

# 基于点线特征的单目同步定位与建图初始化方法

钱琦<sup>1</sup>, 白庭铭<sup>1</sup>, 闭云峰<sup>1</sup>, 乔程昱<sup>1</sup>, 项志宇<sup>1,2\*</sup>

<sup>1</sup>浙江大学信息与电子工程学院, 浙江 杭州 310027;

<sup>2</sup>浙江省信息处理与通信网络重点实验室, 浙江 杭州 310027

**摘要** 针对目前单目同步定位与建图(SLAM)初始化方法鲁棒性差、精度低的问题,提出了一种基于点线特征的单目 SLAM 初始化方法。首先,在两帧图像之间提取线特征,并进行特征匹配。然后,通过最大化投影线特征的重叠长度优化两帧图像之间的初始旋转和平移矩阵。最后,用滑动窗口增加初始化的图像帧数,利用多帧图像的信息和约束关系以及全局光束平差法优化初始地图和估计的相机关键帧位姿。在 TUM 和 OpenLORIS 数据集上的测试结果表明,相比传统初始化方法,本方法的鲁棒性更好、精度更高,能在具有挑战性的场景中快速完成高精度初始化。

**关键词** 图像处理; 运动估计和光流; 多帧图像处理; 单目视觉初始化; 定位与建图; 点线特征

中图分类号 TP391

文献标志码 A

doi: 10.3788/AOS202141.1215002

## Monocular Simultaneous Localization and Mapping Initialization Method Based on Point and Line Features

Qian Qi<sup>1</sup>, Bai Tingming<sup>1</sup>, Bi Yunfeng<sup>1</sup>, Qiao Chengyu<sup>1</sup>, Xiang Zhiyu<sup>1,2\*</sup>

<sup>1</sup> College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China;

<sup>2</sup> Zhejiang Provincial Key Laboratory of Information Processing Communication and Networking, Hangzhou, Zhejiang 310027, China

**Abstract** Aiming at the problem of poor robustness and accuracy in the initialization of monocular visual simultaneous localization and mapping (SLAM), this paper proposes a robust initialization method based on point and line features. First, the line features are extracted and matched between two frames. Then, the initial rotation matrix and translation matrix between the two frames are optimized by maximizing the overlap length of the projected line features. Finally, a sliding window is used to increase the number of initial image frames, and the initial map and the estimated keyframe pose are optimized based on information and constraint of multi-frame images and the global bundle adjustment method. The test results on the TUM and OpenLORIS datasets show that compared with traditional initialization methods, the method is more robust and accurate, and can quickly complete high-precision initialization in challenging scenarios.

**Key words** image processing; motion estimation and optical flow; multiframe image processing; monocular visual initialization; localization and mapping; point and line features

**OCIS codes** 100.2000; 110.4153; 110.4155

### 1 引言

视觉同步定位与建图(SLAM)是计算机视觉领

域的重要研究方向,广泛应用于无人驾驶、机器人以及增强现实(AR)/虚拟现实(VR)等领域。单目 SLAM 的成本低、应用场景广泛,但存在尺度不确

收稿日期: 2020-12-01; 修回日期: 2021-01-11; 录用日期: 2021-01-22

基金项目: 国家自然科学基金浙江两化融合联合基金(U1709214)

\* E-mail: xiangzy@zju.edu.cn

定以及漂移误差累积等缺点。此外,相比多目 SLAM,单目 SLAM 的初始化难度更大,容易发生初始化失败或初始化精度不高的情况。初始化是单目 SLAM 中的重要步骤,主要用于设定初始尺度和创建初始地图,而 SLAM 系统的性能与初始化的精度及鲁棒性息息相关。早期的视觉 SLAM 系统都是基于扩展的卡尔曼滤波进行初始化,初始化三维地图点时的不确定性较大,需要在后续输入帧中持续更新地图点<sup>[1-3]</sup>。基于优化的方法则是从增量运动恢复结构(SFM)中恢复位姿并完成 SLAM 的初始化<sup>[4-5]</sup>。直接法 SLAM 系统初始化时基于灰度不变假设,通过最小化光度误差计算初始两帧图像的位姿,如大尺度直接法单目 SLAM (LSD-SLAM)<sup>[6]</sup>,但该方法在复杂光照情况下不稳定且图像序列需要经过光度校准<sup>[7-8]</sup>。非直接法 SLAM 初始化使用序列的前两帧图像进行初始化,在两帧图像之间提取特征并进行匹配。一些视觉里程计算法在初始化时假设观察到的是一个平面场景,然后通过单应性矩阵初始化位姿<sup>[9]</sup>;另一些算法则使用本质矩阵计算初始位姿,通用性较强。如 ORB (Oriented FAST and rotated brief)-SLAM<sup>[10]</sup> 系统用本质矩阵  $E$  和单应性矩阵  $H$  混合的方法计算位姿,但只有模型的不确定性很低时才能重建成功。该方法对图像帧的要求较高,如果找不到两帧能够有效重建地图的三维点,SLAM 系统就无法运行,且 SLAM 系统初始化的时间较长。

