

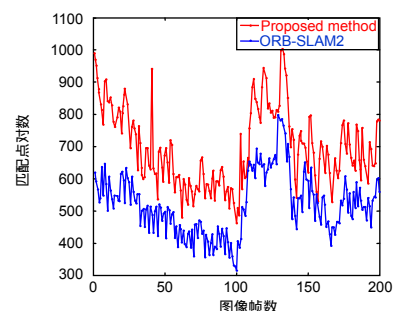


DOI: 10.12086/oe.2018.180244

三阶段局部双目光束法平差视觉里程计

赵彤*, 刘洁瑜, 李卓

火箭军工程大学控制科学与工程系, 陕西 西安 710025



摘要: 针对光束法对初始值依赖性大以及双目相机模型的特点, 在 ORB-SLAM2 算法基础提出了一种三阶段局部双目光束法平差算法。在帧间位姿跟踪阶段, 为降低累积误差对匀速模型下 3D-2D 匹配的影响, 引入环形匹配机制进一步剔除误匹配, 并将关键帧地图点与当前帧 3D-2D 投影匹配; 在跟踪局部地图阶段插入关键帧时, 增加双目相机基距参数优化环节; 在局部地图优化阶段, 将最近的两关键帧间的普通帧也作为局部帧进行优化。KITTI 数据集实验表明, 三阶段局部双目光束法平差与 ORB-SLAM2 相比, 构造了更多精确的 3D-2D 匹配, 增加了优化约束条件, 提高了运动估计与优化精度。

关键词: ORB-SLAM2 算法; 三阶段局部双目光束法平差; 帧间位姿跟踪; 跟踪局部地图; 局部地图优化词

中图分类号: TP242

文献标志码: A

引用格式: 赵彤, 刘洁瑜, 李卓. 三阶段局部双目光束法平差视觉里程计[J]. 光电工程, 2018, 45(11): 180244

Visual odometry with three-stage local binocular BA

Zhao Tong*, Liu Jieyu, Li Zhuo

Department of Control Science and Engineering, Rocket Force University of Engineering, Xi'an, Shaanxi 710025, China

Abstract: In this paper, a three-stage local binocular BA (bundle adjustment) is proposed based on the ORB-SLAM2 algorithm, which is based on the large value of the initial value and the binocular camera model. In order to reduce the influence of cumulative error on 3D-2D matching in the uniform model, the ring matching mechanism is introduced to eliminate the mismatched again and match the key frame map point with the current frame 3D-2D projection. In the tracking part of the local map optimization phase, the normal frame between the two nearest key frames is also optimized as the local frame when the key phase is inserted into the key frame. KITTI data set experiments show that the three-stage local binocular beam method has more accurate 3D-2D matching compared with ORB-SLAM2, which improves the optimization constraint and improves the motion estimation and optimization precision.

Keywords: ORB-SLAM2 algorithm; three-stage local binocular BA; inter-frame pose tracking; local map tracking; local map optimization

Citation: Zhao T, Liu J Y, Li Z. Visual odometry with three-stage local binocular BA[J]. *Opto-Electronic Engineering*, 2018, 45(11): 180244

收稿日期: 2018-05-09; 收到修改稿日期: 2018-07-24

基金项目: 国家自然科学基金资助项目(61203007, 61304001)

作者简介: 赵彤(1994-), 男, 硕士研究生, 主要从事机器视觉, 深度学习的研究。E-mail: 601080018@qq.com

1 引言

视觉里程计(Visual odometry, VO)一般将单帧运动估计级联获取摄像机的全局导航信息,从而使得误差不断累积。为获取大规模复杂环境下全局一致的导航结果,基于图优化的 VSLAM 成为研究热点,图优化的节点是需要优化的变量,边是与优化变量相关的约束条件^[1]。由于实时性的需求,局部光束法(local bundle adjustment, LBA)采用时间滑窗的形式建立多帧图像序列间的特征点匹配关系,有效解决了优化计算代价随时间推移逐步增加的问题,LBA 也成为了运动估计优化中常见模块。Klein 和 Murray^[2]提出的平行跟踪和地图构建算法是 VSLAM 发展历程上的一个重要突破。其后的几年间 VSLAM 愈加完善,涵盖了信息关联视图^[3]、图优化通用框架^[4]、基于词袋^[5]的位置识别等,出现了众多精度较高的 VSLAM 系统^[6-8]。Bellavia 和 Badino 对关键环节进行了改进,分别提出了技巧性较强的特征点回路链匹配^[6]以及多帧特征整合^[7]的策略,提高了运动估计的精度。Cvišić I 研究出一种基于点特征选取和跟踪的双目视觉系统^[8],该系统使用五点法来估计旋转变换并利用三点法来估计平

移量,有效抑制了漂移。上述方法基本上都是基于光束法平差(bundle adjustment, BA)进行运动估计与优化。

BA 优化的前提条件是提供较为准确的匹配和位姿初始值,对初始值存在较大的依赖性。与此同时,如何有效利用局部图像序列间优化约束条件是提高运动估计与优化精度的重要途径。对此,许允喜^[9]提出了一种基于多帧序列运动估计的实时视觉导航算法。罗杨宇^[10]采用光束平差对相机运动参数进行分段优化。卢维提出了一种两阶段局部双目光束法平差算法(two-stage local binocular bundle adjustment, TLBBA),第一阶段使用两帧 LBA 优化单步运动参数,第二阶段采用基于滑动窗口的多帧 LBA^[11]。

2016年 Mur-Artal 提出的 ORB-SLAM2^[12]开源算法计算效率高,具备在 CPU 配置下实时运行的能力,可以为室内环境下手持载体、工业环境下飞行器以及城市环境下驾驶的车辆进行包括地图再建、回环检测以及重定位等功能的视觉导航(如图 1 所示)。ORB-SLAM2 算法具备单目、双目以及深度相机接口,在此基础上继续开发得到了广泛的研究^[13-14]。

