引文格式:张守建,刘皓泽,陆海涛,等.基于3D高斯溅射的大规模场景重定位方法[J].武汉大学学报(信息科学版),2026,51(4):689-702.DOI:10.13203/j.whugis20250276

Citation:ZHANG Shoujian,LIU Haoze,LU Haitao,et al.Large-Scale Scene Relocation via 3D Gaussian Splatting in Urban Enviroment[J].Geomatics and Information Science of Wuhan University,2026,51(4):689-702.DOI:10.13203/j.whugis20250276

基于3D高斯溅射的大规模场景重定位方法

张守建[1] 刘皓泽[1] 陆海涛[1] 陈海吉尔[1] 徐 博[1] 赵 磊[2]

1 武汉大学测绘学院,湖北 武汉,430079

2 武汉大学珞珈实验室,湖北 武汉,430079

摘 要:为了获取复杂环境下精确位姿信息,提出一种利用3D高斯溅射技术(3D Gaussian splatting,3DGS)的大规模重定位框架,使用单目相机实现厘米级定位。首先,利用多传感器数据构建城市复杂环境高精度3DGS地图,然后,利用基于深度学习的特征点提取与匹配方法和的视觉前端结合,实现了高效、鲁棒的特征点提取和匹配,最后通过求解3D到2D点对的方法,迭代渲染优化得到精确的位姿信息。在开源街景数据集KITTI上选取3种场景道路对所提方法进行了测试,实验结果表明,在城镇道路、林荫大道和交通密集型公路上的平均定位精度分别为0.026 m、0.029 m和0.081 m,优于其他常规方法。

关键词:3D高斯溅射;重投影误差;重定位;户外场景;高精度位姿

中图分类号:P228 **文献标识码**:A **收稿日期**:2025-09-20

DOI:10.13203/j.whugis20250276 **文章编号**:1671-8860(2026)04-0689-14

Large-Scale Scene Relocation via 3D Gaussian Splatting in Urban Enviroment

ZHANG Shoujian[1] *LIU Haoze*[1] *LU Haitao*[1] *CHEN Haijier*[1] *XU Bo*[1] *ZHAO Lei*[2]

1 School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China

2 Luojia Laboratory, Wuhan University, Wuhan 430079, China

Abstract: **Objectives**: In the field of autonomous driving, acquiring high-precision pose information is a critical factor in ensuring its safety and reliability. Although global navigation satellite system (GNSS) can provide absolute pose references, its positioning accuracy significantly degrades in complex scenarios due to signal obstruction and multipath effects, making it difficult to meet the stringent requirements of autonomous driving systems. To address this challenge, we have conducted systematic exploration into precise pose acquisition in large-scale, complex, and dynamic environments. **Methods**: First, based on multi-sensor fusion technology, we developed a reconstruction system tailored for dynamic urban environments, focusing on three representative large-scale outdoor scenarios: urban roads, tree-lined avenues, and high-traffic highways. A high-precision 3D Gaussian splatter (3DGS) map was constructed for these scenarios. In the visual front-end, we integrated the SuperPoint feature extraction algorithm with the SuperGlue feature matching algorithm to achieve high-quality feature point detection and matching. Additionally, we introduced the perspective-*n*-point (P*n*P) algorithm, which estimates the initial camera pose by matching feature points with corresponding points in the 3D map. To optimize positioning results and improve the system's real-time performance, we proposed an iterative optimization strategy based on progressive rendering. This strategy utilizes the geometric structure and radiometric field characteristics of the 3DGS map, combining step-by-step rendering and optimization to continuously adjust the camera pose estimation, ultimately achieving progressive optimization of real-time high-precision positioning results. **Results**: The experimental validation further demonstrates the following results: (1) In dynamic urban environments, the reconstruc-

基金项目:湖北省重大攻关项目(JD)(2023BAA025);国家自然科学基金面上项目(42374016)。

第一作者:张守建,博士,副教授,主要从事卫星、机器人导航领域研究。shjzhang@sgg.whu.edu.cn

通信作者:徐博,博士,副研究员。boxu1995@whu.edu.cn

tion system exhibits exceptional performance, effectively rendering continuous and realistic lighting and texture effects in real-time. Particularly in complex urban settings, the system handles issues such as lighting changes and occlusions caused by dynamic objects effectively. (2) The feature extraction and matching algorithms, SuperPoint and SuperGlue, significantly outperform traditional methods in terms of both the quantity and quality of feature points. Compared to traditional methods, the number of extracted feature points has been greatly increased, with a feature point distribution rate reaching 70%, more than twice that of conventional methods. (3) The system demonstrates excellent positioning accuracy in three typical outdoor scenarios. On urban roads, tree-lined avenues, and high-traffic highways, the system achieves average positioning accuracies of 0.026 m, 0.029 m, and 0.081 m, respectively, which are significantly better than other representative methods. Additionally, the system achieves a frequency of 96% for absolute displacement and rotation errors smaller than 10 cm and 1°, respectively, indicating high stability in positioning accuracy and the ability to provide reliable real-time positioning results in various environments. (4) Ablation experiments further validate the effectiveness of the system's design. The complete model achieves the best results in both typical scenarios, with average absolute translation and rotation errors of 3.22 cm and 0.69°, respectively. Compared to experiments where dynamic occlusion handling, SuperPoint feature extraction, and iterative optimization strategies were removed, the system's overall design demonstrates superior performance. (5) The system's real-time performance has been verified. Time analysis shows that the system processes each frame of images in approximately 1 s, which meets the real-time requirements and is suitable for real-time application scenarios. **Conclusions**: In large-scale outdoor environments, the visual relocalization technology based on the 3DGS map enables high-precision location information retrieval, effectively compensating for the limitations of existing positioning methods under GNSS signal occlusion or in complex environments. Through multi-sensor fusion and precise map construction, the system performs exceptionally well in dynamic urban scenarios. Particularly in complex urban environments, the issues of lighting changes and dynamic object occlusion are effectively handled, significantly enhancing the system's robustness and adaptability. The front-end feature extraction and matching methods greatly improve the system's stability and accuracy. In complex dynamic scenes, the distribution of feature points is more uniform, further enhancing positioning precision. Compared to various methods, the system demonstrates optimal performance. Ablation experiments further validate the necessity and effectiveness of these techniques. The system's processing speed meets real-time requirements, ensuring efficient and reliable real-time localization.

Key words: 3D Gaussian splatting; reprojection error; relocalization; outdoor scenes; high-precision pose

在复杂环境中实现稳定且高精度的定位,是推动自动驾驶等无人系统发展的关键环节。当前的车辆定位通常依赖全球导航卫星系统(global navigation satellite system, GNSS)、激光雷达(light detection and ranging,LiDAR)、惯性测量单元(inertial measurement unit,IMU)以及摄像头等多源传感器。然而,GNSS在城市街区常受到遮挡与多径效应的影响,导致定位结果不可靠[1];同时,基于IMU、LiDAR或相机的惯性导航与里程计方法在退化场景中也容易出现累积误差[2-4]。为克服这些问题,研究者普遍引入预先构建的高精度地图作为辅助信息,这些地图包含绝对位置与姿态的关键信息,有助于实现精准定位[5]。

基于地图的定位方式大体可分为激光雷达匹配和视觉重定位两类。相较于高成本的激光雷达方案,视觉重定位在成本、部署与维护方面更具优势,同时相机能够捕获丰富的纹理与光照信息,有助于提升环境感知与定位精度。现有的视觉重定位方法主要包括基于特征匹配、场景坐标回归[6]、位姿回归[7]以及直接图像对齐[8]等方法。其中,基于二维(two dimension,2D)-三维(three dimension,3D)特征匹配的方法仍占据主导地位,该方法通过从2D图像中提取特征点,并与3D地图特征进行匹配来估计相机位姿。然而,这类方法的整体性能高度依赖于3D地图的表示能力。若地图仅包含稀疏特征点或纯几何模型,往往难以在复杂光照、动态遮挡或无纹理区域中保持稳定的匹配效果。因此,提升地图表

