映射构成了稀疏的点云模型，即实现了稀疏重建。这一步骤通常从一个初始的图像对开始，通过逐步添加更多的图像进行增量式重建，并利用优化算法 (如 Bundle Adjustment) 来最小化重投影误差。为了进一步提高重建的精度，COLMAP 还会进行全局优化，包括全局约束调整 (Global Bundle Adjustment) 和全局位姿图优化 (Global Pose Graph Optimization)，以确保所有图像的相机参数和三维点云具有一致性。需要注意的是，只有特征点被成功匹配和相机参数成功解算的 n_0 张有效图像才能进入稀疏重建环节。

3、稠密重建：仅仅只有特征点深度信息的点云是稀疏的，COLMAP 将通过 MVS 算法逐像素估计深度并融合生成稠密点云。这一步骤通常基于光度一致性 (Photometric) 与几何一致性 (Geometric) 联合优化图像的深度和法向信息，以生成高分辨率的深度图。通俗来说，光度一致性要求同一空间点在不同角度的图像上有着相似的局部像素分布；几何一致性则要求不同图像上的对应点应满足极线约束 (Epipolar Constraint)。这些信息为图像上非特征点的其余像素提供了深度约束，能够从中还原与求解更加丰富的三维信息。最终，不同视角图像的深度信息被投影融合到统一的全局坐标系下，就获得了稠密的点云模型。

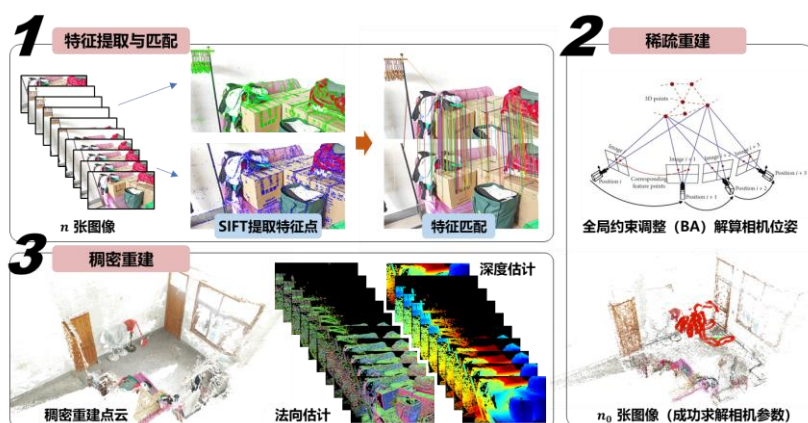


图 1 基于 COLMAP 的三维重建流程

Fig. 1 COLMAP-based 3D reconstruction process

1.2 图像三维重建影响因素分析

图像作为 COLMAP 重建的原始数据，拍摄质量与图像数量将直接影响重建效果。COLMAP 重建原理表明，特征提取与匹配作为重建的第一步，只有成功匹配的 n_0 张图像才对重建结果有贡献，并显著影响重建的耗时、效率和重建点云的稠密程度。因此，这意味着图像数据的采集需要满足两个要求：1) 确保图像具有易于辨识的特征点，即清晰的成像效果和丰富的纹理信息；2) 图像之间有一定重叠，保证不同视角的特征点能够成功匹配。这两条基本原则给图像的拍摄提出了要求：一方面需要尽量避免动态模糊带来的干扰，如手持拍摄抖动或快速移动可能导致图像模糊和特征点提取失败；另一方面需要保持拍摄时具有一致的光照条件，成像设备采用相同的曝光、ISO 等参数，避免亮度突变导致误匹配率上升。

综上所述，拍摄方式、总图像数量 n 、有效图像数量 n_0 和图像重叠程度等参数都会影响重建的时间和效果。因此，本研究将围绕这些影响因素进行图像数据集的采集和重建效率、精度评估方法的构建。

2 基于巡检视频的图像三维重建与评估方法

2.1 巡检视频的采集与重建

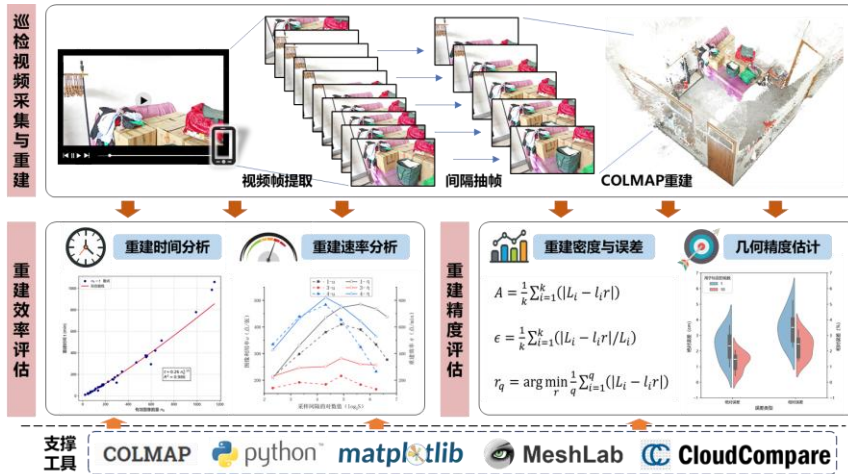


图 2 基于巡检视频的图像三维重建与评估流程

Fig. 2 Evaluation process for image-based 3D reconstruction from inspection videos