Tang 等<sup>[11]</sup>提出了用秩 1 矩阵分解法进行初始化的 G-SLAM 系统,该系统计算的初始地图不涉及模型选择问题,且能同时优化相机位姿和地图点,保证了系统在线性时间内能够快速收敛,但能被秩 1 矩阵分解法分解的特征必须被所有帧观察到。可用 KLT (Kanade-Lucas-Tomasi) 光流法进行特征跟踪,但算法的性能表现非常依赖 KLT 光流法的跟踪精度。Lee 等<sup>[12-14]</sup>借鉴 G-SLAM 的思想,基于极线约束实现了所有特征在全部帧中的跟踪,并提出了有效计算几何插值基本矩阵的方法。虽然该方法能在全部帧中得到所有特征的像素坐标,但插值法得到的坐标不能提供额外的信息,即对于位姿约束不会有额外贡献。光束平差(BA)法可对关键帧进行优化,通常应用于 SLAM 系统的后端优化模块中。对地图进行全局 BA 优化时通常使用位姿图,将优化问题转换成非线性最小二乘问题,从而用  $g^2o$ (General graph optimization)<sup>[15]</sup> 或 Ceres<sup>[16]</sup> 库迭代求解最小化线性目标函数。目前,基于点线特

征的 SLAM 系统在初始化时只使用点特征。但在一些弱纹理结构化的场景中,线特征的数量较多,如果只用点特征计算本质矩阵和单应性矩阵,会导致匹配特征的数量过少、初始化失败。Zhang<sup>[17]</sup>基于两个视角观察到的线匹配对估计两帧图像之间的相对位姿,在点特征较少的弱纹理环境中大大增加了初始化方法的鲁棒性。

本文提出了一种基于点线特征的单目 SLAM 多帧图像初始化方法,能在低视差、结构化弱纹理和近似纯旋转等场景中完成初始化。通过一个滑动窗口,用多帧图像的信息完成初始化,并在初始化时用线特征优化旋转矩阵  $R$  和平移矩阵  $t$ 。为了使构建的初始地图更精准,在滑动窗口中用 BA 法优化位姿以及地图的点、线坐标。在 TUM 和 OpenLORIS 数据集上的测试结果表明,相比 ORB-SLAM 和 LSD-SLAM 系统的初始化方法,基于点线特征的多帧初始化方法能提高 SLAM 系统初始化的鲁棒性及精度。

## 2 基本原理

### 2.1 基于线特征优化帧间的初始位姿

假设两帧图像中有一对匹配的线段,由点的对极几何可知,三维空间中的点  $P$  在第一帧图像中的像素点  $p_1$  在第二帧图像中对应一条极线  $l_2$ ,如图 1 所示,其中  $C_1$  和  $C_2$  为两个相机的光心。受线段检测误差和遮挡等因素的影响,两帧图像中检测到的匹配线段长度及端点往往无法一一对应。但对于初始化中间隔较小的两帧图像,由于视角变化不大,匹配线段的重叠部分较多。对于第一帧图像中一条线段的两个端点  $s_1$  和  $e_1$ ,在第二帧图像中会对应两条极线  $l'_s$  和  $l'_e$ 。由于端点之间没有限制一一对应,因此  $l'_s$  和  $l'_e$  不一定都会过第二帧图像中检测出的线段端点,如图 2 所示。

空间中一条直线在图像中的像也是一条直

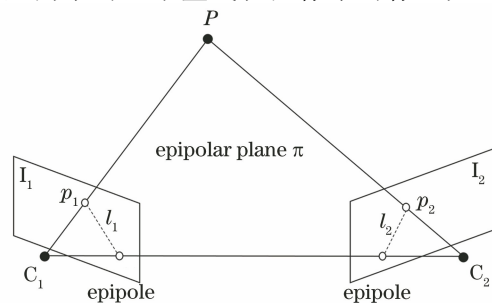


图 1 点的对极几何

Fig. 1 Epipolar geometry of a point

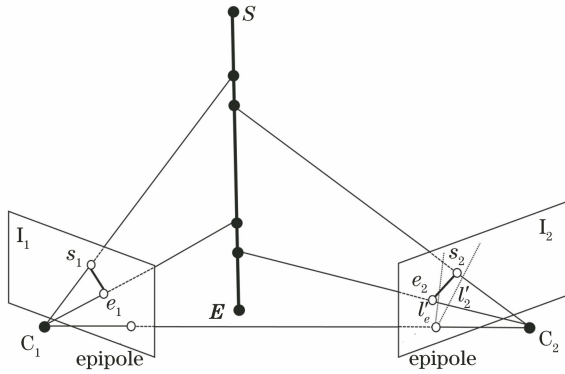


图 2 线段的对极几何约束

Fig. 2 Epipolar geometry of a line segment

线,虽然在第二帧图像中没有检测到  $s_1$  和  $e_1$  对应的端点  $s'_1$  和  $e'_1$ ,但  $s'_1$  和  $e'_1$  肯定在以  $s_2$  和  $e_2$  为端点连成的直线  $l_{s_2e_2}$  上,如图 3 所示。可以发现,第一帧图像中检测到的线段在第二帧图像上的投影通过两条极线和第二帧图像中的线段  $l_{s_2e_2}$  相交,得到  $s'_1$  和  $e'_1$ 。由于初始化时两帧图像之间的旋转角较小,因此对两帧图像检测到的线段起点和终点通过约束进行区分,可表示为

$$\| (e'_1 - s'_1) \cdot (e_2 - s_2) \| > 0, \quad (1)$$

式中,  $s'_1$  和  $e'_1$  为第一帧图像中检测线段在第二帧图像中投影线段的起点和终点像素坐标,  $s_2$  和  $e_2$  为第二帧图像中检测线段的起点和终点像素坐标。