示的表达力与渲染能力，成为进一步增强 2D-3D 匹配定位精度与鲁棒性的关键途径。

随着技术的进步，用于匹配定位的地图表示方法也在不断变化。与传统的矢量地图或点云地图等固定格式不同，神经辐射场（neural radiance fields，NeRF）[9]和 3D 高斯溅射（3D Gaussian splatting，3DGS）等新型表示方法能更细致地描绘场景信息[10-11]。这些创新方法能够从任意视角生成高保真图像，为基于这些先进地图表示的重定位研究提供了新机遇[12]。NeRF 提供了一种连续体积场景建模与新视角合成范式，在位姿估计中展现出极高精度。然而，其反投影与渲染过程计算开销巨大，训练与推理效率不足，且在动态场景中的适应性有限。相比之下，3DGS 具备更高的渲染效率与更好的动态适应性，同时能够自然地融合多模态信息，因此更适合实时定位等应用场景[13-14]。该技术支持快速渲染，并能无缝集成多模态数据，使其在自动驾驶等实时应用中具有显著优势。ZHAI 等[15]首次引入 3DGS 表示进行重定位，实现了高效渲染与位姿估计；XIN 等[16]通过隐式特征度量对齐进一步提升了复杂环境下的鲁棒性；BOTASHEV 等[17]则结合由粗到细的优化策略与模糊核设计，有效缓解了非凸优化带来的不稳定问题；而 ZELLER 等[18]通过最小化渲染深度图与实测深度图的差异，在低纹理场景下依然能够获得高精度的位姿估计。与上述方法不同，本文针对复杂环境下车辆定位的挑战，提出了一种基于 3D 高斯地图的高精度视觉重定位方法，该方法并未依赖可学习姿态先验或可微渲染优化，而是基于静态构建的 3DGS 场景表示，结合透视-n-点（perspective-n-point，PnP）进行几何求解，实现从图像观测到位姿的直接推断。随后，通过基于渲染一致性的迭代优化策略，在不引入梯度反传或联合场景更新的情况下，自适应地细化相机位姿，从而获得全局最优解。该设计在保持系统轻量化与实时性的同时，兼具几何优化的确定性与基于渲染评估的全局鲁棒性，体现出更高的计算效率与收敛稳定性，实现了迭代优化下的高精度位姿估计，为无人系统在复杂环境中的稳定可靠定位提供了有力支持。

1　3DGS 地图构建方法

在自动驾驶所面临的大规模室外场景中，由于单一视角的限制、光照对比度过高以及动态目标数量的增加，传统的 3DGS 图像构建方法容易出现视角盲区。这不仅影响了深度信息和物体 3D 感的获取，还会导致图像出现过曝或阴影细节丢失，从而产生伪影现象，进而降低 3D 重建的精度。为应对这一问题，本文采用了一种面向动态城市场景的重建框架——OmniRe[19]系统。该系统具备全景式的建模能力，能够处理场景中的多种动态要素，并通过融合激光雷达与图像信息，有效克服单一视点的局限性，提供更加可靠的深度信息，生成清晰且无模糊的渲染结果。与此同时，通过构建动态神经场景图，利用高斯表示对场景中的不同动态主体进行建模，并为各类动态实体分配专属的局部规范化空间，实现了对复杂动态场景的全面捕捉与重建。该方法能够以高保真度还原场景，保持细节的完整性，从而满足自动驾驶高精度定位所需的高质量 3DGS 地图模型构建需求。

在具体实现中，为了有效表征各类动态对象，所提出的高斯场景图由天空节点、背景节点、刚性节点和非刚性节点组成。其中，背景节点采用一组静态高斯 G_b 表示，并通过激光雷达点的累积以及随机生成的补充点进行初始化[20]；刚性节点在局部空间中被表示为 $\bar{G}_v$，随后映射至全局世界坐标系的过程可以表示为：

$$G_v = T_v(t) \otimes \bar{G}_v \tag{1}$$

式中，G_v 表示在世界坐标系的位置；$T_v(t)$ 表示 t 时刻二者的位姿变换矩阵。非刚性节点进一步细分为两类，一类为用于表示行走或奔跑行人的蒙皮多人线性模型（skinned multi-person linear model，SMPL）节点，另一类为用于表示分布外非刚性可变形实例的可变形节点。两类节点在世界坐标系中表示为：

$$\begin{cases} G_{\mathrm{SMPL}}(t) = T_h(t) \otimes \mathrm{LBS}(\theta(t), \bar{G}_{\mathrm{SMPL}}) \\ G_{\mathrm{deform}}(t) = T_h(t) \otimes (\bar{G}_{\mathrm{deform}} \otimes F_\ell(\bar{G}_{\mathrm{deform}}, e_h, t)) \end{cases} \tag{2}$$

式中，$G_{\mathrm{SMPL}}(t)$ 与 $G_{\mathrm{deform}}(t)$ 分别表示两类节点在时间 t 的世界空间中的 3D 高斯分布；$T_h(t) \in \mathrm{SE}(3)$，表示节点 h 在时间 t 的全局姿势；LBS() 表示线性混合蒙皮（linear blend skinning，LBS）计算；$\theta(t)$ 表示随时间 t 变化的人体姿势的模型参数；$\bar{G}_{\mathrm{SMPL}}$ 表示从静息姿势（Da 姿势）实例化的 SMPL 模板网格 M_h 的 3D 高斯分布；$\bar{G}_{\mathrm{deform}}$ 为规范空间中的初始高斯分布，表示节点在未变形状态下的位置和形状属性；$F_\ell()$ 表示参数为 ℓ 的通用形变网络；e_h 表示节点的嵌入参数，这 3 类节点会被统一转换为世界坐标系下的高斯表示，然后通过光栅化器进行渲染与拼接。同时，天空节

点则采用优化后的环境纹理贴图进行建模,并独立渲染。最终,来自天空节点的渲染结果C_{sky}会与其他节点的整体渲染结果C_G相结合,从而生成最终输出,表示为[21]:

$$C = C_G + (1 - O_G)C_{\text{sky}} \tag{3}$$

式中,C表示最终渲染得到的图像颜色;$O_G = \sum_{i=1}^{N}\alpha_i\prod_{j=1}^{i-1}(1-\alpha_j)$表示渲染得到的不透明度掩码,$\alpha_i$、$\alpha_j$分别表示第$i$和第$j$个高斯粒子的不透明度。为了获得最优的整体渲染效果,将上述所有参数在同一阶段中进行联合优化,其优化函数定义为:

$$\zeta = \zeta_{\text{img}} + \lambda_{\text{depth}}\zeta_{\text{depth}} + \lambda_{\text{opacity}}\zeta_{\text{opacity}} + \zeta_{\text{reg}} \tag{4}$$

式中,ζ_{img}为基于图像的重建损失,具体定义如下:

$$\zeta_{\text{img}} = (1 - \lambda_r)\zeta_1 + \lambda_r\zeta_{\text{SSIM}} \tag{5}$$

其中,ζ_1表示用于度量渲染图像与目标图像在像素层面差异的L1损失;ζ_{SSIM}表示渲染图像的结构相似性(structure similarity index measure,SSIM)损失,对于由渲染动态区域检测到的像素,因其时空稀疏性导致监督信号不足,对该区域加权放大图像损失,权重设为5;λ_r在实验中设置为0.2;$\lambda_{\text{depth}}\zeta_{\text{depth}}$为深度一致性约束,其中$\zeta_{\text{depth}}$为渲染得到的高斯深度值与来自LiDAR数据的稀疏深度信号之间的差异,表示施加于不同高斯表示上的正则化约束,λ_{depth}设定为0.1;$\lambda_{\text{opacity}}\zeta_{\text{opacity}}$为不透明度约束,$\lambda_{\text{opacity}}$设定为0.05;$\zeta_{\text{reg}}$为应用于不同高斯表示的各种正则化项。$\zeta_{\text{depth}}$和$\zeta_{\text{opacity}}$的计算公式如下:

$$\zeta_{\text{depth}} = \frac{1}{hw}\sum_{i=1}^{h}\sum_{j=1}^{w}\left\| D_{ij}^{s} - \hat{D}_{ij} \right\|_1 \tag{6}$$

$$\zeta_{\text{opacity}} = -\frac{1}{hw}\sum O_G \lg O_G - \frac{1}{hw}\sum M_{\text{sky}} \lg(1 - O_G) \tag{7}$$

式中,D^s为稀疏LiDAR深度图的逆深度;$\hat{D}$为渲染预测的逆深度图;M_{sky}为天空掩膜;h表示图像的高度;w表示图像的宽度。

2 增强型特征点匹配方法

2.1 特征提取

在本系统中,前端的核心任务是特征点的提取与匹配[22]。传统方法如尺度不变特征变换+快速近似近邻算法库和ORB(oriented FAST and rotated BRIEF)+暴力匹配(brute-force matcher,BFMatcher)[23],在一定程度上能够满足实时性要求,但其性能受限于特征点选择,对噪声和遮挡较为敏感,容易造成信息丢失。尤其在自动驾驶所面对的大规模复杂场景中,这类方法往往难以在稀疏候选库中找到最优匹配图像,同时也难以保证足够数量和精度的特征匹配点,限制了系统的整体鲁棒性。

针对上述问题,本文采用SuperPoint+SuperGlue[24-25]的深度特征提取与匹配框架。SuperPoint用于在图像中提取特征点及其描述符,而SuperGlue则实现跨图像的特征匹配。该组合方法将特征点检测、特征描述与匹配三者有机融合,构建了一个端到端的深度网络前端模块[26]。与传统方法相比,该方法在车辆高速运动导致的图像模糊,以及光照缺失或遮挡等恶劣条件下依然能够保持稳定表现,并在特征点数量与匹配精度方面显著优于传统方法。