本文将在 ORB-SLAM2 算法基础上,设计三阶段

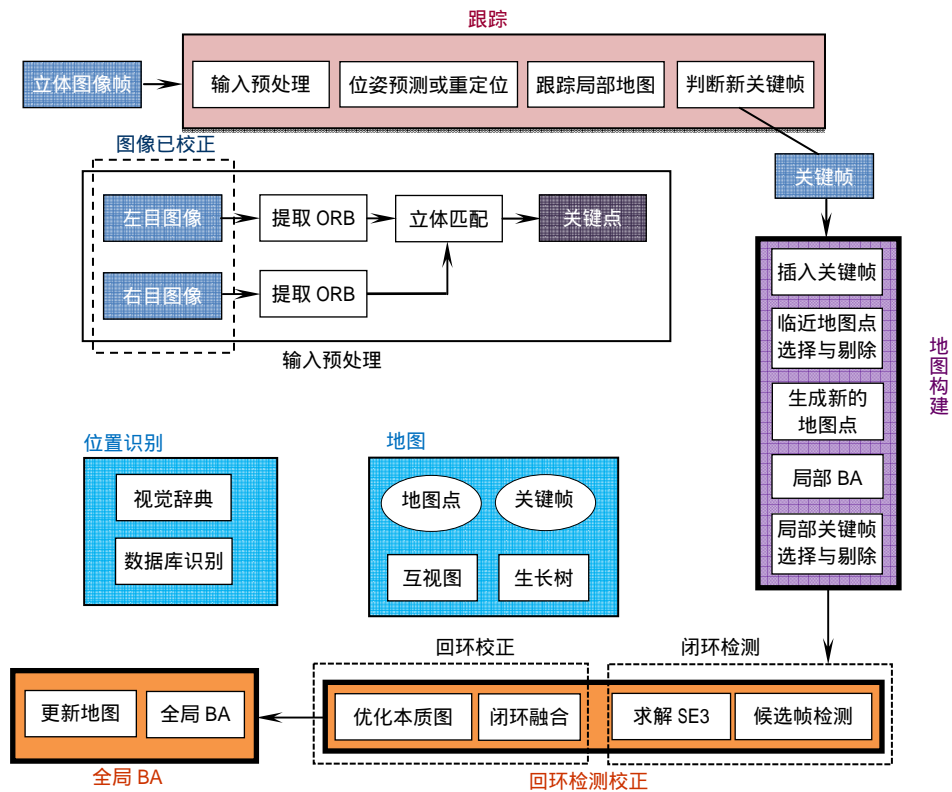


图 1 ORB-SLAM2 双目视觉导航框图

Fig. 1 ORB-SLAM2 binocular visual navigation block diagram

局部双目光束法平差算法，充分考虑匹配和位姿初始值以及优化的约束条件，分阶段优化，逐步提高导航精度。

2 重投影误差与光束法平差

如图 2 所示，假设有相机位姿 $T_{i=1,2,\dots,m}$ 和 3D 特征点 $y_{j=1,2,\dots,n}$ ，且在相机 T_i 的图像中观测提取到 y_j 的所在位置为 z_{ij} ，根据相机投影模型 y_j 在 T_i 中投影预测位置为 $\hat{z}(T_i, y_j)$ ，则观测位置和预测位置间误差(重投影误差)为

$$e_{ij}(T_i, y_j) = z_{ij} - \hat{z}(T_i, y_j) \quad (1)$$

相机线性投影模型为

$$\hat{z}(x) = \text{proj}(x) = K \cdot \frac{1}{x_3} x \quad (2)$$

式中： $x = [x_1, x_2, x_3]^T \in \mathbb{R}^3$, $K = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \end{bmatrix}$ 。(在式(2)

中， K 阵除了 x_3 项，因此此处表达比一般的 K 阵少了 $[0,0,1]$ 这行。)

则 y_j 在 T_i 中的投影坐标 x_{ij} 为

$$x_{ij} = [I_{3 \times 3} \quad 0_{3 \times 1}] \cdot T_i \cdot y_j$$

$$\hat{z}(T_i, y_j) = \text{proj}([I_{3 \times 3} \quad 0_{3 \times 1}] \cdot T_i \cdot y_j) \quad (3)$$

因此 y_j 在 T_i 图像中预测位置为 $\hat{z}(T_i, y_j)$ ，观测位置和预测位置间误差(重投影误差) $e_{ij}(T_i, y_j)$ 关于相机位姿 ξ 的雅克比矩阵为

$$\left. \frac{\partial e_{ij}(\exp(\xi_i^\wedge) T_i, y_j)}{\partial \xi_i} \right|_{\xi_i=0} = - \left. \frac{\partial \hat{z}(x)}{\partial x} \right|_{x=[I_{3 \times 3} \quad 0_{3 \times 1}] T_i y_j} \cdot \left. \frac{\partial x(\exp(\xi_i^\wedge) T_i, y_j)}{\partial \xi_i} \right|_{\xi_i=0} \quad (4)$$

根据李群李代数相关知识，令 $\dot{x}_{ij} = T_i \cdot \dot{y}_j$ ，则有：

$$\left. \frac{\partial (\exp(\xi_i^\wedge) T_i, y_j)}{\partial \xi_i} \right|_{\xi_i=0} = \begin{bmatrix} -x_{ij}^\wedge & I_{3 \times 3} \\ 0 & 0 \end{bmatrix}$$

$$\Rightarrow \frac{\partial (\exp(\xi_i^\wedge) T_i, y_j)}{\partial \xi_i} = \begin{bmatrix} -x_{ij}^\wedge & I_{3 \times 3} \end{bmatrix} \quad (5)$$

因此，重投影关于相机位姿的雅克比矩阵为

$$\left. \frac{\partial e_{ij}(\exp(\xi_i^\wedge) T_i, y_j)}{\partial \xi_i} \right|_{\xi_i=0} = - \begin{bmatrix} \frac{f_u}{x_3} & 0 & -\frac{f_u x_1}{x_3^2} \\ 0 & \frac{f_v}{x_3} & -\frac{f_v x_2}{x_3^2} \end{bmatrix} \cdot \begin{bmatrix} -x_{ij}^\wedge & I_{3 \times 3} \end{bmatrix} \quad (6)$$

式中：

$$x = [x_1, x_2, x_3]^T = R_i y_j + t_i = T_i \begin{bmatrix} y_j \\ 1 \end{bmatrix}$$

同时，重投影误差对于 3D 特征点 y_j 的雅克比矩阵为

$$\frac{\partial e_{ij}(T_i, y_j)}{\partial y_j} = - \frac{\partial \hat{z}(x)}{\partial x} \cdot \frac{\partial x(T_i, y_j)}{\partial y_j}$$

$$= - \frac{\partial \hat{z}(x)}{\partial x} \cdot \frac{\partial (R_i y_j + t_i)}{\partial y_j}$$

$$= - \begin{bmatrix} \frac{f_u}{x_3} & 0 & -\frac{f_u x_1}{x_3^2} \\ 0 & \frac{f_v}{x_3} & -\frac{f_v x_2}{x_3^2} \end{bmatrix} \cdot R_i \quad (7)$$