图 2 展示了本研究完整的技术路线。为了评估和量化室内场景中 COLMAP 三维重建的效率和精度，首先需要对典型室内场景进行图像数据的采集。为了保证图像之间有所重叠，并覆盖完整的室内场景，本研究采用连续拍摄的巡检视频进行均匀间隔的抽帧以模拟完整的图像采集过程。通过控制不同的抽帧间隔 S （即两张相邻图像之间相隔的帧数），能够获得

不同重叠程度和数量的连续图像数据集。这些数据集在相同的计算环境下使用 COLMAP 进行重建，以探究图像数量和重叠程度等因素对重建效率和精度的影响。

为了确保重建的成功率和不同室内场景重建效果的可比性，遵循图像三维重建基本原则，本研究规定了统一的巡检视频拍摄方式和相机参数。如图 3 所示，从室内场景的一角开始，距离墙面或障碍物 2-3m，使用成像设备进行同一方向往复的“S”形拍摄，并遍历室内场景的六个面。同时，需要控制拍摄的平稳度和速度。不稳定或者过快会产生模糊的图像帧；过慢的拍摄速度会增加视频数据的信息冗余。在本研究中，相机移动速度约为 0.5 m/s，旋转速度约为 16 °/s，一个不超过 20 m² 的室内场景大约需要拍摄 3 min 以内的巡检视频。同时在数据采集的过程中，保证视频的分辨率、帧率，相机的 ISO、曝光，环境的光照条件、布置等保持不变。经过测试，这样的拍摄方式能够较大幅度保证重建的成功率，获得较完整的场景。

批注 [Author2]: 请增加说明，如何控制拍摄平稳和稳定的移动速度。

批注 [Author3]: 室内黑暗区域是否引入一些图像处理操作。

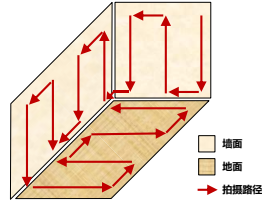


图 3 巡检视频采集方式

Fig. 3 Collection method of inspection videos

2.2 重建效率与精度的评估方法

效率与精度是评估图像三维重建性能的关键。在实际应用中，点云的重建时间、稠密性以及几何信息的准确性都是研究的重点。因此，本节围绕这三项目标，构建图像三维重建效率与精度的综合评估方法。

批注 [Author4]: 这三个评价指标是作者自行定义的还是相关行业规定，或是来自于相关研究的经验？

2.2.1 有效图像比例

成功重建是讨论效率和精度的前提。COLMAP 的重建原理表明，当抽帧间隔 S 增大时，总图像数量显著减少，相邻图像间的重叠程度降低，导致图像间特征匹配效果变差，即有效匹配的图像数量减少，影响重建场景的完整性，甚至无法重建。因此，有效图像数量 n_0 与总图像数量 n 的比值是量化是否成功重建的关键参数，定义为有效图像比例 E ，即

$$E = \frac{n_0}{n} \quad (1)$$

该值从侧面上反映了图像间的重叠程度和匹配效果，对于相同规模的图像数据集， E 越大，则数据集质量越高，重建完整性越好；反之当 E 较小时，重建不完整，视为重建失败。

2.2.2 重建时间函数

COLMAP 的重建时间与图像数量密切相关。直观来看，当图像数量越多，重建时间越长。但这一数量规律并非简单的可推导关系，重建过程涉及不同参数（如每张图像的特征点

数、特征匹配情况等)的影响,计算复杂且难以估计。忽略次要因素,影响重建时间 t 最主要的变量是总图像数量和有效图像数量,因此在本研究中通过拟合 t 与 n 和 n_0 的函数关系 T 来估计重建时间,即