具体而言,SuperPoint是一种基于自监督训练的深度网络,能够同时检测大规模特征点并生成高维、定长的描述符[27]。该网络由编码器、特征点解码器和描述符解码器组成:在编码阶段,特征点检测与描述符提取共享同一编码结构;在解码阶段,则依据任务需求采用不同的网络结构,以分别学习特征点和描述符的独立参数[24]。

2.2 特征匹配

SuperGlue是一种新型的图神经网络算法[28],能够在特征匹配过程中高效剔除离群点。其核心思想是利用可微分的最优传输理论进行特征匹配,并结合注意力机制,将2D特征点信息与聚合策略相融合,从而获得最终图像特征点的整体精确匹配结果。这样,SuperGlue不仅能够处理传统方法提取的特征点与描述符,而且能高效利用深度学习生成的特征信息,从而获得更为准确的匹配结果。

3 地图融合的重定位方法

整体来看,重定位系统主要分为3部分:3DGS地图渲染、增强型特征提取与匹配以及精细化视觉重定位。系统整体结构如图1所示。

3.1 初始匹配以确定粗略位置

在确定目标位置的初始位姿时,首先利用SuperGlue算法在参考图像库中对目标查询图像与参考红绿蓝(red green blue,RGB)图像进行特征匹配,并结合空间上下文信息与全局特征对匹配结果进行优化,选取相似度最高的参考图像I。其中,参考图像库由3DGS生成的多幅RGB图像构成。当获取该参考图像对应的六自由度位姿参数后,将其作为目标点的初始粗略位姿。

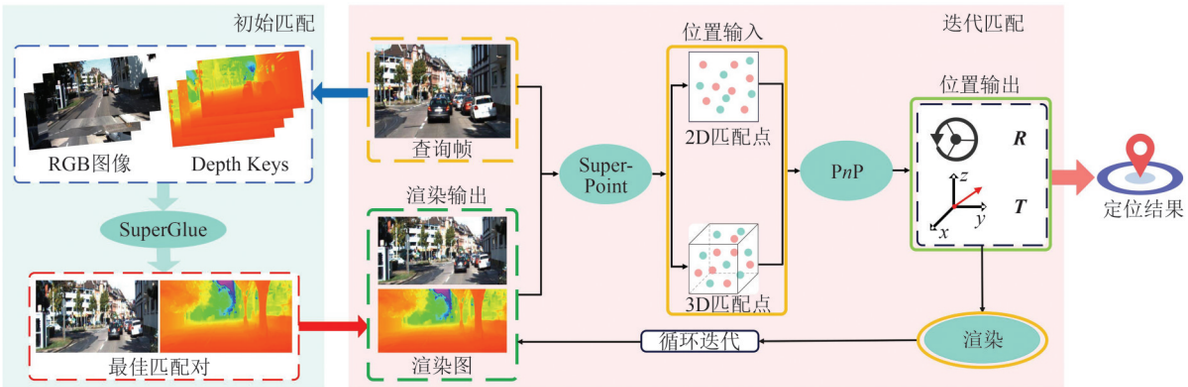


图 1　融合 3DGS 地图的重定位系统

Fig. 1　Relocalization System with Integrated 3DGS Maps

3.2　PnP 精确姿态解算

在粗定位阶段，目标图像会与参考图像进行匹配。参考图像由 3DGS 地图生成的 RGB 图像与深度图像组成。对于每一个空间点，能够在相机坐标系下获得其 3D 坐标(X_G, Y_G, Z_G)，这些坐标由空间点的齐次坐标与归一化平面投影得到的特征点共同计算而来。在此基础上，采用 PnP 方法，根据初始渲染得到的参考图像，进一步求解目标图像位置的精确相对位姿。

高效 PnP（efficient PnP，EPnP）是一种高效求解 PnP 问题的方法[29]，其目标函数可表示为：

$$E_{\mathrm{EP}n\mathrm{P}}=\sum_{i=1}^{N}\left\| \boldsymbol{u}_i-\pi\left(\boldsymbol{R}_{\mathrm{WC}}\boldsymbol{P}_i+\boldsymbol{t}_{\mathrm{WC}}\right)\right\|^2 \tag{8}$$

式中，$\boldsymbol{P}_i\in\boldsymbol{R}^3$ 为世界坐标系下的空间点；$\boldsymbol{u}_i=\left[u_i\ v_i\right]^{\mathrm{T}}$ 为对应的图像观测；$\boldsymbol{R}_{\mathrm{WC}}$ 与 $\boldsymbol{t}_{\mathrm{WC}}$ 分别表示相机在世界坐标系（world coordinate，WC）中的旋转和平移；$\pi(\ )$ 为相机投影函数。EPnP 通过将 3D 点表示为 4 个虚拟控制点的加权线性组合，从而将求解过程线性化，实现了在保持精度的同时显著提高位姿估计效率。

该方法计算量较小，在处理大量特征点时具有较高的计算效率。其核心思想是利用非共面的虚拟控制点，通过线性加权的方式表示相机坐标系下的任意特征点。因此，算法首先需要确定虚拟控制点与特征点在相机坐标系中的位置，并据此建立世界坐标系与相机坐标系之间的关系，可以表示为：

$$\begin{cases}\boldsymbol{P}_i^k=\sum_{j=1}^{4}a_{ij}\boldsymbol{C}_j^k\\ \boldsymbol{P}_i^m=\sum_{j=1}^{4}a_{ij}\boldsymbol{C}_j^m\\ i=1,2,\cdots,n\end{cases} \tag{9}$$

式中，$\boldsymbol{P}_i^k$ 与 $\boldsymbol{P}_i^m$ 分别表示标记点在 WC 系和对应相机坐标系中的位置；$\boldsymbol{C}_j^k$ 与 $\boldsymbol{C}_j^m$ 分别表示虚拟控制点在世界坐标系和相机坐标系中的位置；i 对应一对匹配点；j 表示虚拟控制点；a_{ij} 为每个标记点对应的加权系数，且所有系数之和为 1。为了求解相机坐标系下虚拟控制点的位置，构建方程表达式为：

$$Z_c\begin{bmatrix}u_i\\ v_i\\ 1\end{bmatrix}=\boldsymbol{K}\begin{bmatrix}\boldsymbol{R} & \boldsymbol{t}\\ 0 & 1\end{bmatrix}\begin{bmatrix}X_i^m\\ Y_i^m\\ Z_i^m\\ 1\end{bmatrix} \tag{10}$$

式中，$\boldsymbol{K}$ 为相机内参矩阵；$\begin{bmatrix}X_i^m & Y_i^m & Z_i^m & 1\end{bmatrix}^{\mathrm{T}}=\boldsymbol{C}_j^m$ 为待求解未知量，即相机坐标系下虚拟控制点的坐标；u_i、v_i 为像素坐标系下坐标；$\boldsymbol{R}$ 为旋转矩阵；$\boldsymbol{t}$ 为位移矩阵。随后，利用绝对定位方法[30]求解目标点的六自由度位姿，其优化目标可表示为：

$$E_{\mathrm{ICP}}=\sum_{i=1}^{N}\left(\boldsymbol{n}_i^{\mathrm{T}}\left(\boldsymbol{R}p_i+\boldsymbol{t}-q_i\right)\right)^2 \tag{11}$$

式中，ICP（iterative closest point）：迭代最近点；p_i 与 q_i 分别表示源点云与目标点云中对应点的坐标；$\boldsymbol{n}_i$ 为目标点处的法向量。该方法通过最小化点到平面的距离误差，实现源点云与目标点云的精确对齐与位姿估计。

为了充分利用更有效的信息并降低匹配特征点间噪声的影响，进一步构建最小二乘优化问题，其表达式为：

$$E_{\mathrm{BA}}=\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left\| u_{ij}-\pi\left(R_iP_j+t_i\right)\right\|^2 \tag{12}$$

式中，BA（bundle adjustment）：光束法平差；P_j 为三维空间点；u_{ij} 为像素观测，该优化通过联合调整相机位姿与空间点位置，以最小化所有观测的重投影误差，从而获得全局一致的最优估计结果。

将像素坐标与空间点位置关系矩阵的误差以代数方式最小化，调整位姿估计值，其优化模

型可以表示为[31]:

$$T^{*}=\arg\min_{T}\frac{1}{2}\sum_{i=1}^{n}\left\|u_i-\frac{1}{s_i}\boldsymbol{K}\boldsymbol{T}P_i\right\|_2^2 \tag{13}$$

式中,T^{*}表示优化后的位姿;T表示当前的位姿;u_i表示特征点的像素坐标;$\boldsymbol{T}$表示变换矩阵。

3.3 迭代渲染确定精确位姿

考虑到单次PnP求解的定位精度有限,本文对初步位姿进行基于渲染反馈的迭代精化,如图2所示。具体而言,每次迭代将最新估计位姿输入3DGS地图,渲染生成对应的RGB图像与深度图,并与观测图像进行像素级配准优化以更新位姿。