光束法平差是通过调整相机位姿 $T_{i=1,2,\dots,m}$ 和 3D 特征点 $y_{j=1,2,\dots,n}$ 使得重投影误差之和最小，此优化问题为

$$\min_{T_i, y_j} \sum_{ij} e_{ij}^T W_{ij} e_{ij} = e^T W e \quad (8)$$

由于并非相机每个位姿 T_i 都可以观测到所有 y_j ，因此雅克比矩阵具有稀疏性，线性求解器高效计算更

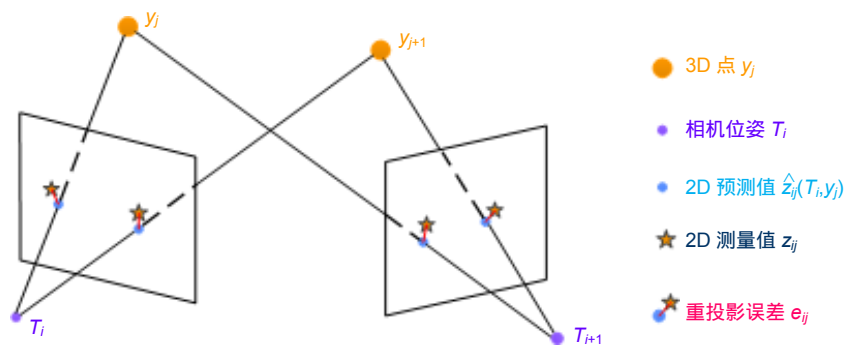


图 2 重投影误差
Fig. 2 Reprojection error

新量。同时式(8)包含了 3D 点在所有相机位姿 T_i 图像中的重投影误差,也被称为全局 BA,当图像帧数和 3D 点数量较多时,需要优化的参数空间维度和雅克比矩阵 J 规模会很大。为了满足实时性要求,局部 BA 将图像序列限定在一定的时间或者空间窗口内,降低了参数的空间维数,使得计算代价不随图像序列的增长而增加,又能充分利用相邻帧间的约束关系^[15]。

3 三阶段局部双目光束法平差

设计三阶段局部双目光束法平差算法的流程如图 3 所示,具体包括以下步骤:

1) 初值设定。车辆运动过程中实时采集图像序列,在跟踪上一帧成功情况下进入帧间位姿跟踪,首先根据匀速模型得出当前帧位姿的假设值,将上一帧

以及关键帧中的地图点与当前帧进行 3D-2D 投影匹配、提纯、排序,若不能找到足够的匹配,则进入跟踪参考帧模式,采用 BoW 方式在参考帧中寻找与当前帧特征的匹配关系。

2) 帧间位姿跟踪。根据步骤 1)获得的优化初值进行基于 MoBA 的最小化重投影误差,获得初步优化后的相机位姿。

3) 跟踪局部地图。跟踪局部地图首先进行局部地图的更新,进而根据步骤 2)获得的摄像机位姿来将局部地图点与当前帧投影匹配,最后利用非线性优化重投影误差进一步提高摄像机位姿优化精度。当插入关键帧时,对双目相机基距参数进行优化。

4) 局部地图优化。对最近两关键帧间的普通帧以及和当前关键帧相连的关键帧所构成的局部帧及地图

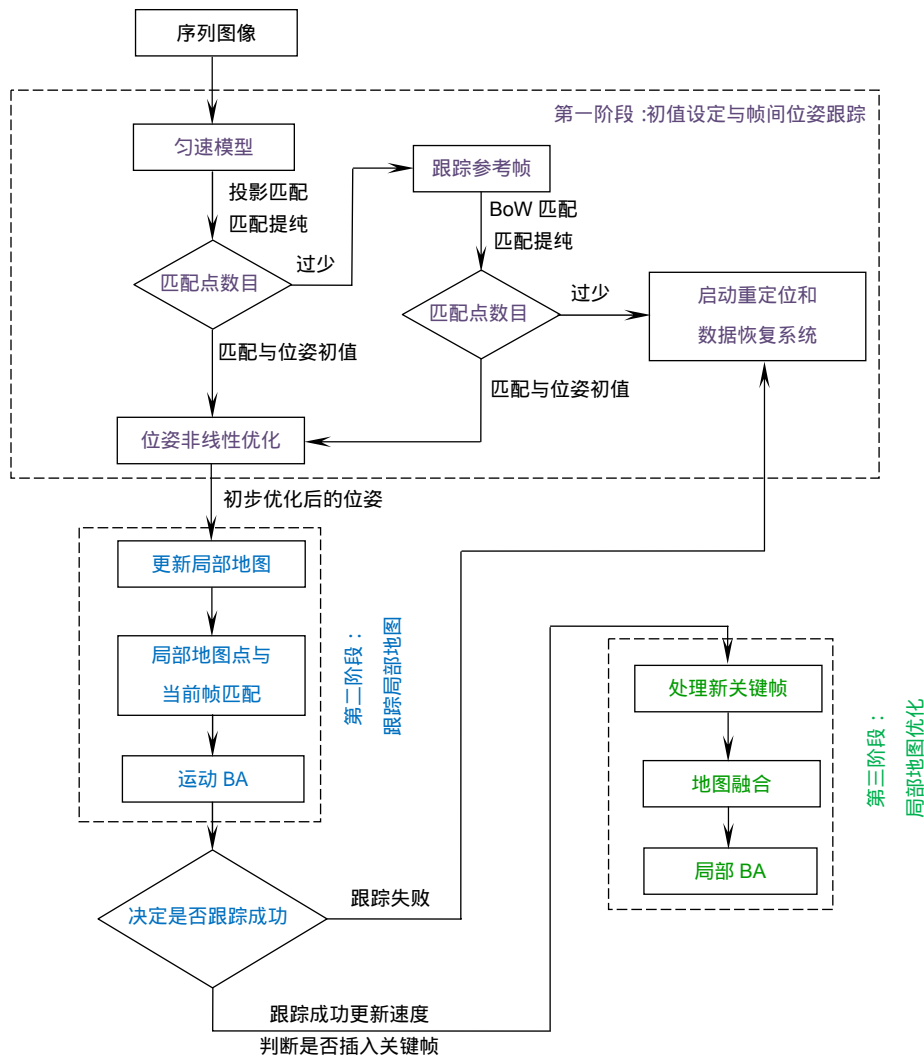


图 3 三阶段局部双目光束法平差算法流程图

Fig. 3 Visual odometry with three-stage local binocular BA algorithm flowchart

点进行 LBA 优化。

下面对每个步骤进行具体的阐述。

3.1 初值设定

本文提取的图像特征为文献 [16] 提出的 ORB-LATCH 特征，定义为 $x^i = (u_L, v_L, u_R)$ ，其中 (u_L, v_L) 为特征点在左目上的图像坐标， u_R 为在右目上的图像坐标。同时采取和 ORB-SLAM2 一样的稀疏地图构建方法，地图点是根据特征点景深以及当前帧位姿所构造的具有世界坐标系的 3D 点，关键帧包含相机位姿和在图像帧上提取到的 ORB-LATCH 特征。