$$t = T(n, n_0) \quad (2)$$

2.2.3 重建速率与图像利用率

为了在有限的时间内获得尽可能稠密的点云,定义重建速率 v 为单位时间内生成点云的数量,作为重建效率的重要指标之一,即

$$v = N/t \quad (3)$$

式中 N 为重建点云中点的数量。 v 越大,表明重建的时间利用率越高,重建的效益也越高。

类似地,为了使尽可能少的图像获得尽可能稠密的点云,定义图像利用率 u 为平均每张图像生成的点云数量,即

$$u = N/n_0 \quad (4)$$

式中分母为有效图像数量 n_0 ,因为只有有效匹配的图像才真正被用于稠密重建。 u 越大,表明图像的价值越高,数据集的质量也越高,信息冗余越少。

2.2.4 重建几何精度

重建几何精度反映了重建点云与真实场景的尺寸误差。值得注意的是, COLMAP 重建过程仅仅利用了图像的 RGB 信息,缺乏尺度信息,这使得重建点云与真实尺寸之间存在一个比例差异。假设在某重建点云中,测量了 k 个具有代表意义的长度值,分别为 l_1, l_2, \dots, l_k , 对应真实场景中的长度值分别为 L_1, L_2, \dots, L_k 。此时,若真实尺寸与重建点云的缩放比例为 r ,则重建点云的平均绝对误差 A 和平均相对误差 ϵ 可以通过式(5)和(6)计算:

$$A = \frac{1}{k} \sum_{i=1}^k (|L_i - l_i r|) \quad (5)$$

$$\epsilon = \frac{1}{k} \sum_{i=1}^k (|L_i - l_i r|/L_i) \quad (6)$$

但是,真实的缩放比例 r 无从得知,需要通过已知尺寸的物体对重建点云进行标定。

从测定的 k 组长度对应关系中任选 q 组 ($1 \leq q \leq k$, 角标记为 $1 \sim q$), 则可根据式(7)计算最优的缩放比例 r_q , 代入式(5)和(6)可计算相应的误差指标,记为 A_q 和 ϵ_q 。

$$r_q = \arg \min_r \frac{1}{q} \sum_{i=1}^q (|L_i - l_i r|) \quad (7)$$

3 实验设计与结果分析

3.1 实验设计与参数配置

为了获得巡检视频重建效率和精度的普遍规律,本研究对 7 个不同的室内环境进行了重建实验。7 个场景的巡检视频由 4 名研究人员使用消费级相机(智能手机)采集完成,包括客厅、卧室、宿舍、办公室、仓库、走廊等常见的室内场景,拍摄按照 2.1 规定的方式进行,并保证光照条件不变,统一使用 1920*1080 (1080P) 的分辨率进行拍摄,每秒 30 帧,最终视频时长根据场景大小从 100s 到 180s 不等。随后,通过每 5、10、20、30、50、75、

批注 [Author5]: 重建时间 t 具体是通过计算 COLMAP 中重建记录计算的还是根据 n 和 n_0 之间的函数关系计算? 如果是函数关系请添加详细计算方法。

批注 [Author6]: 缩放比例的计算确实需要通过前期采样得到。请说明,缩放比例是 COLMAP 的固定内参(一次采样计算全部通用)还是会根据不同的场景需要各自的特定计算。此外,明确最优缩放比例是否为全局统一缩放比例

100 帧的间隔对上述巡检视频进行抽帧以构建原始的图像数据集。本研究中用于三维重建的实验环境如下：CPU 为 2 颗 Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz，显卡为 1 张 NVIDIA GeForce RTX 3090，操作系统为 Ubuntu 18.04.6 LTS，COLMAP 为 3.8 cuda 版。此外，重建参数按照软件默认配置，重建过程的所有输出结果保存在本地的日志文件下。

完成重建后，对每个重建案例提取图像数量 n 、有效图像数量 n_0 、重建时间 t 和重建点云中点的数量 N 等关键参数，用于重建效率的计算。对于重建成功的案例，还需要对真实场景与重建点云分别进行测距，保证每个场景有 10 对“真实尺寸-点云距离”的长度对应关系，用于重建几何精度的计算。

3.2 重建时间与效率分析

3.2.1 有效图像比例与重建效果

实验结果表明，有效图像比例能够反映图像三维重建的成功情况。如图 4 所示，以场景 <6> 为例，抽帧间隔为 5、10 和 20 帧能够较好实现场景的重建，间隔 30、50 和 75 帧会导致重建不完整，而间隔 100 帧则直接导致重建失败。经计算，这 7 个数据集对应的有效图像比例 E 分别为 99%、95%、83%、31%、31%、28% 和 0%，与重建场景的完整性规律一致。在本研究中，规定有效图像比例不低于 80% 为成功重建，其余重建不完整的实验不参与重建效率和精度的分析。在 49 个重建案例中，有 14 个案例的有效图像比例低于 80%，不予以考虑。并且这 14 个案例中有 9 个的抽帧间隔为 75 或 100 帧，这说明较大的抽帧间隔会导致图像间的重叠程度较小，严重影响重建完整度和成功率。