本文在迭代精化阶段采用单尺度优化策略,即仅在原始分辨率下进行渲染、匹配与位姿更新,不使用多尺度或金字塔层级结构。该设置在保证计算效率的同时,已能在精确初值条件下获得稳定且快速的收敛效果。

在优化过程中,使用Huber鲁棒核函数减弱异常匹配点的影响,并对输入图像进行动态物体及遮挡区域剔除,以提高优化的稳定性与鲁棒性。迭代过程以位姿变化小于1 cm或0.1°作为收敛条件,若未满足,则继续迭代,最多执行10次。对比实验表明,该方法在大规模自动驾驶场景下能够基于3DGS地图和视觉观测实现快速、高精度的重定位,验证了所提迭代优化策略的有效性与鲁棒性。

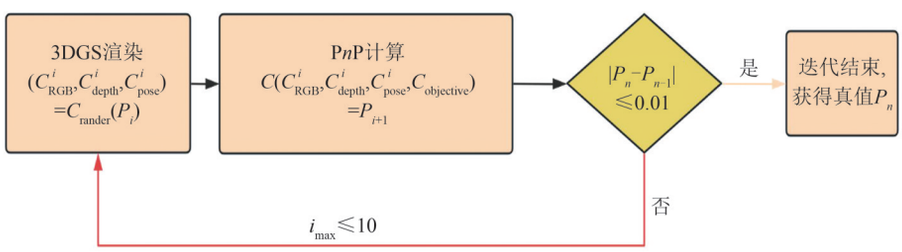


图2 系统迭代步骤

Fig. 2 System Iteration Steps

4 实验分析

本文使用卡尔斯鲁厄理工学院与丰田技术研究院提供的KITTI数据集进行3DGS地图构建及重定位实验评估。KITTI数据集是面向自动驾驶研究的标准数据集,包含丰富的真实驾驶场景。在实验中,首先利用相机、IMU惯性测量设备、Velodyne HDL-64E S2激光雷达传感器数据以及场景物体的MASK信息,完成大规模场景的高精度3DGS地图构建;然后仅使用相机提供的单幅图像进行重定位。重定位的真实位姿参考值基于数据集提供的载体位置真值及相机外参矩阵计算得出。

本文的定位算法在配备Intel Core i7 CPU、16 GB内存及NVIDIA RTX 4090 GPU(24 GB显存)的系统上实现,算法基于Python与PyTorch开发,并使用自定义CUDA内核加速微渲染方法中的光栅化与反向传播过程。该配置保证了算法的实时性能,对于重定位系统的实际应用至关重要。

4.1 3DGS地图构建

为了验证所提方法在大规模真实场景中的定位性能,首先基于开源数据集进行了3DGS地图的构建与可视化。图3展示了系统在KITTI数据集中的城市和林荫道路场景的建图效果。可以看出,所生成的3DGS地图能够高保真地重建道路结构、建筑立面及周围环境的空间分布,具有连续且逼真的光照与纹理表现。相比传统点云或网格地图,该表示在保证细节精度的同时显著提升了渲染效率与视觉一致性。该高精度环境模型为后续基于单幅图像的重定位提供了可靠的几何与外观先验。



(a) 城镇道路 (b) 林荫道路

图3 3DGS城镇、林荫道路建图效果

Fig. 3 Mapping Effect of 3DGS on Towns andTree-Lined Roads

4.2　图像特征点匹配比较

首先基于已建立的大规模室外3DGS地图渲染部分2D RGB图像及其对应的深度图，作为重定位查询图像的初始匹配基础地图库。为了保证基础地图库的轻量性以及查询图像匹配的实时性，基础地图库采用一定程度的稀疏渲染，即在相邻渲染位置达到一定距离后仅渲染一幅图像，而非密集渲染。然而，在大规模室外场景中，稀疏渲染会导致查询图像与图库中即便是最优匹配的RGB图像存在较大视角差异。因此，即便在视角差异较大的情况下，仍需确保查询图像能够准确选择最优匹配图像。为此，本文在前端特征点提取与匹配模块中采用了SuperPoint + SuperGlue的组合方法。

图4对比展示了传统方法(ORB+BFMatcher)与本文方法(SuperPoint+SuperGlue)在3种典型场景下的特征提取与匹配效果。从图4可以看出，在大规模场景下，组合方法的特征点提取与匹配能力明显优于传统方法。据统计，ORB+BFMatcher在匹配数量上并不输于SuperPoint+SuperGlue方法，但是在准确性上远远差于SuperPoint+SuperGlue方法，其中3个场景传统方法特征点的分布率仅为40%左右，但是SuperPoint+SuperGlue可以达到70%，同时从图4中匹配情况也可看出，ORB+BFMatcher的错误率非常明显。

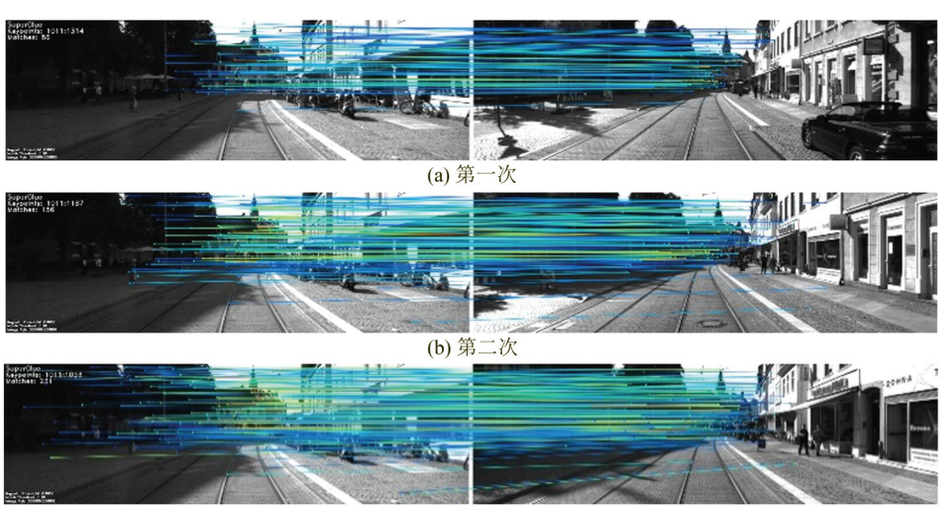


(a) ORB+BFMatcher　　(b) SuperPoint+SuperGlue

图4　ORB+BFMatcher与SuperPoint+SuperGlue匹配结果的比较

Fig. 4　Comparison of Matching Results Between ORB+BFMatcher and SuperPoint+SuperGlue

此外，随着3DGS的逐次迭代，渲染图像与查询图像之间的特征匹配效果也逐渐提升，如图5所示。在迭代过程中，可以观察到，3DGS地图渲染的新图像与查询图像之间的视角差距逐步缩小，图像间匹配的特征点数量逐渐增加，从而使相对位移计算的精度逐步提升，这与重定位过程的迭代推进相一致。



(a) 第一次

(b) 第二次

(c) 第三次

图5　迭代渲染中的SuperGlue匹配结果的比较

Fig. 5　Comparison of Matching Results of SuperGlue in Iterative Rendering

表1通过特征点匹配数量、特征匹配平均置信度以及特征点分布均匀性等指标定量说明了迭代对特征点提取和匹配的改进效果。其中，特征点匹配数量反映两幅图像之间的特征点匹配

总数;特征匹配平均置信度表示所有匹配点的可靠性;特征点分布均匀性则为特征点分布标准差归一化后与1的差值,越接近1,表明特征点在两幅图像中分布越均匀。

表1 不同迭代次数匹配效果的比较

Table 1 Comparison of Matching Effects with Different Iteration Times

迭代次数	特征点匹配数量/个	特征匹配的平均置信度	特征点分布的均匀性
1	58	0.735	0.354
2	126	0.826	0.466
3	257	0.892	0.598
4	383	0.957	0.701

4.3 重定位结果

本文在KITTI数据集的3个典型场景中对所提方法的重定位精度进行了测试,分别为城镇道路、林荫道路以及车流密集高速公路。图6展示了每个场景中选定序列若干图像的重定位结果与参考真值之间的误差分布。

从图6可以看出,城镇道路的定位精度最佳,林荫道路定位精度次之。这可能与图7的对比结果有关,图7(a)、7(b)显示,在渲染森林等光影复杂多变的场景时,相较于城镇道路,渲染图像存在一定程度的模糊(红框标识),这影响了特征点的提取及后续相对定位计算的精度。在车流密集的高速公路场景中,重定位计算的平均位置误差和平均角度误差在3类场景中相对较大,其原因可能是:虽然3DGS构建方法有效避免了动态物体的重影效应,但在特征退化的高速公路环境中,图像间的特征点匹配较少(图7(c)),且匹配分布均匀性较差,从而导致最终平均误差增大。