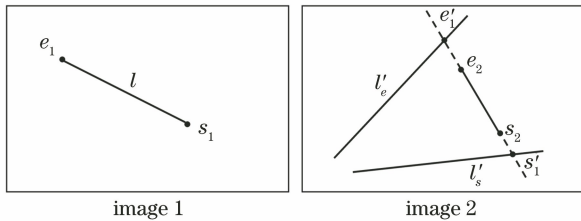


图 3 匹配线段特征的重叠长度

Fig. 3 Overlap length of the matched line segment feature

满足约束条件时,第二帧图像中的线段  $l_{s_2e_2}$  和  $l_{s'_1e'_1}$  有 4 种情况,如图 4 所示。将两条线段的重叠长度记为  $\mathcal{L}'$ ,当两帧图像之间的  $\mathbf{R}$  和  $\mathbf{t}$  准确且两帧图像之间的视角变化不大时,  $s_1$  和  $s'_1$  重合,  $e_1$  和  $e'_1$  重

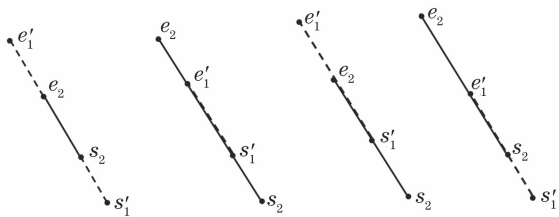


图 4 线段重叠的 4 种情况

Fig. 4 Four cases of overlapping line segments

合,即线段的重叠长度应该尽可能大,  $\mathcal{L}'$  可定义为

$$\mathcal{L}' = \min(\| e'_1 - s'_1 \|, \| e_2 - s_2 \|, \| e_2 - s'_1 \|, \| e'_1 - s_2 \|). \quad (2)$$

若第二帧图像中的线段  $l_{s_2e_2}$  和  $l_{s'_1e'_1}$  没有重叠,则  $\mathcal{L}'$  为负值,可表示为

$$\mathcal{L}' = -\min(\| e'_1 - s_2 \|, \| e_2 - s'_1 \|). \quad (3)$$

根据图像的对称性原理,可通过第二帧图像中检测的线段端点计算其在第一帧图像中的投影,得到  $\mathcal{L}$ 。由于长线段和短线段的重叠长度在约束两帧图像之间的位姿时应具有相同的权重,因此对  $\mathcal{L}$  和  $\mathcal{L}'$  进行归一化,分别除以检测线段长度  $l$  和  $l'$ ,得到优化重叠长度和检测线段长度的比值,对于图 4 中有重叠的情况,该比值在 0 到 1 之间。文献[17]将用两帧图像匹配的线特征优化运动位姿问题描述为给定  $n$  对匹配的线特征  $\{(l_i, l'_i) | i=1, \dots, n\}$ ,通过最小化  $\mathcal{F}$  估计相机的运动参数  $(\mathbf{R}, \mathbf{t})$ ,  $\mathcal{F}$  可表示为

$$\mathcal{F} = \sum_{i=1}^n \left[ \left(1 - \frac{\mathcal{L}_i}{l_i}\right)^2 + \left(1 - \frac{\mathcal{L}'_i}{l'_i}\right)^2 \right]. \quad (4)$$

文献[17]中用下降单纯形法<sup>[18]</sup>解决该优化问题,而在 SLAM 中,由于初始化时两帧图像之间根据点特征可计算出本质矩阵  $\mathbf{E}$  和单应性矩阵  $\mathbf{H}$ 。因此,将  $\mathbf{E}$  或  $\mathbf{H}$  作为待优化变量,利用 Ceres 库<sup>[16]</sup>最小化函数  $\mathcal{F}$ ,然后将优化后的  $\mathbf{E}$  或  $\mathbf{H}$  分解为  $(\mathbf{R}, \mathbf{t})$ 。