非线性优化对初值存在依赖，需要设定相机的位姿初值以及匹配初值，即三阶段局部双目光束法平差算法第一阶段优化的初值。ORB-SLAM2 中采用多种匹配策略和技巧来提高初值的准确性，利用匀速模型对相机当前帧位姿进行预测，从而为上一帧所观察到的地图点在当前帧中的投影搜索区域提供指导，并与在当前帧中预测位置附近的特征进行匹配，最后通过距离阈值、比例阈值、角度投票剔除误匹配。在此基础上，本文根据双目视觉特点引入四边形环形匹配机制，根据是否满足环形匹配约束再次提纯特征匹配，保证匹配正确性。对前后帧的两对图像构建如图 4 所示的四边形环形匹配结构^[17]，箭头方向为匹配次序。

四边形环形匹配约束包含水平约束和垂直约束，水平方向上约束为

$$v_1 = v_2, v_3 = v_4, \quad (9)$$

$$|u_1 - u_2|, |u_3 - u_4| \leq d_{\max}, \quad (10)$$

式中 d_{\max} 是水平方向上帧间变化的距离阈值。

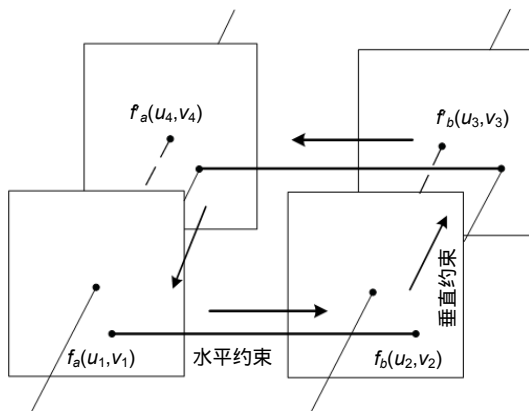


图 4 双目环形匹配结构

Fig. 4 Binocular ring matching structure

在垂直方向上，同一特征点在前后帧图像中位置变化相同，垂直约束为

$$\left| \frac{v_4 - v_1}{u_4 - u_1} - \frac{v_3 - v_2}{u_3 - u_2} \right| < \delta. \quad (11)$$

在四边形环形匹配中，水平方向可以依据双目立体匹配获取联系，在前后帧垂直方向上的匹配可依据匀速模型指导搜索匹配，整个环形匹配过程简单高效。

3.2 帧间位姿跟踪

帧间位姿跟踪作为三阶段局部双目光束法平差的第一阶段，主要任务是对摄像机当前帧的位姿进行初次优化。ORB-SLAM2 初始位姿估计时，仅利用上一帧的地图近点进行 3D-2D 投影，三阶段局部双目光束法平差增加关键帧地图点在当前帧投影，构造更多的匹配点对(如图 5 所示，其中蓝色空心点为关键帧的地图点，红色实心点为上一帧的地图点)。这种综合考虑上一帧与关键帧的地图点的方式，可以增加当前帧位姿优化的约束条件，提高位姿估计精度。

通过最小化世界坐标系下 3D 地图点 X^i (上一帧和关键帧地图点)和匹配的特征点 x^i 之间重投影误差来实现对相机的旋转 $R \in SO(3)$ 和位置 $t \in \mathbb{R}^3$ 的优化。

$$\{R, t\} = \operatorname{argmin}_{R, t} \sum_i \rho \left(\|x^i - \pi(RX^i + t)\|_{\Sigma}^2 \right), \quad (12)$$

式中： ρ 是具备鲁棒性的 Huber 代价函数， Σ 是与特征点尺度相关的协方差矩阵。

优化的目标函数为

$$F = \sum_i \rho \left(\|x^i - \pi(RX^i + t)\|_{\Sigma}^2 \right), \quad (13)$$

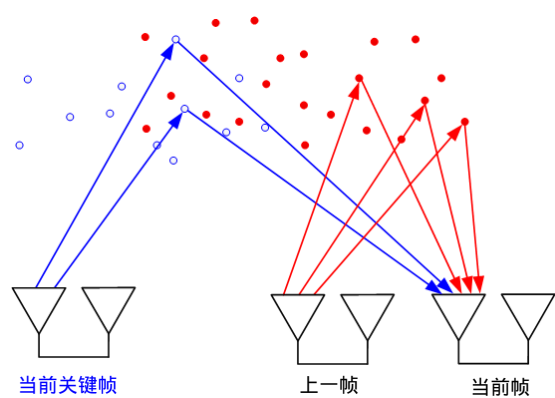


图 5 考虑关键帧地图点的帧间位姿跟踪

Fig. 5 Consider inter-frame pose tracking of keyframe map points

式(13)中投射函数 π 定义为

$$\pi \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \\ f_x \frac{X-B}{Z} + c_x \end{pmatrix}, \quad (14)$$

式中: (f_x, f_y) 是摄像机焦距, (c_x, c_y) 是投影中心, B 是基距长度, 以上参数均通过相机的标定来获取。

BA 优化分为两种: 仅优化相机位姿的 MoBA(Motion BA)和同时优化运动和结构的 LBA。MoBA 构建一元边, 在优化相机位姿过程中不改变地图点参数信息, LBA 则对地图点和相机位姿同时进行优化。帧间位姿跟踪通过 MoBA 来最小化重投影误差, 优化的初值由 3.1 节生成。

3.3 跟踪局部地图

跟踪局部地图是 TLBBA 算法的第二阶段, 主要任务是进一步优化摄像机位姿并决定何时插入关键帧。主要包括更新局部地图、投影匹配、最小化重投影误差以及判断是否插入关键帧四部分。更新局部地图完成对局部地图点和局部关键帧的更新, 投影匹配可以获取局部地图点与当前帧的 3D-2D 匹配, 从而构成优化的匹配初值, 位姿初值为帧间位姿跟踪后的相机位姿, 通过 MoBA 来最小化重投影误差, 最后根据跟踪到的近点数量来判断是否插入关键帧。

由于双目相机需要足够长度的基距以有效应对车辆行驶环境, 且立体匹配和三维重建均需要精确的基距参数。因此, 考虑相机标定参数在运动过程中可能发生变化, 在每次插入关键帧对双目相机的基距进行优化。构造双目左右相机间的重投影误差, 此时将相