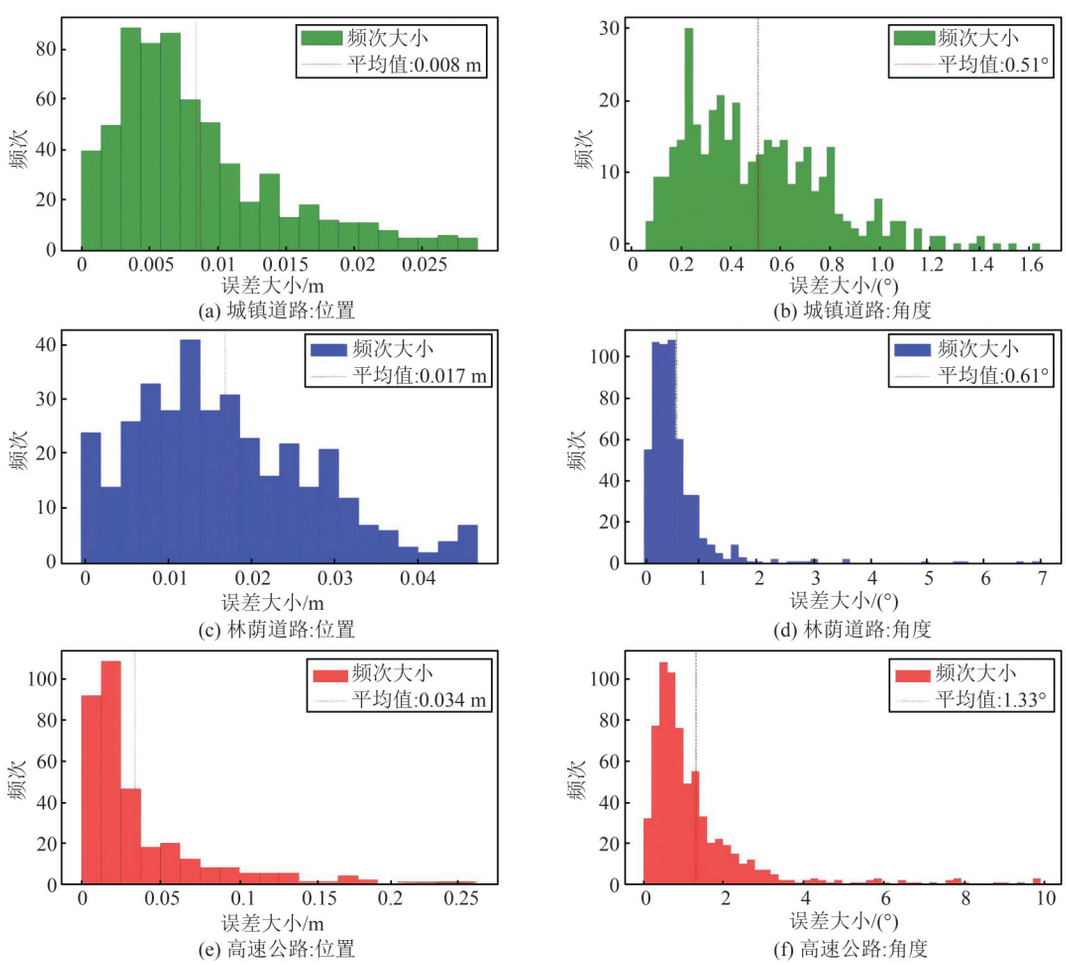
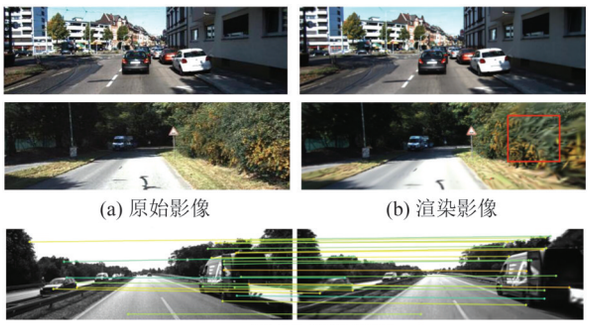


图6 城镇道路、林荫道路和高速公路场景中位置和角度定位误差的直方图

Fig. 6 Histograms of Position and Angle Positioning Errors in Urban Roads, Tree-Lined Paths, and Highway Scenes

分别在3个序列中随机选择一段长度为300 m的时间区间,将该区间内车辆轨迹中摄像头采集的图像按时间顺序排列,依次作为查询图像进行重定位,并对得到的轨迹结果进行统计分析。表2汇总了3类场景下重定位结果与参考真值的绝对轨迹误差,包括均方根误差(root mean square error,RMSE)、标准差、平均值、中位数、最大值与最小值。



(a) 原始影像　(b) 渲染影像

(c) 高速公路匹配图

图7　精度影响分析

Fig. 7　Precision Impact Analysis

表2　不同场景下的连续轨迹误差/m

Table 2　Continuous Trajectory Errors in Different Scenarios/m

序列	RMSE	标准差	平均值	中位数	最大值	最小值
城镇道路	0.026	0.018	0.019	0.015	0.209	0.003
林荫道路	0.027	0.023	0.014	0.013	0.246	0.008
高速公路	0.079	0.069	0.038	0.018	0.522	0.006

由表2可知,城镇道路的绝对轨迹误差RMSE和标准差最小,分别为0.026 m和0.018 m;林荫道路的平均轨迹误差最小,为0.013 m。车流密集道路的绝对轨迹误差统计值相对较大,因此城镇道路和林荫道路的RMSE均接近0.02 m,精度最佳;高速公路场景下效果略差,但RMSE仍接近0.07 m。结合图8~10轨迹图可见,3类场景整体定位精度较高。同时,3个场景的误差标准差分别为0.018 m、0.023 m和0.069 m,显示绝对轨迹误差具有良好的鲁棒性。综合来看,该系统在车辆连续运行的情况下,仍能满足高精度定位的需求。

将重定位计算得到的轨迹与真实轨迹进行对比,结果如图8至图10所示。图中对估计轨迹与真实轨迹最大偏差的位置进行了放大处理,可以观察到,在位移误差方面,车流密集高速公路场景的序列末端Z轴存在明显异常,误差达到0.7 m,这也是该场景平均定位精度明显低于其他两个场景的原因。此时高速公路场景中移动车辆过多,虽然3DGS地图渲染方法能够在很大程度上避免运动模糊,但在极端情况下仍会出现部分模糊,从而影响重定位精度。

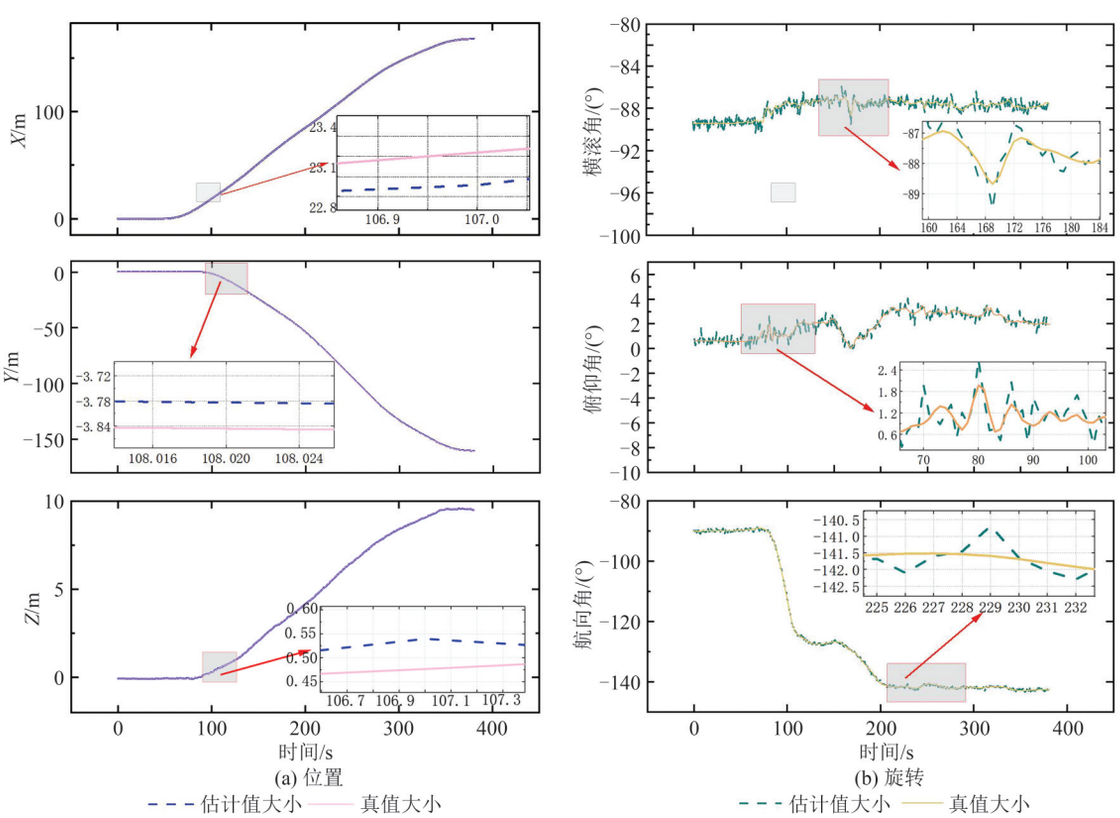


图8　城镇道路预测轨迹与真实轨迹误差比较(序列1)