## 2.2 多帧图像的初始化

针对传统初始化方法在部分场景(如近似纯旋转、移动物体干扰较多、小视差场景)下求解误差大的问题,提出了一种基于 BA 法优化的多帧图像初始化方法。用滑动窗口将用于构建初始地图的关键帧数量增加至  $N$  帧。当输入新图像后,将其与滑动窗口内的第一帧(参考帧)图像进行特征匹配,若匹配的特征数量超过设定的阈值  $\sigma$ ,则将其加入滑动窗口,并与其他所有帧进行特征匹配。每一帧与第一帧之间都通过点对计算本质矩阵和单应性矩阵,并用 2.1 节中的线特征优化方法得到  $(\mathbf{R}, \mathbf{t})$ 。然后,重构地图点与地图线的坐标,记录所有关键帧和特征信息,作为初始地图。若输入帧与第一帧图像的匹配特征数量少于  $\sigma$ ,则舍弃该帧图像。若连续六帧图像都被舍弃,则表明目前滑动窗口内关键帧包含的信息不够多,舍弃所有关键帧,重新初始化滑动窗口。当滑动窗口内的有效帧数为  $N$  时,对整个滑动窗口内的关键帧和地图点进行位姿图优化。

以  $N=3$  为例,首先,在每两帧图像之间用 ORB 特征点进行匹配,并用五点法或八点法粗略计算出帧间的本质矩阵和单应性矩阵。然后,基于

LSD<sup>[19]</sup> 提取线特征和线段二进制描述符 (LBD)<sup>[20]</sup> 进行线特征匹配, 并用线特征优化位姿。最后, 重构出特征点和特征线的三维坐标, 并将所有三维坐标在以第一帧相机光心为原点的坐标系下进行表示。根据关键帧和三维地图点、线之间的共视关系, 将地图点和地图线分为四类, 如图 5 所示。其中,  $F_0$  能被第一帧和第二帧图像观察到, 可用于优化第二帧

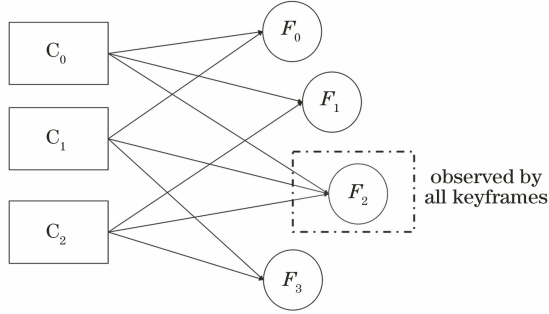


图 5 关键帧和地图点、线的共视关系

Fig. 5 Common view relationship of keyframe, map points and lines

图像的位姿;  $F_1$  能被第一帧和第三帧图像观察到, 可用于优化第三帧图像的位姿;  $F_2$  能被三帧图像观察到, 可用于优化三帧图像之间的位姿, 提供的信息最多;  $F_3$  能被第二帧和第三帧图像观察到, 可用于优化第二帧和第三帧图像的相对位姿。

所有关键帧的位姿和三维点坐标、线坐标都在第一帧图像坐标系下 (上标 (0) 表示以第一帧相机的光心作为原点), 对所有关键帧位姿和地图点坐标进行全局 BA 优化, 优化的状态向量可表示为

$$\mathbf{x} = [\xi_i^{(0)}, \mathbf{P}_j^{(0)}, \mathbf{O}_k^{(0)}], \quad i = 1, \dots, N - 1, \\ j = 0, \dots, M_p - 1, \quad k = 0, \dots, M_l - 1, \quad (5)$$

式中,  $\xi_i^{(0)}$  为关键帧的位姿,  $\mathbf{P}_j^{(0)}$  为地图点的三维坐标,  $\mathbf{O}_k^{(0)}$  为线特征的正交表示, 有四个自由度,  $N$  为滑动窗口内关键帧的数量,  $M_p$  为地图点数量,  $M_l$  为地图线的数量。误差函数为地图点与地图线投影到所有关键帧图像上的重投影误差, 重投影误差示意图如图 6 所示。其中,  $p_1$  和  $p_2$  分别为图像中检测出的点特征,  $p$  和  $q$  为检测线特征的两个端点,  $p'$  和  $q'$  为投影线特征的两个端点,  $d$  为点到直线的距离。

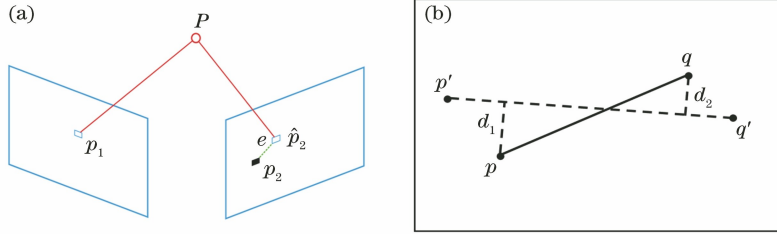


图 6 重投影误差示意图。(a) 点特征; (b) 线特征

Fig. 6 Schematic diagram of the reprojection error. (a) Point feature; (b) line feature

对于某个地图点坐标  $\mathbf{P}_j = [X_j, Y_j, Z_j]^T$ , 其在某一关键帧的投影像素坐标  $\tilde{\mathbf{u}}_j = [u_j, v_j]^T$ , 像素位置与空间点位置关系可表示为

$$s_j \begin{bmatrix} u_j \\ v_j \\ 1 \end{bmatrix} = \mathbf{K} \exp(\xi^\wedge) \begin{bmatrix} X_j \\ Y_j \\ Z_j \\ 1 \end{bmatrix}, \quad (6)$$