机基距 B 看作变量, 通过最小化重投影误差, 求解最优的基距 B , 则重投影误差 $e_{ij}(T_i, y_j)$ 对基距 B 的导数为

$$\begin{aligned} \frac{\partial e_{ij}(T_i, y_j)}{\partial B} &= \frac{\partial (z_{ij} - \hat{z}(T_i, y_j))}{\partial B} \\ &= -\frac{\partial \hat{z}(T_i, y_j)}{\partial B} = -\frac{\partial [u_L \quad v_L \quad u_R]^T}{\partial B} \\ &= -\frac{\partial \left[f_x \frac{X}{Z} + c_x \quad f_y \frac{Y}{Z} + c_y \quad f_x \frac{X-B}{Z} + c_x \right]^T}{\partial B} \\ &= \begin{bmatrix} 0 & 0 & \frac{f_x}{Z} \end{bmatrix}^T. \end{aligned} \quad (15)$$

通过最小化此函数, 计算出 B 的优化值。对关键帧时刻双目相机的基距进行优化后运用于普通帧的三角测量, 可获取更加准确的地图点坐标。

3.4 局部地图优化

关键帧策略保证了优化的计算效率, 也能够通过关键帧间的三角测量构建精确的 3D 地图点。ORB-SLAM2 中将当前关键帧具有共视关系^[3]的关键帧定义为局部帧, 通过优化局部帧和局部地图点提高关键帧位姿估计精度。

如图 6 所示, 三阶段局部双目光束法平差将最近两关键帧间的普通帧也纳入局部帧进行优化, 为下一步的相机位姿跟踪提供更多数量的精确局部地图点, 提高相机位姿跟踪的精确性。

在三阶段局部双目光束法平差的第三阶段, 对局部帧及其地图点进行基于 LBA 的局部地图优化, 对具备信息互视关联的关键帧 K_L 和在这些关键帧中能观察到的地图点 P_L 进行优化。除 K_L 之外的其他所有能观察到 P_L 中点的关键帧 K_F 为代价函数提供约束, 但

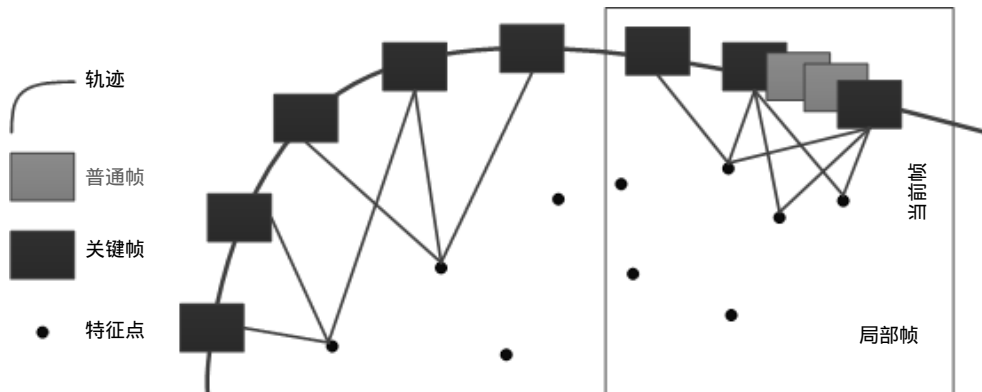


图 6 考虑普通帧的局部地图优化

Fig. 6 Local map optimization considering ordinary frames

在优化中保持固定值。定义 χ_k 为 P_L 中的地图点与关键帧 k 中特征点之间的匹配点集, 从而转化为如式(16)的优化问题。

$$\begin{aligned} & \{x^i, R_l, t_l | i \in P_L, l \in K_L\} \\ & = \arg \min_{x^i, R_l, t_l} \sum_{k \in K_L} \sum_{j \in \chi_k} \rho(E(k, j)), \quad (16) \end{aligned}$$

式中 $E(k, j) = \|x^j - \pi(R_k X^j + t_k)\|^2$ 。

经过上述四个步骤, 相机位姿和地图得到了优化, 为下一帧位姿跟踪提供更加精确的地图点和位姿初始估计。

与 ORB-SLAM2 相比, 三阶段局部双目光束法平差在初值设定和帧间位姿跟踪阶段引入四边形环形匹配机制进一步剔除误匹配, 增加关键帧地图点在当前帧的投影; 在跟踪局部地图阶段对插入关键帧时刻双

目相机的基距进行优化; 在局部地图优化阶段, 将最近的两关键帧间的普通帧也视为局部帧。

4 实验分析

为对比分析本文算法、TLBBA 以及 ORB-SLAM2 算法性能, 实验采用与测试评价 TLBBA^[11,18]、ORB-SLAM2 相一致的 KITTI 数据集。测试设备为 DELL OPTIPLEX 7010 台式电脑(Intel Core i5-3470 CPU, 主频 3.20 GHz, 4.00 G 内存)。

首先分析基于本文算法的 VO 导航轨迹效果, 并与导航车实际轨迹对比。由于篇幅所限, 这里仅展示在 01、02、08、09 序列下的导航轨迹结果(如图 7 所示), 分别对应高速、城市道路、小镇、乡村环境。红色实线为 OXTS RT 3000 IMU/GPS 测量系统得到的车