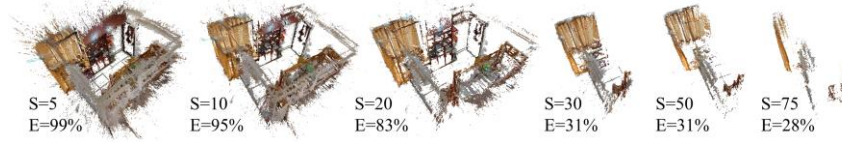


图 4 重建效果与有效图像比例的关系

Fig. 4 Relationship between reconstruction quality and effective image proportion (E)

3.2.2 重建时间与图像数量

由于 $n_0 = E \cdot n$ ，并且在上一节中进行了 $E \geq 80\%$ 的筛选，因此 n 与 n_0 之间存在高度相关性。图 5 绘制了 $n - n_0$ 的散点分布和线性拟合结果，其中 n 与 n_0 的皮尔逊相关系数 $\rho_{n,n_0} = 0.998$ ，显著性水平 $P = 9.14 \times 10^{-42} < 0.05$ ，可以推断出 n 与 n_0 高度线性相关，在拟合重建时间函数时不需要考虑二者的耦合关系。

考虑到有效图像数量 n_0 对重建结果的影响更直接，对 $n_0 - t$ 进行拟合。如图 6 所示，重建时间随图像数量单调增加，且增长趋势逐渐加快，基于幂律函数拟合的结果为：

$$t = 0.17n_0^{1.15} \quad (9)$$

其中，拟合优度 R^2 为 0.986。

从图 6 可以看出，指数 1.15 是对重建时间的保守估计，当 n_0 较大时， t 的拟合结果

批注 [Author7]: 请说明运行内存大小，这是至关重要的。

批注 [Author8]: 建议修改为抽帧间隔 S ，再一次的说明可以更好的与图 4 对应。

批注 [Author9]: 可以简单说明为什么使用皮尔逊相关系数。

明显偏小。这是由于大部分 $n_0 - t$ 的对应关系都集中在 n_0 较小的区域内，相应区间的权重更大，从而影响了 n_0 较大时的拟合结果。可以预见的是，当图像数量继续增加，重建时间的增长会进一步加速，这也意味图像三维重建并不适合直接利用大规模图像数据集进行三维重建。

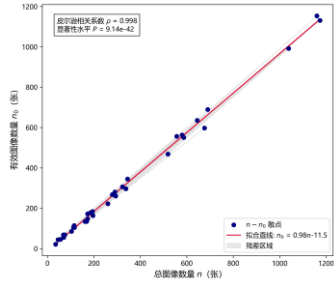


图 5 $n - n_0$ 相关性分析

Fig. 5 Correlation analysis of $n - n_0$

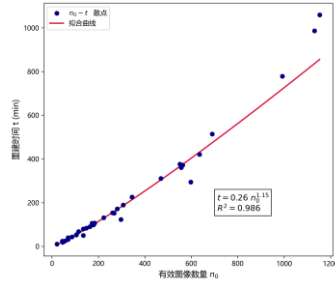


图 6 $n_0 - t$ 拟合函数

Fig. 6 Fitting function of $n_0 - t$

3.2.3 重建速率、图像利用率与图像重叠程度

选择在所有抽帧间隔上重建效果都较好的三个场景（场景<1>、<3>和<4>）进行重建速率和图像利用率的分析。以抽帧间隔的对数值 $\log_2 S$ 为横坐标，图像利用率 u 和重建速率 v 为纵坐标，图 7 展示了这两项指标随抽帧间隔的变化趋势。结果表明，无论是图像利用率还是重建速率，都随抽帧间隔的增大呈现出先增大后减小的变化趋势，其峰值大多出现在 20-30 帧范围内，并显著高于其他点。这说明了重建效率受图像重叠程度的影响大：当图像数量少时，图像重叠程度低，特征匹配数量少，导致重建速率和图像利用率较低，甚至可能导致重建不成功。反之，当图像数量多时，图像重叠程度大，特征匹配数量多，重建成功率高，但造成了信息冗余，增加计算负担，显著降低了重建速率和图像利用率。

此外，根据式（2）（3）（4）和 3.2.2 的实验结果，可得 u 、 v 与 N 、 T 应满足式（7）：

$$\frac{N}{v} = T\left(\frac{N}{u}\right) \quad (7)$$

不难发现，由于重建时间函数 T 是单调递增的，重建速率与图像利用率在本质上是统一的，仅仅是从不同层面反映了重建的效率。在本研究的实验条件下（相机移动速度约为 0.5 m/s，旋转速度约为 16 °/s，视频采集为 30 帧/s），20-30 帧的抽帧间隔下对应的图像重叠程度能够使 u 和 v 最大，在保证重建效果的前提下重建效率最高，最有利于图像三维重建在工程实际中的应用。