式中,  $s_j$  为成像平面的深度,  $\mathbf{K}$  为相机的内参矩阵,  $\xi^\wedge$  为地图点坐标所在坐标系到该帧相机坐标系转换矩阵的李代数。可以发现, 点特征的重投影误差  $\mathbf{e}_p = \mathbf{u}_j - \tilde{\mathbf{u}}_j$ ,  $\mathbf{u}_j$  为该帧图像中与  $\mathbf{P}_j$  对应特征点的像素坐标。线特征的重投影误差为检测线段端点到投影直线的距离, 可表示为

$$\mathbf{e}_l = \left[ \frac{\mathbf{p}_1^T \mathbf{l}'}{\sqrt{l_1^2 + l_2^2}}, \frac{\mathbf{q}_1^T \mathbf{l}'}{\sqrt{l_1^2 + l_2^2}} \right]^T, \quad (7)$$

式中,  $\mathbf{p}_l, \mathbf{q}_l$  为检测线段两个端点的像素坐标,  $\mathbf{l}'$  为

投影直线的法向量,  $l_1, l_2$  为法向量  $\mathbf{l}'$  的第一、第二参数。则整体优化方程可表示为

$$\{\xi^*, \mathbf{P}^*, \mathbf{O}^*\} = \underset{\xi_i^{(0)}, \mathbf{P}_j^{(0)}, \mathbf{O}_k^{(0)}}{\operatorname{argmin}} \left[ \sum_{i,j} \rho(\mathbf{e}_{p_{ij}}^T \boldsymbol{\Omega}_{ij}^{-1} \mathbf{e}_{p_{ij}}) + \sum_{i,k} \rho(\mathbf{e}_{l_{ik}}^T \boldsymbol{\Omega}_{ik}^{-1} \mathbf{e}_{l_{ik}}) \right], \quad (8)$$

式中,  $\mathbf{e}_{p_{ij}}$  为第  $j$  个地图点在第  $i$  个关键帧图像上的重投影误差,  $\mathbf{e}_{l_{ik}}$  为第  $k$  条地图线在第  $i$  个关键帧图像上的重投影误差,  $\mathbf{x}^* = [\xi^*, \mathbf{P}^*, \mathbf{O}^*]$  为优化后最终得到的状态向量,  $\boldsymbol{\Omega}$  为信息矩阵,  $\rho$  为鲁棒核函数。全局优化后可得到精确的初始局部地图, 包括地图中关键帧的位姿及地图点、地图线坐标, 为建图及定位提供了精确的数据。

### 3 实验结果及分析

#### 3.1 实验数据集

TUM 数据集<sup>[21]</sup> 包含 RGB (Red, Green, Blue)

深度(RGBD)图像、单目图像、视觉惯性(VI)图像等不同类型的数据集,实验使用的是 RGBD 图像数据集。TUM-RGBD 数据集主要包括室内场景图像,既有纹理丰富的办公室场景,也有弱纹理或结构化信息较少的场景,有些序列中还存在大量的动态物体,因此在 TUM-RGBD 数据集上进行测试对 SLAM 系统的鲁棒性提出了很高的要求。

OpenLORIS 数据集<sup>[22]</sup>中的图像全部采集自真实生活环境,分为物体(Object)数据和场景(Scene)数据,可用于物体识别和 SLAM 系统的评估。OpenLORIS-Scene 数据集全部由真实机器人录制,包含办公室、走廊、家庭、咖啡厅和超市五类场景,彩色图像、深度图像、双目鱼眼图像、惯导(IMU)和里程计五类数据以及由动作捕捉系统或高精度激光雷达提供的机器人轨迹真值。相比其他 SLAM 数据集,OpenLORIS 数据集最大的特点在于场景重入,即同一场景会有多段数据,录制间隔从半天到几周不等,包含了日常生活导致的场景变动。

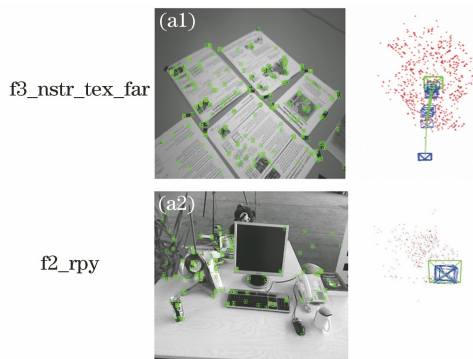
### 3.2 基于线特征的初始位姿估计结果

为了验证用两帧图像匹配的线特征优化初始位姿的效果,在 TUM 数据集上进行测试,测试环境为 Intel Core i5-8600K CPU@3.60 GHz×6。以单目 ORB-SLAM 为基础,在其初始化模块加入线特征优化方法,并将其和原始单目 ORB-SLAM 进行对比。实验中用 LSD 提取 200 根线特征,用 ORB 提取 300 个点特征,且线特征只用在初始化模块中。表 1 为在 TUM 数据集的三个序列中加入线特征(Line

表 1 线特征优化前后系统的定位误差

Table 1 Positioning of the system before and after line feature optimization unit: cm