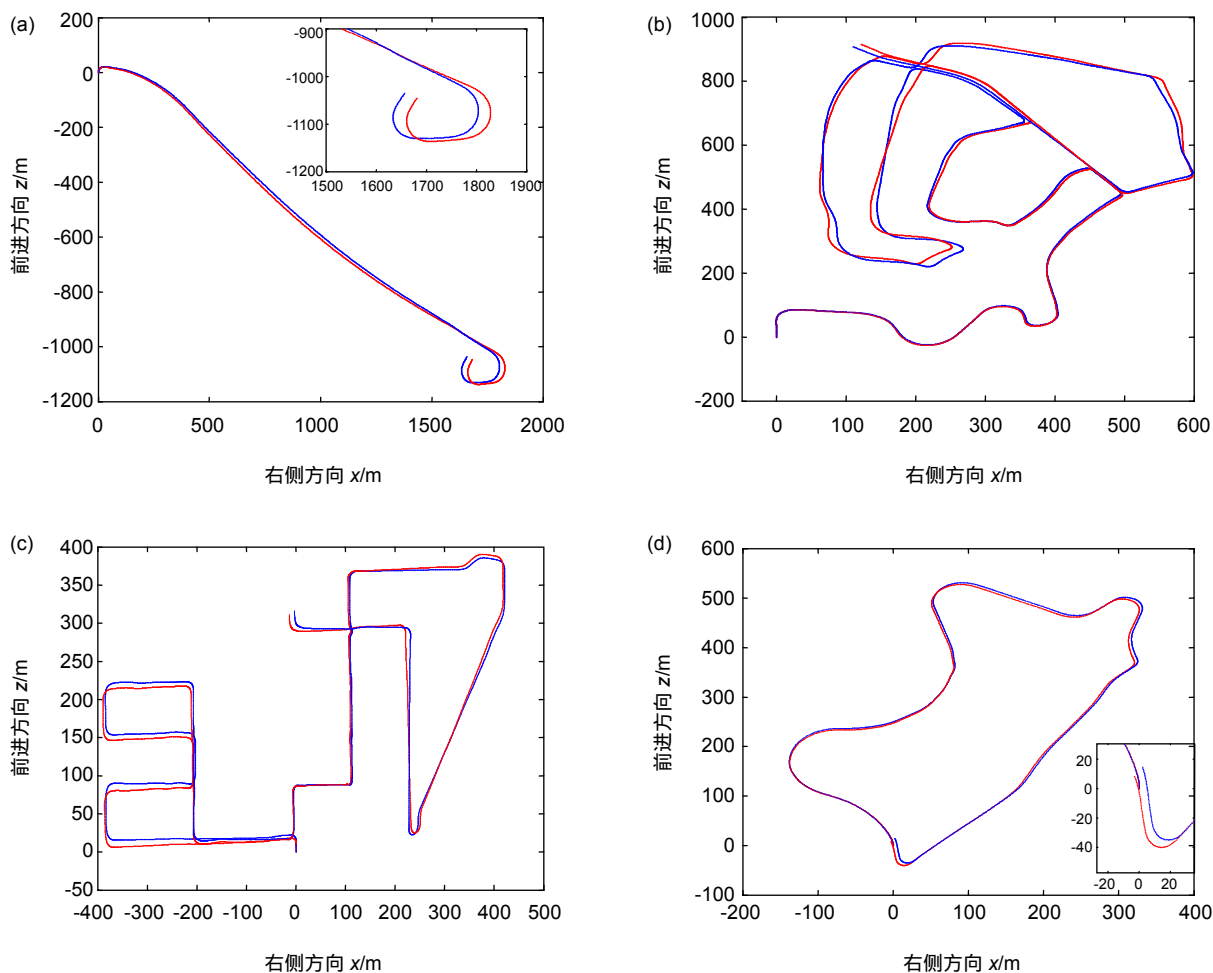


图 7 VO 导航轨迹与导航车实际轨迹对比图。(a) 01 序列车辆行驶轨迹; (b) 02 序列车辆行驶轨迹; (c) 08 序列车辆行驶轨迹; (d) 09 序列车辆行驶轨迹

Fig. 7 The comparison between the navigation track and the actual track of the navigation vehicle. (a) 01 sequence vehicle trajectory; (b) 02 sequence vehicle trajectory; (c) 08 sequence vehicle trajectory; (d) 09 sequence vehicle trajectory

辆行驶轨迹(真值), 蓝色虚线为基于本文算法 VO 所得到的导航轨迹。

可以看出, 不论导航车在 KITTI 测试序列的何种环境下行驶, 基于本文算法 VO 导航效果良好。图 7 中仅展示了基于本文算法的 VO 导航轨迹, 无法有效与 TLBBA 以及 ORB-SLAM2 进行对比评价。为进一步有效评估分析本文算法、TLBBA 以及 ORB-SLAM2 的 VO 导航精度, 在实验时关闭 ORB-SLAM2 的闭环检测功能并采用文献[19]中平均平移误差 t_{rel} 和平均旋转误差 r_{rel} 作为精度衡量指标。实验结果如表 1 所示, 其中误差较低的数据通过粗体突出显示。

从表 1 可以看出, 在大多数序列中基于本文算法的 VO 导航精度优于 TLBBA 和 ORB-SLAM2。其中 01 序列下本文算法与 TLBBA、ORB-SLAM2 均存在较大的平移误差和较小的旋转误差, 主要原因是导航车的高速行驶用于跟踪的特征近点较少但远点丰富。在实时性方面, 基于 TLBBA 并采用 GPU 加速 SIFT 特征构建的双目 VO 在 10 序列下的平均每帧处理用时为 49 ms^[3], 本文算法和 ORB-SLAM2 在测试序列下平均每帧耗时约为 70 ms, 能够满足 KITTI 数据集 10 Hz 的帧频要求。

图 8(a)和图 8(b)分别为位姿估计平移误差和旋转误差曲线图, 其中红色实线为基于本文算法 VO 所得轨迹误差曲线, 蓝色实线为基于 ORB-SLAM2 所得轨迹误差曲线, 黑色实线为基于 TLBBA 算法 VO 所得轨迹误差曲线。

从实验结果可以得出: 本文算法与 TLBBA、ORB-SLAM2 相比, 其平移和旋转误差均有所减小。在 01~10 测试序列下得到各算法导航误差: 基于 TLBBA 算法 VO 平均平移误差为 1.0372%, 平均旋转误差为 0.3667°/100 m; 基于 ORB-SLAM2 算法 VO 平均平移误差为 0.8595%, 平均旋转误差为 0.2724°/100 m; 基于本文算法 VO 平均平移误差为 0.8163%, 平均旋转误差为 0.2617°/100 m, 相比较于 ORB-SLAM2 平均平移误差下降了 0.0432%, 平均旋转误差下降了 0.0107°/100 m。

分析导航精度提高的原因, 主要是三阶段局部双目光束法平差算法在帧间位姿跟踪阶段将关键帧地图点考虑在内, 同时在初值设定阶段引入四边形环匹配剔除误匹配。在局部地图优化阶段将最近的两关键帧间的普通帧也作为局部帧, 构造了更多的匹配点对和局部地图点。上述方法均有效地增加了位姿估计时可靠的优化约束, 有利于提升位姿估计与优化精度。图 9 为 05 序列中普通帧初始位姿估计时所利用到的匹配点对数量, 可见三阶段局部双目光束法平差所使用的匹配点数量更多, 有利于位姿估计精度的提高。

与此同时, 本实验所取用的 KITTI 数据集双目相机的基距标定参数在 0.54 m 上下浮动, 变化量只在毫米数量级。在各序列标定中相机基距参数变化很小, 因此在 KITTI 数据集下优化基距 B 的方法对位姿估计的精度提高不够明显, 但是可以预见在应用其他低成本相机时, 对双目相机的基距 B 进行优化是必要的。

表 1 本文算法、ORB-SLAM2、TLBBA 实验结果对比
Table 1 The experiment result of algorithm, ORB-SLAM2, TLBBA