批注 [Author10]: 一般性结论，建议凝练。

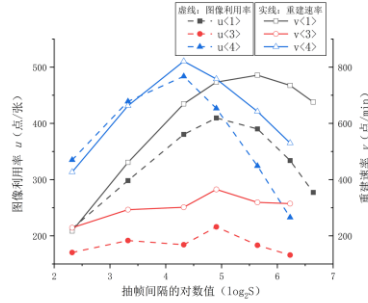


图 7 重建速率与图像利用率的变化曲线

Fig. 7 The variation curves of reconstruction velocity and image utilization efficiency

3.2.4 重建时间、效率与图像重叠程度的制约关系

在此基础上，计算所有场景的重建时间和重建速率，并进行无量纲化处理，以探究重建时间、效率（速率）与图像重叠程度间更加普遍的变化规律。规定抽帧间隔为 100、75、50、30、20、10、5 帧为 7 种等级的抽帧频率，即 7 种不同图像重叠程度，分别记为等级 1~7，抽帧间隔越小，图像数量越多，重叠程度越高，对应的等级数字就越大。将不同场景无量纲化后的指标进行平均，绘制重建效率和时间的变化曲线如图 8 所示。可以将图 8 分为三个区域 A、B 和 C：A 表示图像数据集的重叠程度低，重建时间短，重建效率较高，但重建点云较为稀疏，重建成功率不稳定，适合快速重建任务；B 是最理想的参数分区，图像重叠程度适中，重建时间较短，重建效率最高，并且能够保证良好的重建成功率，适用于大多数重建任务；C 则由于图像重叠程度过高，导致信息冗余浪费，重建时间急剧增加，效率急剧降低，是最不推荐的参数选择，但适合需要稠密点云的重建任务。表 1 总结了 A、B、C 的参数选择与各项指标的对比结果。

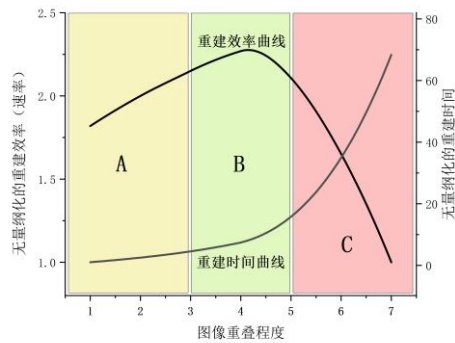


图 8 重建效率、时间与图像重叠程度的关系

Fig. 8 Relationship between reconstruction efficiency, time, and image overlap ratio

表 1 分区 A、B 和 C 的参数选择与指标对比

Table 1 Comparison of parameters and indicators in region A, B, and C

区域	图像重 叠程度	重建 成功率	重建 时间	点云 密度	重建 效率	适用范围
A	低	不稳定	短	稀疏	中	适合需要快速重建的任务。
B	中	高	较短	正常	高	最佳参数选择, 适用于大多数任务。
C	高	高	长	稠密	低	不推荐, 适合需要稠密点云的任务。

3.3 重建的几何精度分析

重建几何精度的分析将围绕以下两个方面展开: 1) 不同抽帧间隔(重建点云密度)对重建几何精度的影响; 2) 不同室内场景下重建几何精度的一般规律。

值得注意的是, 针对每个实验案例, 本研究测定了 10 对“真实尺寸-点云距离”对应关系用于重建几何精度的分析, 最理想的指标是 10 组长度关系都用于缩放比例标定的误差 A_{10} 和 ϵ_{10} 。但不排除特殊情况下, 先验的长度信息有限, 如只有 1 组长度可以用于缩放比例标定, 这将导致重建误差的放大。因此规定基于 10 组长度对应关系逐一标定的计算指标 (A_1 和 ϵ_1) 的平均值 \bar{A}_1 和 $\bar{\epsilon}_1$ 为最恶劣条件下重建结果的几何精度指标。

3.3.1 同一场景不同点云重建密度的几何精度

以场景<4>为例, 表 2 展示了在不同抽帧间隔(重建密度)下 10 组长度进行标定和 1 组长度进行标定的重建误差, 其中抽帧间隔 100 帧重建不成功, 不予以考虑。使用 10 组长度进行标定时, 场景<4>的绝对误差平均值为 1.03 cm, 相对误差平均值为 1.68%; 使用 1 组长度进行标定的对应误差分别为 1.62 cm 和 2.67%。图 9 对重建误差分布进行了可视化, 可以看出: 1) 使用 10 组长度进行标定的重建误差显著小于仅用 1 组长度标注的重建误差; 2) 对于同一场景而言, 只要重建成功, 不同抽帧间隔对于重建误差的影响不明显。除了抽帧间隔 75 帧时点云过于稀疏引起误差较大波动以外, 其余误差都集中分布在平均值附近。

表 2 同一场景不同点云密度下的重建误差

Table 2 Reconstruction error under different point cloud densities of the same scene