TUM sequence	ORB-SLAM	ORB-SLAM+Line init
f2_xyz	1.61	1.24
f2_desk	1.54	1.48
f2_desk_person	1.22	1.16



init)优化前后系统的定位精度,其中,定位精度是整个序列的绝对平移误差(ATE),多次测量得到整个序列的均方根误差(RMSE)。结果表明,用线特征优化初始化模块能有效提高 SLAM 的初始化精度,改善整个系统的定位精度。

### 3.3 基于点特征的初始化结果

为了验证多帧初始化方法的有效性,将多帧初始化方法加入 ORB-SLAM 系统中,并与原始 ORB-SLAM 系统和 LSD-SLAM 系统在 TUM-RGBD 数据集上进行了测试实验,实验中只采用了点特征,目的是探究基于滑动窗口的多帧初始化方法对系统初始化性能的影响。ORB-SLAM 系统采用传统初始化方法,即通过两帧图像的信息计算本质矩阵或单应性矩阵,进而恢复 $(R, t)$ ;LSD-SLAM 系统利用直接法随机初始化地图。表 2 为不同系统在 9 个序列上的测试结果,将每个序列运行 25 次后取平均值。其中,ambiguity 表示全程无法初始化成功。可以发现,多帧初始化方法能提升 SLAM 系统初始化的鲁棒性和精度,在传统初始化方法无法完成初始化的场景中也能顺利进行初始化。

图 7 为本系统和 ORB-SLAM 系统在 f2\_rpy、

表 2 不同方法在 TUM 数据集上的 ATE

Table 2 ATE of different algorithms on

TUM-RGBD sequence	TUM-RGBD sequences		unit: cm
	LSD-SLAM	ORB-SLAM	Ours
f1_xyz	9.04	1.38	0.96
f2_xyz	2.21	0.98	0.93
f2_rpy	ambiguity	ambiguity	1.48
f2_desk_person	4.60	0.89	1.12
f3_long_office	38.12	1.42	1.38
f3_nstr_tex_near	6.50	1.34	1.83
f3_nstr_tex_far	18.32	ambiguity	3.65
f3_sitting	7.71	2.76	2.11
f3_walking	15.20	ambiguity	2.84



图 7 两种方法在 TUM 数据集上的初始化结果。(a)本方法;(b)传统方法

Fig. 7 Initialization results of two methods on the TUM dataset. (a) Our method; (b) traditional method

f3\_nstr\_tex\_far 序列上的初始化结果,其中,传统初始化方法没有初始化成功,但本方法初始化成功并创建了局部地图,这表明使用多帧图像的信息可以提高初始化的鲁棒性和精度。图 8 为滑动窗口大小  $N$  对初始化方法的影响,结果表明, $N > 3$  时,初始化精度基本趋于稳定,但收敛时间会增加,因此实验选择  $N = 3$ 。

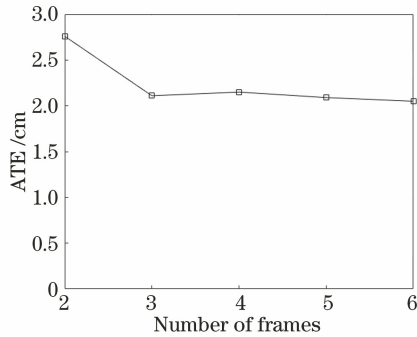


图 8 系统的 ATE 随  $N$  的变化曲线

Fig. 8 Variation curve of ATE of the system with  $N$

### 3.4 融合点线特征的多帧图像初始化结果

将基于点线特征的多帧初始化 SLAM 系统和传统 PL-SLAM (Point and line SLAM)、ORB-SLAM 系统进行对比,其中,PL-SLAM 系统为在 ORB-SLAM 系统上加入线特征,且线特征只用于跟踪及优化部分的 SLAM 系统(与初始化过程无关),也是目前常见的一种 PL-SLAM 框架。在 TUM<sup>[21]</sup> 和 OpenLORIS 数据集<sup>[22]</sup> 上的测试结果如表 3 和表 4 所示,可以发现,由于 TUM 数据集中有些序列的初始化难度较高,ORB-SLAM 和 PL-SLAM 系统全程都无法完成初始化。在 OpenLORIS 数据集中,图像的尺度较大,且 office 系列的初始化难度较低,本方法对 SLAM 系统初始化精度的提升不明显;而 cafe 序列初始化时视差小,本方法的提升效果更显著。corridor 序列的初

表 3 不同方法在 TUM 数据集上的 ATE

Table 3 ATE of different methods on the TUM dataset  
unit: cm

TUM-RGBD	PL-SLAM	ORB-SLAM	Ours
f1_xyz	3.02	1.38	1.00
f2_xyz	1.13	0.98	0.97
f2_rpy	ambiguity	ambiguity	1.48
f2_desk_person	1.63	0.89	1.34
f3_nstr_tex_far	ambiguity	ambiguity	3.65
f3_sitting	3.00	2.76	2.65
f3_walking	ambiguity	ambiguity	2.45