序列号	本文算法			ORB-SLAM2 算法			TLBBA 算法		
	$t_{rel}/\%$	r_{rel} /(rad/100 m)	平均每帧 耗时/s	$t_{rel}/\%$	r_{rel} /(rad/100 m)	平均每帧 耗时/s	$t_{rel}/\%$	r_{rel} /(rad/100 m)	平均每帧 耗时/s
00	0.805	0.47	0.070	0.847	0.49	0.068	0.984	0.72	
01	1.295	0.38	0.076	1.362	0.39	0.075	2.383	0.70	
02	0.811	0.50	0.069	0.852	0.52	0.067	0.976	0.60	
03	0.663	0.22	0.071	0.705	0.24	0.071	1.055	0.69	
04	0.457	0.30	0.071	0.476	0.31	0.070	1.215	0.35	
05	0.525	0.36	0.073	0.553	0.38	0.072	0.747	0.59	
06	0.781	0.41	0.070	0.831	0.43	0.069	1.146	0.64	
07	0.793	0.65	0.080	0.826	0.69	0.079	0.843	1.01	
08	0.980	0.54	0.075	1.021	0.52	0.074	1.135	0.61	
09	0.829	0.47	0.073	0.882	0.46	0.071	1.045	0.49	
10	0.567	0.40	0.078	0.604	0.42	0.077	0.543	0.75	0.049

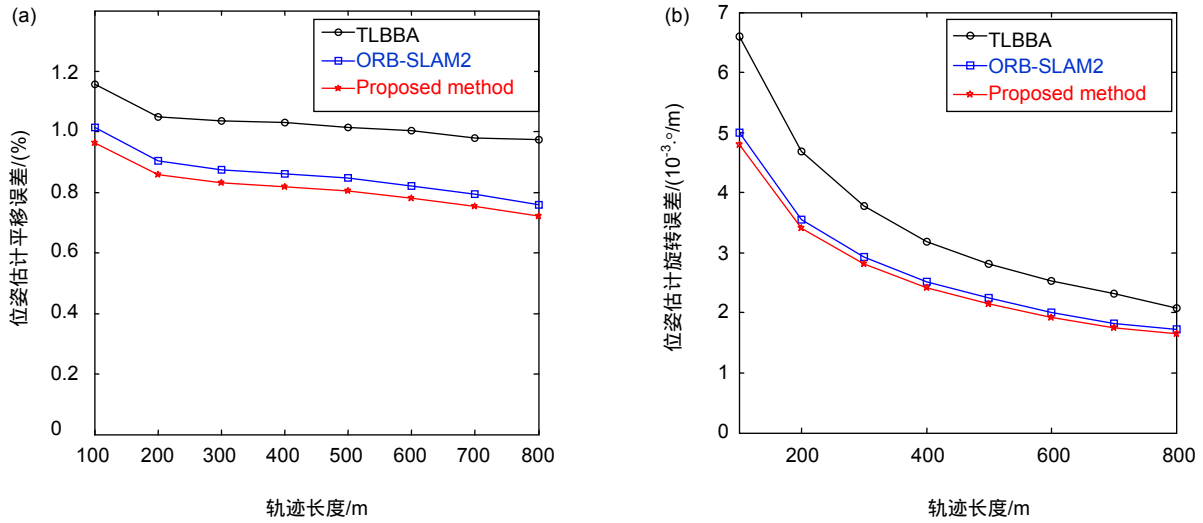


图 8 KITTI 数据集测试序列实验结果。(a) 位姿估计平移误差曲线; (b) 位姿估计旋转误差曲线
 Fig. 8 KITTI data set test sequence experimental results. (a) Pose estimation translation error curve; (b) Pose estimation rotation error curve

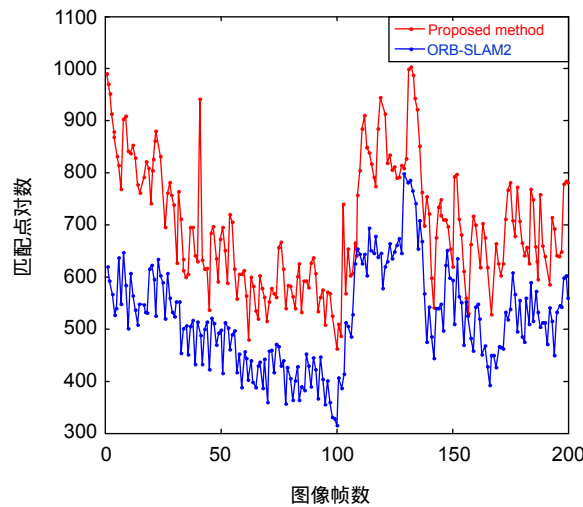


图 9 05 序列下匹配点对数目对比
 Fig. 9 Comparison of matching point pairs under KITTI 05 sequence

归纳总结上述实验结果，得出结论：与 ORB-SLAM2 相比，本文算法可构造更多精确的匹配点对和局部地图点，增加了位姿估计时的优化约束，提高了位姿估计与优化精度。在实时性方面，基于本文算法 VO 满足 KITTI 数据集 10 Hz 的帧频要求。

5 结论

本文在 ORB-SLAM2 基础上设计了三阶段局部双

目光束法平差算法，在初值设定时引入四边形环形匹配进一步剔除误匹配，在帧间位姿跟踪阶段引入关键帧地图点在当前帧的投影，增加当前帧初始位姿估计的优化约束。在局部地图优化阶段将最近的两关键帧间的普通帧也作为局部帧，构造了更多的匹配点对和局部地图点。三阶段局部双目光束法平差可构造更多精确的匹配点对和局部地图点，增加了位姿估计时的优化约束，提高了位姿估计与优化精度。