抽帧间隔 (帧)	10 组长度标定			随机 1 组长度标定的平均		
	r_{10}	$A_{10}(\text{cm})$	$\epsilon_{10}(\%)$	\bar{r}_1	$\bar{A}_1(\text{cm})$	$\bar{\epsilon}_1(\%)$
5	14.02	1.07	1.53	13.98	1.66	2.47
10	12.84	0.79	1.26	12.84	1.53	2.14
20	12.79	0.43	1.37	12.71	1.24	2.26
30	12.42	0.98	1.70	12.39	1.37	2.48
50	14.20	0.94	1.72	14.22	1.42	2.62
75	13.36	1.97	2.49	13.42	2.47	4.02
100	-	-	-	-	-	-
平均		1.03	1.68		1.62	2.67

批注 [Author11]: 既然是精度分析, 建议除了量化数值对比分析外, 添加对应的重建可视化结果分析 (可以在可视化图片上标识出差距)。

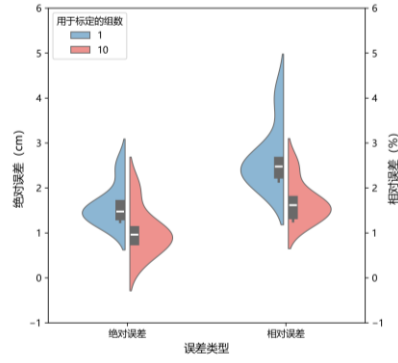


图 9 同一场景不同点云密度下的重建误差分布

Fig. 9 Error distribution of 3D reconstruction under different point cloud densities of the same scene

上述结果表明：图像三维重建的几何精度对重建密度不敏感。通过不同重建密度的点云对比，能够发现图像三维重建对于纹理变化的区域具有较强的识别能力，这使得即便点云稠密程度发生变化，物体的轮廓信息仍然能够得到有效重建，而不会对重建的几何精度造成显著影响，这一表现与 COLMAP 中基于 SIFT 等算法进行图像特征点提取的原理高度吻合。

3.3.2 不同场景的几何精度

针对不同场景，选择最优重建效率对应的重建点云作为几何精度计算的研究对象，共 7 个场景。表 4 分别展示了 7 个场景的 10 组长度标定和 1 组长度标定的误差结果。从表中可以看出，室内场景图像三维重建的几何精度为 cm 级。使用 10 组长度进行标定时，绝对误差平均值为 1.29 cm，相对误差平均值为 2.27 %；使用 1 组长度进行标定的对应指标分别为 2.29 cm 和 3.49 %。图 10 对所有场景的重建误差分布进行了可视化，虽然 7 个实验案例的原始巡检视频来自于不同场景、不同拍摄者和不同成像设备，但重建误差都集中分布在平均值附近，这证明厘米级的精度结果对不同室内场景具有鲁棒性和普遍性。

表 4 不同场景的重建误差范围

Table 4 Error range of 3D reconstruction in different scenarios

场景序号	10 组长度标定			随机 1 组长度标定的平均		
	r_{10}	$A_{10}(\text{cm})$	$\epsilon_{10}(\%)$	\bar{r}_1	$\bar{A}_1(\text{cm})$	$\bar{\epsilon}_1(\%)$
1	14.01	0.58	1.42	13.99	0.98	2.26
2	22.82	1.62	2.57	23.22	3.64	3.51
3	8.63	1.23	1.85	8.77	2.00	3.09
4	12.79	0.43	1.37	12.71	1.24	2.26
5	9.51	1.47	2.44	9.64	2.81	3.85
6	9.93	1.68	2.90	3.92	2.33	4.23
7	13.79	2.04	3.36	13.97	3.07	5.22
平均		1.29	2.27		2.29	3.49

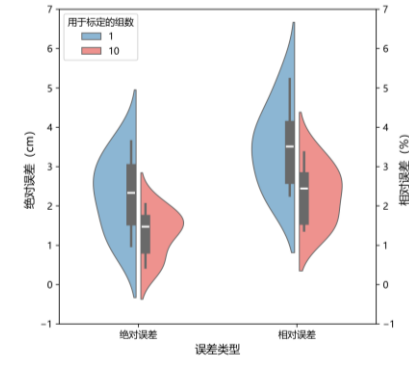


图 10 不同场景的重建误差分布

Fig. 10 Error distribution of 3D reconstruction in different scenarios

此外，尺寸缩放比例 r_{10} 和 r_1 随场景波动较大，并没有呈现出明显的规律性。这一比例差异源于 COLAMP 重建的固有问题，无法直接估计。图 10 也表明，用 10 组真实尺寸标定可以获得更优秀和更稳定的重建误差，这启示在一定程度内提供更多先验的尺寸信息能够减少图像三维重建中缩放比例带来的几何精度影响。