表 4 不同方法在 OpenLORIS 数据集上的 ATE

Table 4 ATE of different methods on the OpenLORIS dataset  
unit: cm

OpenLORIS sequence	PL-SLAM	ORB-SLAM	Ours
office1	0.061	0.059	0.053
office2	0.146	0.154	0.137
office3	0.041	0.040	0.031
office4	0.141	0.137	0.134
office5	ambiguity	ambiguity	0.035
office6	0.063	0.058	0.061
office7	0.043	0.043	0.041
cafe1	0.163	0.167	0.120
cafe2	0.255	0.267	0.128
corridor1-1-part1	0.654	0.671	0.231
corridor1-1-part2	0.633	0.624	0.280
corridor1-1-part3	0.267	0.290	0.223
corridor1-1-part4	0.403	0.415	0.121

始化难度大,图像中的特征少,传统初始化方法极易初始化失败或精度差,而本方法能明显提升 SLAM 系统初始化的鲁棒性和精度。

图 9 为本方法和 ORB-SLAM 系统在 cafe2 序列上的轨迹图,其中,图 9(a1)、图 9(b1)为本方法、ORB-SLAM 系统与真值轨迹的对比图,图 9(a2)、图 9(b2)为两种方法在 cafe2 序列整体轨迹中一小段轨迹的局部放大图。可以发现,本方法生成的轨迹更接近真值。图 10 为本方法和 ORB-SLAM 系统在 corridor1-1 序列中的两个初始化场景,可以发现,传统初始化方法建立的初始化地图存在漂移现象,且传统方法的初始关键帧位姿误差较大。而本方法能精确建立一张初始化地图,对后续的定位和跟踪起到关键作用。

## 4 结 论

提出了一种基于点线特征的多帧初始化方法,用一个滑动窗口选择多个关键帧进行初始化,并在 BA 法优化前用两帧图像之间的线特征优化初始位姿。对于两帧图像中的匹配线特征,用最大化重叠长度的方式提高两帧图像之间位姿初值的估计精度。在初始地图上进行全局 BA 优化时,通过最小化点特征和线特征的重投影误差对初始地图的位姿和点线特征三维坐标进行优化。在 TUM 数据集和 OpenLORIS 数据集上的测试结果表明,相比传统非直接法初始化方法以及直接法初始化方法,本方

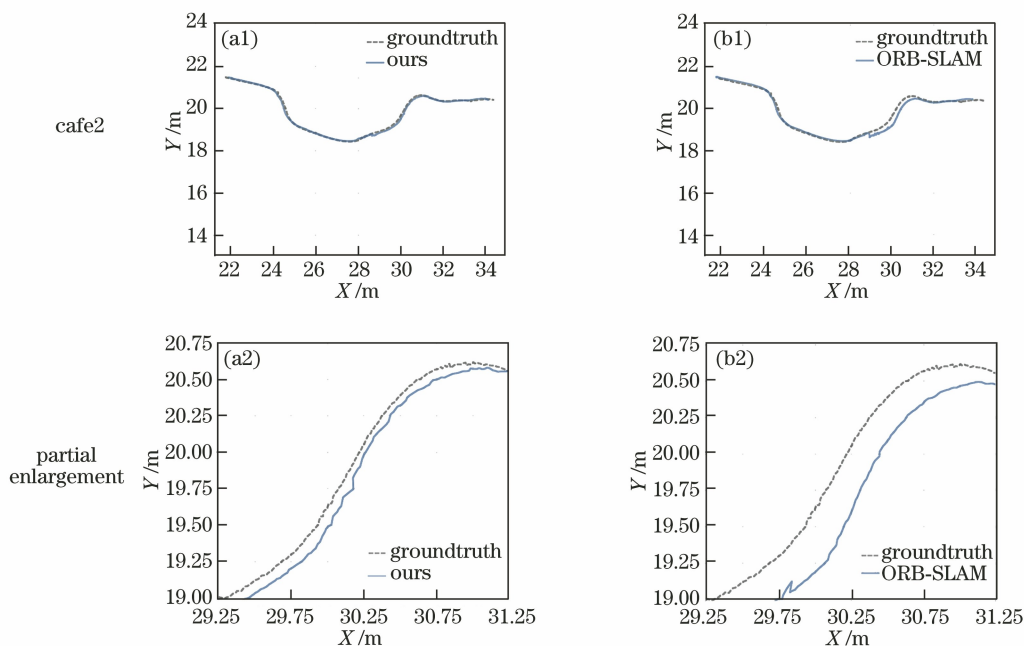


图 9 两种方法在 cafe2 序列上的轨迹图。(a)本方法;(b)传统方法

Fig. 9 Trajectories of the two methods on the cafe2 sequence. (a) Our method; (b) traditional method