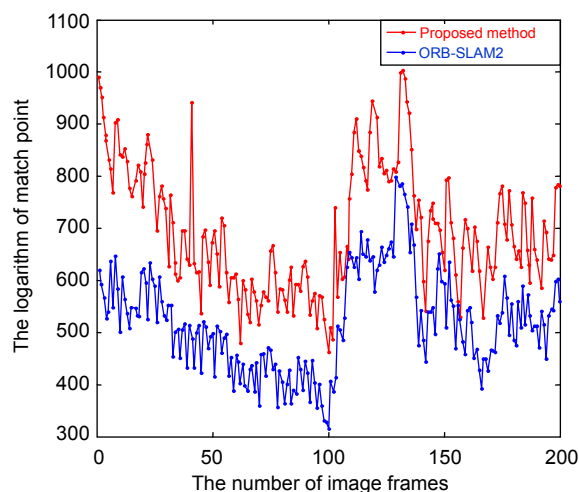
参考文献

- [1] Sibley G. Relative bundle adjustment[R]. Department of Engineering Science, Oxford University Technical Report. Oxford: Department of Engineering Science, Oxford University, 2009.
- [2] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]//*Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2007: 225–234.
- [3] Strasdat H, Davison A J, Montiel J M M, et al. Double window optimisation for constant time visual SLAM[C]//*2011 International Conference on Computer Vision*, 2011: 2352–2359.
- [4] Kümmerle R, Grisetti G, Strasdat H, et al. G²O: A general framework for graph optimization[C]//*2011 IEEE International Conference on Robotics and Automation*, 2011: 3607–3613.
- [5] Galvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences[J]. *IEEE Transactions on Robotics*, 2012, **28**(5): 1188–1197.
- [6] Bellavia F, Fanfani M, Pazzaglia F, et al. Robust selective stereo SLAM without loop closure and bundle adjustment[C]//*Proceedings of the 17th International Conference on Image Analysis and Processing*, 2013, **8156**: 462–471.
- [7] Badino H, Yamamoto A, Kanade T. Visual odometry by multi-frame feature integration[C]//*Proceedings of 2013 IEEE International Conference on Computer Vision Workshops*, 2013: 222–229.
- [8] Cvišić I, Petrović I. Stereo odometry based on careful feature selection and tracking[C]//*European Conference on Mobile Robots. IEEE*, 2015:1–6.
- [9] Xu Y X, Chen F. Real-time stereo visual localization based on multi-frame sequence motion estimation[J]. *Opto-Electronic Engineering*, 2016, **43**(2): 89–94.
许允喜, 陈方. 基于多帧序列运动估计的实时立体视觉定位[J]. *光电工程*, 2016, **43**(2): 89–94.
- [10] Luo Y Y, Liu H L. Research on binocular vision odometer based on bundle adjustment method[J]. *Control and Decision*, 2016, **31**(11): 1936–1944.
罗杨宇, 刘宏林. 基于光束平差法的双目视觉里程计研究[J]. *控制与决策*, 2016, **31**(11): 1936–1944.
- [11] Lu W, Xiang Z Y, Liu J L. High-performance visual odometry with two-stage local binocular BA and GPU[C]//*2013 IEEE Intelligent Vehicles Symposium (IV)*, 2013: 23–26.
- [12] Mur-Artal R, Tardós J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. *IEEE Transactions on Robotics*, 2016, **33**(5): 1255–1262.
- [13] Hou R B, Wei W, Huang T, et al. Indoor robot localization and 3D dense mapping based on ORB-SLAM[J]. *Computer Applications*, 2017, **37**(5): 1439–1444.
侯荣波, 魏武, 黄婷, 等. 基于 ORB-SLAM 的室内机器人定位和三维稠密地图构建[J]. *计算机应用*, 2017, **37**(5): 1439–1444.
- [14] Zhou S L, Wu X Z, Liu G, et al. Integrated navigation method of monocular ORB-SLAM/INS[J]. *Journal of Chinese Inertial Technology*, 2016, **24**(5): 633–637.
周绍磊, 吴修振, 刘刚, 等. 一种单目视觉 ORB-SLAM/INS 组合导航方法[J]. *中国惯性技术学报*, 2016, **24**(5): 633–637.
- [15] Lourakis M I A, Argyros A A. SBA: a software package for generic sparse bundle adjustment[J]. *ACM Transactions on Mathematical Software (TOMS)*, 2009, **36**(1): 2.
- [16] Li Z, Liu J Y, Li H, et al. Feature detection and description algorithm based on ORB-LATCH[J]. *Journal of Computer Applications*, 2017, **37**(6): 1759–1762, 1781.
李卓, 刘洁瑜, 李辉, 等. 基于 ORB-LATCH 的特征检测与描述算法[J]. *计算机应用*, 2017, **37**(6): 1759–1762, 1781.
- [17] Fan J J, Liang H W, Zhu H, et al. Closed quadrilateral feature tracking algorithm based on binocular vision[J]. *Robot*, 2015, **37**(6): 674–682.
樊俊杰, 梁华为, 祝辉, 等. 基于双目视觉的四边形闭环跟踪算法[J]. *机器人*, 2015, **37**(6): 674–682.
- [18] Lu W. Research on key techniques of high-precision and real-time visual localization[D]. Hangzhou: Zhejiang University, 2015.
卢维. 高精度实时视觉定位的关键技术研究[D]. 杭州: 浙江大学, 2015.
- [19] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset[J]. *The International Journal of Robotics Research*, 2013, **32**(11): 1231–1237.

Visual odometry with three-stage local binocular BA

Zhao Tong*, Liu Jieyu, Li Zhuo

Department of Control Science and Engineering, Rocket Force University of Engineering, Xi'an, Shaanxi 710025, China



Comparison of matching point pairs under KITTI 05 sequence

Overview: Visual odometry (VO) generally cascades single-frame motion estimation to obtain global navigation information of the camera, so that errors accumulate. In order to obtain globally consistent navigation results in large-scale complex environments, VSLAM based on graph optimization has become a research hotspot. ORB-SLAM2 is an open source algorithm proposed by Mur-Artal in 2016. The High computational efficiency and the ability to run in real time under CPU configuration make it can be used for Visual navigation of features such as map reconstruction, loop detection, and relocation in many scenes such as handheld carrier in indoor environment, aircraft in industrial environment, vehicles driven in urban environments and so on. The ORB-SLAM2 algorithm has a monocular, binocular and depth camera interface, and has been extensively researched on the basis of this. In this paper, a three-stage local binocular BA is proposed based on the ORB-SLAM2 algorithm, which is based on the large value of the initial value and the binocular camera model. On the basis of the ORB-LATCH feature proposed in, in order to reduce the influence of cumulative error on 3D-2D matching in the uniform model, the ring matching mechanism is introduced. Re-purifying the feature matching according to whether the ring matching constraint is satisfied, ensuring the correctness of the matching, to eliminate the mismatched again and match the key frame map point with the current frame 3D-2D projection. In the tracking of the local map phase, since the binocular camera requires a base distance of sufficient length to effectively cope with the driving environment of the vehicle, both the stereo matching and the three-dimensional reconstruction require accurate base distance parameters. Therefore, considering that the camera calibration parameters may change during the motion, we optimize the base distance of the binocular camera each time a key frame is inserted. In the local map optimization stage, the normal frames between the last two key frames are also included in the local frame for optimization, which provides a larger number of accurate local map points for the next camera pose tracking, and improves the accuracy of the camera pose tracking. Experiment with the algorithm under the KITTI data set. The result shows that the three-stage local binocular beam method has more accurate 3D-2D matching compared with ORB-SLAM2, which improves the optimization constraint and improves the motion estimation and optimization precision. In terms of real-time, the VO based on the algorithm in this paper meets the frame rate requirement of 10 Hz for the KITTI data set.

Citation: Zhao T, Liu J Y, Li Z. Visual odometry with three-stage local binocular BA[J]. *Opto-Electronic Engineering*, 2018, 45(11): 180244

Supported by National Natural Science Foundation of China (61203007, 61304001)

* E-mail: 601080018@qq.com