Fig. 8　Comparison of Prediction Trajectory with Real Trajectory Error for Urban Roads (Sequence 1)

为了验证方法的有效性,将本文方法与5种代表性的重定位方法进行了对比,包括AS(active search)[32]、神经引导随机采样一致(neural-guided random sample consensus, NG-RANSAC)

算法[33]、Regression-only、DSAC++[34]和3DGS-ReLoc[35]。AS是一种基于特征点匹配的方法,通过结合帧到点与点到帧搜索机制,利用协方差关系优化匹配过程,并通过直接线性变换与

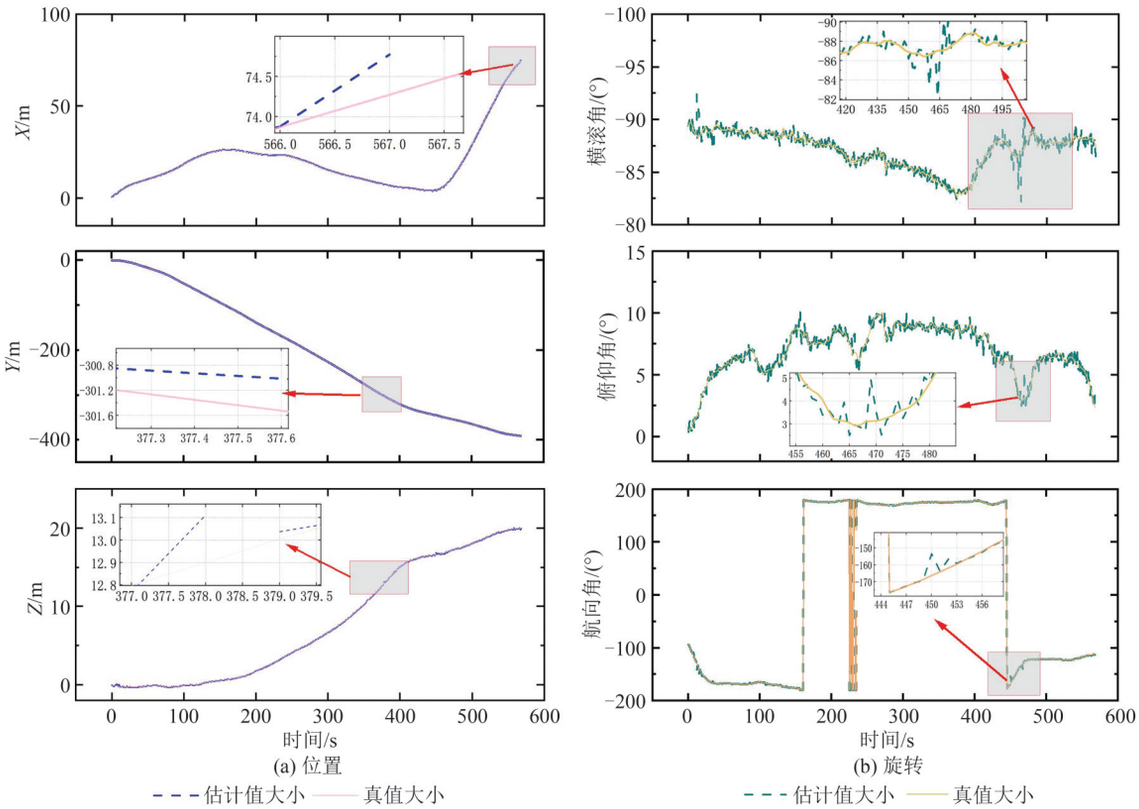


图9 林荫道路预测轨迹与真实轨迹误差比较(序列2)

Fig. 9 Comparison of Prediction Trajectory with Real Trajectory Error for Tree-Lined Paths (Sequence 2)

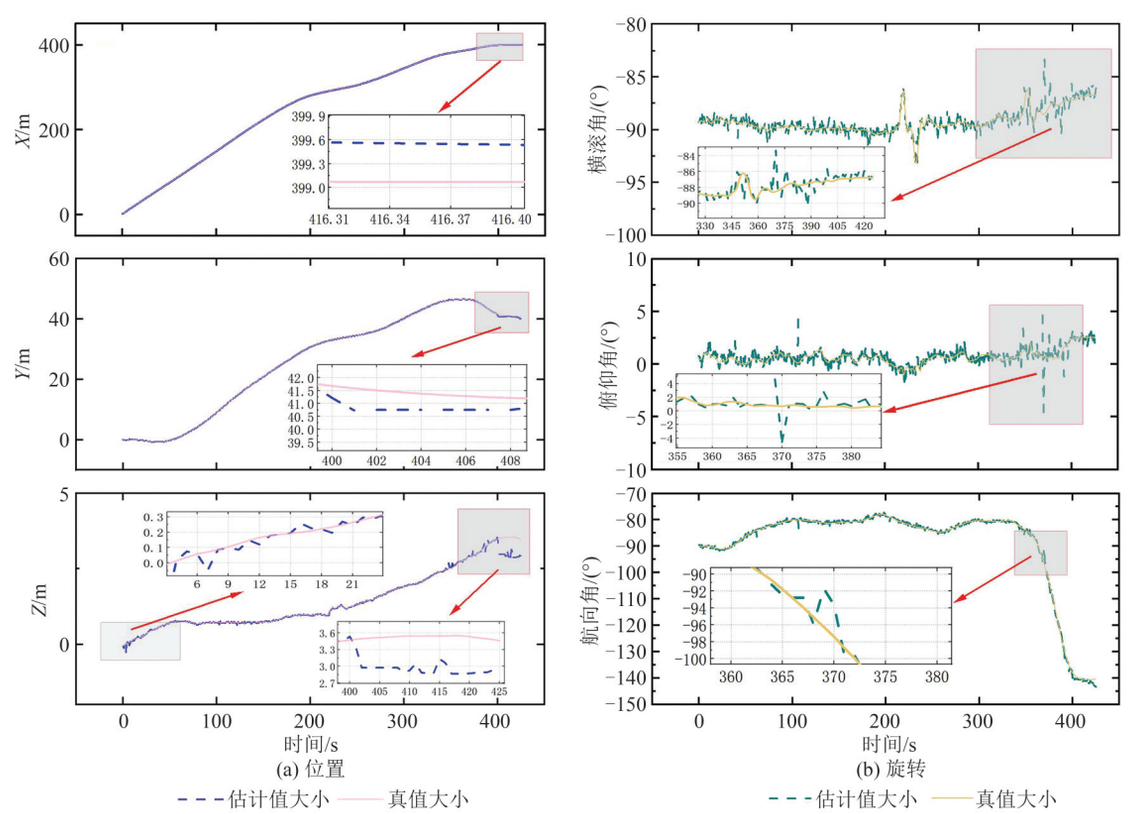


图10 高速公路预测轨迹与真实轨迹误差比较(序列3)

Fig. 10 Comparison of Prediction Trajectory with Real Trajectory Error for Highway Scenes (Sequence 3)

RANSAC算法求解位姿，适用于纹理丰富的场景。NG-RANSAC是经典RANSAC算法的改进版本，通过优化采样策略与计算流程，提高了运行效率与鲁棒性。DSAC++是一种基于深度学习的重定位方法，通过可微分RANSAC算法和卷积神经网络预测场景坐标，再使用P*n*P算法估计相机位姿，具有较高的定位精度和泛化能力。Regression-only方法直接通过神经网络从图像预测相机位姿，无需特征点提取，计算速度快，且抗噪能力强。3DGS-ReLoc利用3D高斯渲染技术进行重定位，理念与本文方法相似，但需要额外传感器（如GPS）提供粗略初始位姿。这些对比涵盖了从传统特征匹配到深度学习的多种技术，充分展示了本文方法的优势。

实验在KITTI、Waymo Open和nuScenes 3个公开数据集上进行评测，涵盖了城镇道路与开阔道路两类典型场景。为确保结果的可复现性与公平性，在各数据集上统一了序列选择、采样间隔、数据划分、视觉前端、检索策略及评价脚本。具体而言，KITTI数据集中选取了00、02、05、09和10 5个序列，其中00与05主要为城镇道路场景，02、09与10则包含更多开阔道路与郊区环境。所有序列以10 Hz采样，并划分为70%训练集、10%验证集和20%测试集。在Waymo Open数据集中，选取San Francisco downtown与Phoenix highway区域序列，分别代表城镇和高速公路场景，采样频率与数据划分策略与KITTI保持一致。nuScenes数据集采用官方提供的Boston与Singapore城市子集，其中前者以城镇道路为主，后者则包含部分开阔道路路段。