3.3.3 图像三维重建的常见缺陷

虽然常见物体的图像三维重建的绝对误差平均值在 3 cm 以内，但从重建点云的可视化结果来看，室内场景中弱纹理、高重复结构和透光材料带来的重建缺陷仍然是不可避免的^[19]。纯白墙面、天花板和地面，重复的瓷砖排布等都会影响局部区域的重建效果。图 11 展示了本实验中常见的重建缺陷：(a) 场景的墙面重建不完整，仅有插座和裂纹这些纹理特征附近有重建的痕迹，并且窗框处受到玻璃窗的影响产生了扭曲；(b) 场景中的白色瓷砖由于缺乏纹理和重复结构导致重建不完整，只有接缝处被重建；(c) 场景的墙面贴有纹理丰富的墙纸，因此得以完整重建，但白色天花板由于缺乏纹理而重建不完整，仅有灯饰被重建；(d) 场景由于窗户透光导致外景被重建产生了大量无意义的噪点。这些缺陷都是目前基于图像的室内场景三维重建中常见的问题与技术难点，在应用和选择标定尺寸时需要尽量避开。

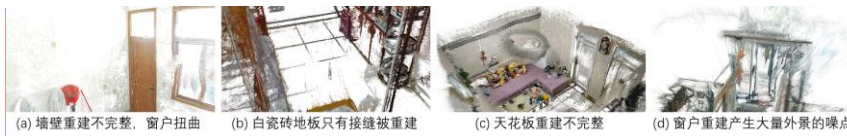


图 11 室内场景中常见的重建缺陷

Fig. 11 Common defects in 3D reconstruction of indoor scenes

4 结论与展望

图像三维重建为计算机的三维感知提供了一种低成本、高灵活性的解决方案，是数字化室内巡检的重要技术之一。COLMAP 开源框架采用了基于 SfM 到 MVS 的传统图像三维重

批注 [Author12]: 应该单独成一个小节，并简单分析原因。

批注 [Author13]: 建议放大图像，2×2 排布会更好。

建算法，因其高效性和稳定性而得到广泛应用。本文针对 COLMAP 巡检视频三维重建在室内场景中的效率和精度进行研究与评估，通过对视频进行不同间隔的抽帧构建不同重叠程度的图像数据集以开展对比实验，并基于所构建的评价体系进行重建效率与精度的量化。结果表明：

- 1) 在本研究的重建环境下，重建时间与成功匹配的有效图像数量的 1.15 次幂呈正比；
- 2) 重建效率随图像重叠程度的增加呈现先提高后降低的趋势，在本研究 0.5 m/s 相机移动速度、16 °/s 相机旋转速度和 30 帧/s 视频采集帧率下最理想图像重叠程度对应的抽帧间隔为 20-30 帧，能够在保证较短重建时间的同时获得最高的重建效率；
- 3) 室内场景图像三维重建的几何精度为 cm 级，同一场景的重建误差受点云密度影响较小，不同室内场景的绝对误差平均值不超过 3 cm，相对误差平均值不超过 4 %。

上述结果揭示了室内场景下图像三维重建时间、效率和精度的一般规律，为图像三维重建在室内场景中的采样策略和巡检应用提供了指导：

- 1) 从时间和效率角度来看，基于 COLMAP 的三维重建时间随图像数量增加而加速增长，因此并不适用于大场景多图像的重建任务。以本研究采用的常见智能手机和“S”形拍摄方式为例，图像数量在 300 张以内较为合适，重建时间较短重建效率最高，这能够应付不超过 30 m² 室内场景。对于室内大场景，建议通过场景拆分、并行计算等手段实现高效重建。
- 2) 从精度角度来看，消费级相机视频重建的厘米级误差能够检测建筑物中关键构件的有无与大幅尺寸偏差，如建筑构件的施工进度跟踪、管道安装的错位检测等巡检任务。但依赖高精重建的尺寸测量将难以胜任。实际巡检任务建议使用精确标定的工业相机进一步提高重建的几何精度，适用于更广泛的巡检需求。

此外，目前针对室内场景中图像三维重建常见缺陷的考虑和评估仍然缺乏，这些问题会导致重建精度的降低，需要进一步研究。在未来，基于图像三维重建的室内巡检可以更多与语义分割^[5]、图像识别^[20]、图像处理^[21]等技术相结合，以实现更高效的巡检应用。

参考文献 (References)

[1] 智研咨询. 2025-2031 年中国智能巡检机器人行业市场现状调研及发展前景预测报告[R]. 北京: 北京智研科信咨询有限公司, 2024.

[2] MADAN H T, MAHESH L, SAJJAD M. ROS Powered Autonomous Mobile Robot for Indoor Applications[C]//2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS). Hassan, India, 2024: 1-7.

[3] 林佳瑞, 陈柯吟, 潘鹏. 建筑工程标准数字化与智能化: 现状与未来[J]. 东南大学学报(自然科学版), 2025,55(1): 16-29.