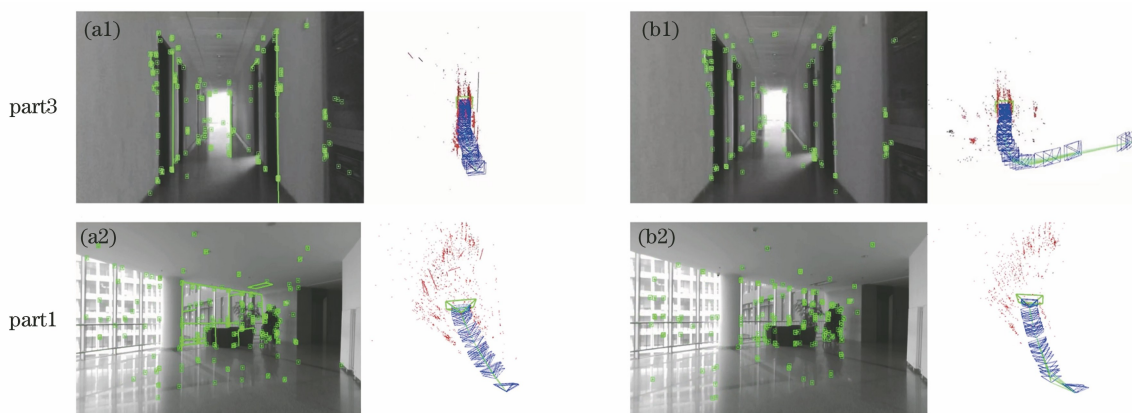


图 10 两种方法在 corridor1-1 序列上的初始化结果。(a)本方法;(b)传统方法

Fig. 10 Initialization results of the two methods on the corridor1-1 sequence. (a) Our method; (b) traditional method

法的效果都更好。这证明了利用多帧图像的信息以及线特征优化初始位姿,可提高 SLAM 系统在初始化过程中的精度和鲁棒性,进而提高后续跟踪、建图的精度,且本方法对 SLAM 系统的实时性没有影响。后续还将针对滑动窗口内关键帧的选择算法进行优化,将基于点线特征的多帧初始化方法应用到其他 SLAM 框架中。

参 考 文 献

[1] Chiuso A, Favaro P, Jin H L, et al. Structure from motion causally integrated over time [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 523-535.  
 [2] Davison A J, Reid I D, Molton N D, et al.

MonoSLAM: real-time single camera SLAM [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.  
 [3] Civera J, Davison A J, Montiel J M M. Inverse depth parametrization for monocular SLAM [J]. IEEE Transactions on Robotics, 2008, 24(5): 932-945.  
 [4] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces [C] // 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, November 13-16, 2007, Nara, Japan. New York: IEEE Press, 2007: 225-234.  
 [5] Royer E, Bom J, Dhome M, et al. Outdoor autonomous navigation using monocular vision [C] // 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, August 2-6, 2005,

- Edmonton, AB, Canada. New York: IEEE Press, 2005: 1253-1258.
- [6] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8690: 834-849.
- [7] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625.
- [8] Bergmann P, Wang R, Cremers D. Online photometric calibration of auto exposure video for realtime visual odometry and SLAM [J]. IEEE Robotics and Automation Letters, 2018, 3(2): 627-634.
- [9] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry [C] // 2014 IEEE International Conference on Robotics and Automation (ICRA), May 31-June 7, 2014, Hong Kong, China. New York: IEEE Press, 2014: 15-22.
- [10] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
- [11] Tang C Z, Wang O, Tan P. GSLAM: initialization-robust monocular visual SLAM via global structure-from-motion[C] // 2017 International Conference on 3D Vision (3DV), October 10-12, 2017, Qingdao, China. New York: IEEE Press, 2017: 155-164.
- [12] Lee S J, Hwang S S. Elaborate monocular point and line SLAM with robust initialization[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1121-1129.
- [13] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge: Cambridge University Press, 2004.
- [14] Goldstein A, Fattal R. Video stabilization using epipolar geometry[J]. ACM Transactions on Graphics, 2012, 31(5): 1-10.
- [15] Kümmerle R, Grisetti G, Strasdat H, et al. g<sup>2</sup>o: a general framework for graph optimization[C]//2011 IEEE International Conference on Robotics and Automation, May 9-13, 2011, Shanghai, China. New York: IEEE Press, 2011: 3607-3613.
- [16] Agarwal S, Mierle K. Ceres solver[EB\OL]. [2020-11-28]. <http://www.ceres-solver.org/>.
- [17] Zhang Z Y. Estimating motion and structure from correspondences of line segments between two perspective images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(12): 1129-1139.
- [18] Nelder J A, Mead R. A simplex method for function minimization[J]. The Computer Journal, 1965, 7(4): 308-313.
- [19] von Gioi R G, Jakubowicz J, Morel J M, et al. LSD: a line segment detector [J]. Image Processing on Line, 2012, 2: 35-55.
- [20] Zhang L L, Koch R. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency [J]. Journal of Visual Communication and Image Representation, 2013, 24(7): 794-805.
- [21] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 7-12, 2012, Vilamoura-Algarve, Portugal. New York: IEEE Press, 2012: 573-580.
- [22] Shi X, Li D, Zhao P, et al. Are we ready for service robots? The OpenLORIS-scene datasets for lifelong SLAM[C]//2020 IEEE International Conference on Robotics and Automation (ICRA), May 31-August 31, 2020, Paris, France. New York: IEEE Press, 2020: 3139-3145.