地图均通过自建的3DGS流程构建，平均每帧包含约1.5×10^6个高斯点。初始化阶段利用深度一致性与相机位姿的联合优化实现全局对齐，损失函数由颜色重建项、几何稀疏性项和不确定度约束项共同组成。为提升鲁棒性，在地图构建过程中通过语义分割与深度一致性检测剔除动态物体及遮挡区域，从而保证场景几何的一致性与渲染稳定性。

需要说明的是，表3中的所有对比方法均基于其原论文提出的模型结构和算法流程，在本文实验环境中统一复现完成。所有方法均在相同的3DGS地图、训练与测试划分、视觉前端、检索策略以及评价脚本下进行评测，以确保结果的公平性与可比性。

此外，本文的主要目标是实现车辆在全局坐标系下的绝对位姿估计，因此在表3中主要展示绝对平移误差与绝对旋转误差这两项核心指标。因此为了保证分析的聚焦性与清晰度，这里仅展示绝对定位精度指标，以直接反映方法在重定位任务中的性能。总体来看，在KITTI数据集上，本文方法在城镇道路和开阔道路的定位误差分别为2.92 cm、0.67°和3.51 cm、0.71°，显著优于其他5种方法。在Waymo Open和nuScenes数据集上，本文方法同样取得了最佳的整体定位性能。

表3　在KITTI、Waymo Open和nuScenes数据集上的定位结果

Table 3　Localization Results on the KITTI, Waymo Open, and nuScenes Datasets

方法	指标	KITTI			Waymo Open			nuScene		
		城镇	开阔	平均值	城镇	开阔	平均值	城镇	开阔	平均值
AS[32]	平移MAE/cm	16.61	21.20	18.91	15.11	19.21	17.12	16.32	20.51	18.43
	旋转MAE/(°)	0.75	0.84	0.79	0.77	0.71	0.74	0.59	0.62	0.61
NG-RANSAC[33]	平移MAE/cm	13.93	16.44	15.24	10.52	11.61	11.12	8.10	10.40	9.251
	旋转MAE/(°)									
Regression-only	平移MAE/cm	10.92	13.43	12.22	12.84	15.42	14.09	9.73	12.22	10.92
	旋转MAE/(°)	0.74	0.79	0.77	0.62	0.72	0.67	0.53	0.56	0.55
DSAC++[34]	平移MAE/cm	8.80	10.24	9.51	7.81	11.30	9.62	6.52	9.94	8.22
	旋转MAE/(°)	0.67	0.72	0.70	0.53	0.63	0.58	0.42	0.48	0.45
3DGS-ReLoc[35]	平移MAE/cm	4.80	5.24	5.02	3.01	4.24	3.57	4.37	3.82	4.11
	旋转MAE/(°)	0.77	0.75	0.76	0.64	0.71	0.68	0.53	0.57	0.55
本文方法	平移MAE/cm	2.92	3.51	3.22	3.10	3.52	3.28	2.60	3.12	2.93
	旋转MAE/(°)	0.67	0.71	0.69	0.55	0.61	0.58	0.42	0.47	0.45

注：平均值表示两类场景的平均精度。

虽然在城镇道路场景中并非在所有数据集上均取得最佳结果，但在更具挑战性的开阔道路场景中仍保持了出色的定位精度。本文还统计了定位误差小于10 cm与1°的测试图像比例，该指标用于

综合比较不同方法的定位性能。由图11可见,本文方法在3种不同数据集上均表现最佳。

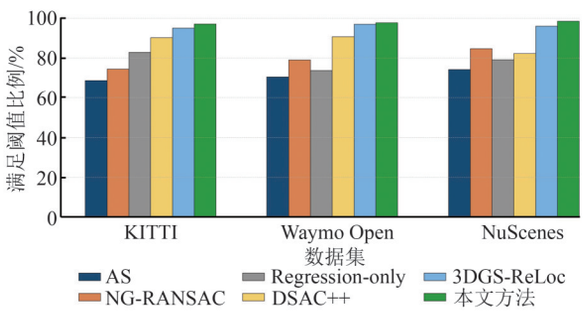


图11 重定位成功图像比例

Fig. 11 Success Rate of Relocalization Images

总体而言,本文方法在所有场景中均实现了最低的定位误差,并在不同环境下表现出良好的稳定性与适应性,表明该方法在各类复杂环境下具有显著优势。

4.4 消融实验

为验证所提系统中各功能模块对定位性能的影响,本文进行了消融实验。实验在KITTI数据集的城镇道路与开阔道路两类典型场景下进行,分别测试了以下4种配置:无动态遮挡处理、无SuperGlue特征匹配、无迭代优化、完整模型。表4统计了不同配置在两类场景下的平移与旋转的平均绝对误差(mean absolute error,MAE)及其平均值,以全面评估各模块对系统精度与稳定性的贡献。

表4 消融分析结果

Table 4 Ablation Analysis Results

配置方式	城镇		开阔		平均值	
	平移MAE/cm	旋转MAE/(°)	平移MAE/cm	旋转MAE/(°)	平移MAE/cm	旋转MAE/(°)
无动态遮挡处理	3.74	0.78	3.60	0.72	3.67	0.75
无SuperGlue特征匹配	3.85	0.82	4.02	0.79	3.94	0.81
无迭代优化	3.43	0.75	3.80	0.76	3.62	0.76
完整模型	2.92	0.67	3.51	0.71	3.22	0.69

由表4可见,完整模型在城镇道路与开阔道路两类场景中均取得最优结果,平移MAE和绝对MAE的平均值分别为3.22 cm和0.69°,验证了系统整体设计的有效性与稳定性。去除动态遮挡处理后,定位精度下降,尤其在城镇道路场景中,由于行人和车辆等动态目标的干扰,误差增幅较为显著。去除SuperGlue模块后,特征匹配的鲁棒性降低,在纹理较弱或光照变化显著的场景中更易出现误匹配,从而导致整体定位精度下降,说明深度特征匹配对于保持全局一致性至关重要。若不进行迭代优化,系统在初始位姿偏差较大时难以充分收敛,表现为平移与旋转误差同步增大。总体而言,各模块在系统性能中均发挥关键作用,其中迭代优化与深度特征匹配模块对精度提升贡献最为显著,充分验证了所提系统的合理性与有效性。

4.5 时间分析

对本文方法在重定位任务中各阶段的时间消耗进行分析。重定位过程主要包括以下步骤:(1)从检索数据库中查询最邻近的锚点图像;(2)利用P*n*P算法计算查询图像与最近锚点图像之间的相对位移;(3)基于该相对位移使用高斯模型渲染新的锚点图像;(4)迭代更新定位结果,直至满足终止条件。表5统计了各阶段的时间消耗,并报告了KITTI数据集中3类场景下的重定位时间。由表5可见,整个重定位过程可在2 s内完成,表明本文方法能够满足大多数自动驾驶场景下重定位任务的实时性要求,并具有实际应用潜力。在所有场景中,迭代渲染约在5次迭代后收敛。然而,随着场景规模增大,锚点数量增加,导致检索时间线性增加,重定位速度下降。通过设计合理的锚点选择策略,减小查询数据库规模,可以有效提升重定位算法的效率。

表5 时间统计分析结果

Table 5 Time Statistical Analysis Results

阶段	时间消耗/ms	执行次数/次		
		城镇道路	林荫道路	高速公路
SuperPoint	10	25	35	45
SuperGlue	30	25	35	45
P*n*P	2	5	5	5
渲染	6	5	5	5
总时间/s		1.04	1.44	1.84

5 结 语

本文基于多传感器数据的3DGS技术构建了室外大规模环境的地图,并引入SuperPoint与SuperGlue算法,在大视角偏差条件下实现了对稀疏基图库的高效检索与鲁棒特征点匹配。结合3DGS地图的渲染特性,设计了迭代逼近式重定位结构,从而获得高精度的位姿估计。在KITTI数据集的多个典型场景序列上对该方法进行了验证分析,结果表明,所提方法能够实现厘米级的定位精度,相较现有方法提升显著。

需要指出的是,由于KITTI数据集的多传感器数据是沿车辆行驶路径单向采集的,其构建的3DGS地图在可高保真渲染的视角范围上存在一定局限。

因此，当迭代过程中出现较大角度偏差时，渲染质量会明显下降，从而影响迭代收敛的精度[36-38]。未来可进一步探讨在有限数据采集条件下如何实现自由视角下的高保真3DGS渲染，并研究降低重定位迭代过程对角度误差敏感性的方法。

参 考 文 献

[1] 张小红，陶贤露，王颖喆，等. 城市场景智能手机GNSS/MEMS融合车载高精度定位[J]. 武汉大学学报(信息科学版)，2022，47(10)：1740-1749.
ZHANG Xiaohong，TAO Xianlu，WANG Yingzhe，et al. MEMS-Enhanced Smartphone GNSS High-Precision Positioning for Vehicular Navigation in Urban Conditions[J]. *Geomatics and Information Science of Wuhan University*，2022，47(10)：1740-1749.