LIN JIARUI, CHEN KEYIN, PAN PENG. Digital and intelligent standards for building and construction engineering: Current status and future[J]. Journal of Southeast University (Natural Science Edition), 2025, 55(1): 16-29. (in Chinese)

[4] ZHOU S J, LIN J R, PAN P, et al. Impact of color and mixing proportion of synthetic point clouds on semantic segmentation[J]. Automation in Construction, 2025, 171: 105963.

- [5] 周绍杰, 潘鹏, 顾栋炼, 等. 融合 BIM 与三维重建的施工进度数字孪生跟踪方法初探[C]//第九届全国 BIM 学术会议, 西安, 中国, 2023: 453-457.
- [6] WIJAYATHUNGA L, RASSAU A, CHAI D. Challenges and solutions for autonomous ground robot scene understanding and navigation in unstructured outdoor environments: A review[J]. *Applied Sciences*, 2023, 13(17): 9877.
- [7] 胡欢, 孙汉旭, 贾庆轩. 基于在线重建的遥操作预测显示系统[J]. *东南大学学报(自然科学版)*, 2015, 45(3): 448-454.
HU HUAN, SUN HANXU, JIA QINGXUAN. On-line reconstruction-based predictive display system for teleoperation[J]. *Journal of Southeast University (Natural Science Edition)*, 2015, 45(3): 448-454. (in Chinese)
- [8] 李新德, 张晓, 朱博. 基于立体视觉的一般物体识别方法[J]. *东南大学学报(自然科学版)*, 2013, 43(4): 711-716.
LI XINDE, ZHANG XIAO, ZHU BO. Generic object recognition method based on stereo vision[J]. *Journal of Southeast University (Natural Science Edition)*, 2013, 43(4): 711-716. (in Chinese)
- [9] 刘志海, 代振锐, 田绍鲁, 等. 非接触式三维重建技术综述[J]. *科学技术与工程*, 2022, 22(23): 9897-9908.
LIU ZHIHAI, DAI ZHENRUI, TIAN SHAOLU, et al. A review of non-contact 3D reconstruction techniques[J]. *Science Technology and Engineering*, 2022, 22(23): 9897-9908. (in Chinese)
- [10] RAO A S, RADANOVIC M, LIU Y, et al. Real-time monitoring of construction sites: Sensors, methods, and applications[J]. *Automation in Construction*, 2022, 136: 104099.
- [11] SHEN Y, PENG P, GAO W. 3D reconstruction from a single family camera[C]//2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI). Würzburg, Germany, 2012: 108-112.
- [12] BAYATI H, NAJAFI A, VAHIDI J, et al. 3D reconstruction of uneven-aged forest in single tree scale using digital camera and SfM-MVS technique[J]. *Scandinavian Journal of Forest Research*, 2021, 36(2-3): 210-220.
- [13] BERRA E F, PEPPA M V. Advances and challenges of UAV SfM MVS photogrammetry and remote sensing: Short review[C]//2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS). Santiago, Chile, 2020: 533-538.
- [14] MA D, FANG H, WANG N, et al. A low-cost 3D reconstruction and measurement system based on structure-from-motion (SfM) and multi-view stereo (MVS) for sewer pipelines[J]. *Tunnelling and Underground Space Technology*, 2023, 141: 105345.
- [15] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas, NV, USA, 2016: 4104-4113.
- [16] PEPE M, ALFIO V S, COSTANTINO D. UAV platforms and the SfM-MVS approach in the 3D surveys and modelling: A review in the cultural heritage field[J]. *Applied Sciences*, 2022, 12(24): 12886.
- [17] HERMANN M, WEINMANN M, NEX F, et al. Depth estimation and 3D reconstruction from UAV-borne imagery: Evaluation on the UseGeo dataset[J]. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 2024, 13: 100065.
- [18] LU P, SHI W, QIAO X. Multi-View Image-Based 3D Reconstruction in Indoor Scenes: A Survey[J]. *ZTE Communications*, 2024, 22(3): 91.
- [19] PINTORE, G., MURA, C., GANOVELLI, F., et al. State-of-the-art in automatic 3D reconstruction of structured indoor environments[C]//Computer Graphics Forum, 2020, Wellington, New Zealand, 39(2), 667-699.
- [20] 郭俊熙, 潘鹏, 林佳瑞. 混凝土结构表观质量缺陷识别算法性能测评与分析[C]//第十届全国 BIM 学术会议, 杭州, 中国, 2024: 489-493.
- [21] 甘磊, 金洪杰, 沈振中. 混凝土表面粗糙度数字图像特征参数影响因素[J]. *东南大学学报(自然科学版)*,

2022, 52(3): 497-505.

GAN LEI, JIN HONGJIE, SHEN ZHENZHONG. Influencing factors of characteristic parameters of digital image on concrete surface roughness[J]. Journal of Southeast University (Natural Science Edition), 2022, 52(3): 497-505. (in Chinese)