[2] KUUTTI S，FALLAH S，KATSAROS K，et al. A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications[J]. *IEEE Internet of Things Journal*，2018，5(2)：829-846.

[3] LU Y Q，MA H J，SMART E，et al. Real-Time Performance-Focused Localization Techniques for Autonomous Vehicle：A Review[J]. *IEEE Transactions on Intelligent Transportation Systems*，2022，23(7)：6082-6100.

[4] 崔晓珍，周琪，武东杰，等. GNSS/IMU与里程计紧-松耦合的因子图融合定位方法[J]. 武汉大学学报(信息科学版)，2024，49(10)：1911-1921.
CUI Xiaozhen，ZHOU Qi，WU Dongjie，et al. Factor Graph Fusion Localization Method with Tight and Loose Coupling of GNSS/IMU and Odometry[J]. *Geomatics and Information Science of Wuhan University*，2024，49(10)：1911-1921.

[5] 李磊磊，梁琳，钟傲，等. 基于GPS的先验地图辅助视觉惯性定位方法[J]. 中国惯性技术学报，2023，31(1)：53-60.
LI Leilei，LIANG Lin，ZHONG Ao，et al. Visual-Inertial Positioning Method Based on GPS Prior Map Assisted[J]. *Journal of Chinese Inertial Technology*，2023，31(1)：53-60.

[6] BRACHMANN E，KRULL A，NOWOZIN S，et al. DSAC Differentiable RANSAC for Camera Localization[C]//IEEE Conference on Computer Vision and Pattern Recognition，Honolulu，HI，2017.

[7] LASKAR Z，MELEKHOV I，KALIA S，et al. Camera Relocalization by Computing Pairwise Relative Poses Using Convolutional Neural Network[C]//IEEE International Conference on Computer Vision Workshops，Venice，Italy，2018.

[8] SARLIN P E，UNAGAR A，LARSSON M，et al. Back to the Feature：Learning Robust Camera Localization from Pixels to Pose[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition，Nashville，TN，USA，2021.

[9] MILDENHALL B，SRINIVASAN P P，TANCIK M，et al. NeRF：Representing Scenes as Neural Radiance Fields for View Synthesis[C]//Computer Vision ECCV，Glasgow，UK，2020.

[10] YEN-CHEN L，FLORENCE P，BARRON J T，et al. INeRF：Inverting Neural Radiance Fields for Pose Estimation[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems，Prague，Czech Republic，2021.

[11] MAGGIOD，ABATEM，SHIJN，et al. Loc-NeRF：Monte CarloLocalizationUsingNeuralRadiance Fields[PP/OL]. arXiv(2022-09-19)[2025-08-20]. https://arxiv.org/abs/2209.09050.

[12] 赵文晔，高井祥，李增科，等. 地图匹配辅助的KF-PF室内定位算法模型[J]. 武汉大学学报(信息科学版)，2018，43(5)：806-812.
ZHAO Wenye，GAO Jingxiang，LI Zengke，et al. An Indoor Positioning System Based on Map-Aided KF-PF Module[J]. *Geomatics and Information Science of Wuhan University*，2018，43(5)：806-812.

[13] MATSUKI H，MURAI R，KELLY P H J，et al. Gaussian Splatting SLAM[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition，Seattle，WA，USA，2024.

[14] SUN Y，WANG X，ZHANG Y F，et al. IComMa：Inverting 3D Gaussian Splatting for Camera Pose Estimation via Comparing and Matching[PP/OL]. arXiv(2023-12-14)[2025-08-20]. https://arxiv.org/abs/2312.09031.

[15] ZHAI H J，ZHANG X Y，ZHAO B M，et al. SplatLoc：3D Gaussian Splatting-Based Visual Localization for Augmented Reality[J]. *IEEE Transactions on Visualization and Computer Graphics*，2025，31(5)：3591-3601.

[16] XIN Z，DAI C K，LI Y，et al. GauLoc：3D Gaussian Splatting-Based Camera Relocalization[J]. *Computer Graphics Forum*，2024，43(7)：e15256.

[17] BOTASHEV K，PYATOV V，FERRER G，et al. GSLoc：Visual Localization with 3D Gaussian Splatting[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems，Abu Dhabi，United Arab Emirates，2024.

[18] ZELLER A J，WU H J. GSplatLoc：Ultra-Precise Camera Localization *via* 3D Gaussian Splatting[PP/

OL]. arXiv(2024-12-28)[2025-08-20]. https://arxiv.org/abs/2412.20056.

[19] CHEN Z Y, YANG J W, HUANG J H, et al. OmniRe: Omni Urban Scene Reconstruction[PP/OL]. arXiv(2024-08-29)[2025-08-20]. https://arxiv.org/abs/2408.16760.

[20] CHEN Y R, GU C, JIANG J Z, et al. Periodic Vibration Gaussian: Dynamic Urban Scene Reconstruction and Real-Time Rendering[PP/OL]. arXiv(2023-11-30)[2025-08-20]. https://arxiv.org/abs/2311.18561.

[21] YANG J W, IVANOVIC B, LITANY O, et al. EmerNeRF: Emergent Spatial-Temporal Scene Decomposition via Self-Supervision[PP/OL]. arXiv(2023-11-03)[2025-08-20]. https://arxiv.org/abs/2311.02077.

[22] 王铉彬, 李星星, 廖健驰, 等. 基于图优化的紧耦合双目视觉/惯性/激光雷达SLAM方法[J]. 测绘学报, 2022, 51(8): 1744-1756.
WANG Xuanbin, LI Xingxing, LIAO Jianchi, et al. Tightly-Coupled Stereo Visual-Inertial-LiDAR SLAM Based on Graph Optimization[J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(8): 1744-1756.

[23] NOBLE F K. Comparison of OpenCV's Feature Detectors and Feature Matchers[C]//The 23rd International Conference on Mechatronics and Machine Vision in Practice, Nanjing, China, 2017.

[24] DETONE D, MALISIEWICZ T, RABINOVICH A. SuperPoint: Self-Supervised Interest Point Detection and Description[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 2018.

[25] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: Learning Feature Matching with Graph Neural Networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020.

[26] SHI Y, CAI J X, SHAVIT Y, et al. ClusterGNN: Cluster-Based Coarse-to-Fine Graph Neural Network for Efficient Feature Matching[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 2022.

[27] MA J Y, JIANG X Y, FAN A X, et al. Image Matching from Handcrafted to Deep Features: A Survey[J]. *International Journal of Computer Vision*, 2021, 129(1): 23-79.

[28] WANG A, PRUKSACHATKUN Y, NANGIA N, et al. SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems[PP/OL]. [2025-08-20]. https://arxiv.org/abs/1905.00537.

[29] 叶沅鑫, 单杰, 熊金鑫, 等. 一种结合SIFT和边缘信息的多源遥感影像匹配方法[J]. 武汉大学学报(信息科学版), 2013, 38(10): 1148-1151.
YE Yuanxin, SHAN Jie, XIONG Jinxin, et al. A Matching Method Combining SIFT and Edge Information for Multi-Source Remote Sensing Images[J]. *Geomatics and Information Science of Wuhan University*, 2013, 38(10): 1148-1151.

[30] BESL P J, MCKAY N D. A Method for Registration of 3-D Shapes[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(2): 239-256.

[31] WANG J Y, RUPPRECHT C, NOVOTNY D. PoseDiffusion: Solving Pose Estimation via Diffusion-Aided Bundle Adjustment[C]//IEEE/CVF International Conference on Computer Vision, Paris, France, 2024.

[32] SATTLER T, LEIBE B, KOBBELT L. Efficient & Effective Prioritized Matching for Large-Scale Image-Based Localization[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(9): 1744-1756.

[33] BRACHMANN E, ROTHER C. Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses[C]//IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019.

[34] BRACHMANN E, ROTHER C. Learning Less Is More: 6D Camera Localization via 3D Surface Regression[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.

[35] JIANG P, BENDAPUDI N, SARIPALLI S, et al. 3DGS-ReLoc: 3D Gaussian Splatting for Map Representation and Visual Localization[J]. *Journal of Autonomous Vehicles and Systems*, 2025, 5(3): 031004.

[36] ZHANG S J, LIU Z A, XU B, et al. Fusion GNSS/INS/Vision with Path Planning Prior for High Precision Navigation in Complex Environment[J]. *IEEE Sensors Journal*, 2025, 25(5): 9045-9055.

[37] XU B, ZHANG S J, KUANG K F, et al. A Unified Cycle-Slip, Multipath Estimation, Detection and Mitigation Method for VIO-Aided PPP in Urban Environments[J]. *GPS Solutions*, 2023, 27(2): 59.

[38] WANG J R, LIU J N, ZHANG S J, et al. Sky-View Images Aided NLOS Detection and Suppression for Tightly Coupled GNSS/INS System in Urban Canyon Areas[J]. *Measurement Science and Technology*, 2024, 35(2): 025112.