DOI:10.13718/j.cnki.xsxb.2021.02.019

半直接法与 IMU 融合的双目视觉里程计①

种一帆, 冀 杰, 宫铭钱, 陈琼红

西南大学 工程技术学院,重庆 400715

摘要:针对基于特征点法的视觉里程计实时性和鲁棒性较差的问题,提出一种基于半直接法的双目视觉惯性里程计算法. 在跟踪线程中将惯性测量数据作为先验,并使用逆光流法跟踪均匀化的特征关键点,以提高特征匹配的速度和鲁棒性,构建高精度的初始化地图,为后续的运动估计提供良好的初值. 使用简化的双目视觉模型构造重投影误差,结合 IMU 误差项构建联合优化模型,并在滑动窗口中进行非线性优化求解. 实验结果显示,该算法在数据集上的定位精度达到主流算法的水平,与 VINS－Fusion 算法相比,此算法拥有更低的 CPU 负载和更高的运行帧率.

关 键 词:实时性;鲁棒性;简化的双目视觉模型;初始化地图;非线性优化

中图分类号:TP242 **文献标志码**:A **文章编号**:1000－5471(2021)02－0112－09

随着计算机和传感器技术的不断发展,视觉 SLAM(即时定位与地图构建)逐渐成为智能驾驶、自主机器人和无人机等新兴领域的核心技术. 视觉里程计根据相机采集到的图像信息计算运动载体的位姿,并选取关键帧用于建图和后端优化,是视觉 SLAM 的核心部分. 图像特征是视觉里程计最重要的定位信息之一,对其提取和跟踪的结果将直接影响整个系统的稳定性. 根据图像信息处理方法的不同,视觉里程计主要分为特征点法和直接法两类[1]. 很多经典的算法通过计算描述子来实现对特征关键点的跟踪和匹配[2-4],但这种组合方式仍存在一些的问题:特征描述子的计算会给 CPU 带来很大压力,算法在特征纹理不明显的场景下无法工作等. LSD-SLAM 以及 DSO 等一系列的直接法算法的提出给研究者提供了解决这些问题的另一种思路[5-6]:基于图像的灰度变化对图像帧中的特征像素进行匹配,这种方式节省了特征提取以及描述子计算的时间,只要图像中存在像素的明暗变化,直接法就可以通过最小化光度误差完成相机运动的估计,然而,直接法对光照变化极其敏感,而且无法应对较大幅度的快速运动. 近年来,研究者的目光逐渐集中到基于半直接法的视觉里程计[7-10],这类算法通过特征对齐和稀疏图像对齐避免了显式的特征匹配过程,节约了描述子计算和匹配的时间,拥有极高的计算效率,同时也避免了像直接法一样需要较高的采样帧率,为在定位精度和计算效率之间找到合理的平衡提供了有效的途径.

图像信息可以提供丰富的运动约束,但当相机镜头被遮挡或是遇到纹理不明显的场景时,纯视觉里程计难以实现精准定位,而惯性测量单元可以在短期内提供可靠的运动估计,但由于 IMU 的测量特性,随着里程的增加累积误差也会变大. 作为目前主流的实时定位技术,视觉惯性里程计能够融合相机的图像信息和 IMU 的运动信息,兼具两种里程计的优势,从而获得较高的定位精度和鲁棒性. 然而,不同传感器的数据融合会增加算法的复杂性,因此需要进一步提升算法的鲁棒性和计算效率.

视觉惯性里程计根据融合框架的不同可以分为松耦合和紧耦合两类. 松耦合的框架中两种传感器信息

① 收稿日期:2020－06－08

基金项目:国家自然科学基金(61304189);中央高校基本业务费专项资金重点项目(XDJK2019B053);汽车主动安全测试技术重庆市工业和信息化重点实验室 2019 年度开放课题(19AKC8).

作者简介:种一帆,硕士研究生,主要从事计算机视觉和视觉里程计的研究.

通信作者:冀 杰,副教授,硕士生导师.

分别单独处理，其中 IMU 模块仅用于辅助视觉模块的位姿估计，这类框架的信息融合通常由扩展卡尔曼滤波完成. 紧耦合的框架则是将 IMU 和视觉约束信息放在同一个非线性优化函数中进行状态估计. 目前的研究结果表明，基于松耦合的算法在计算效率上有一定优势，但在紧耦合框架中，IMU 数据可以对视觉模块进行补充，同时视觉信息也可以矫正 IMU 的零偏，因此一般认为紧耦合算法的定位精度更高. 另外，根据求解思路的不同又可以将视觉惯性里程计分为基于滤波和基于优化的两种方法. MSCKF 是一个经典的基于滤波的紧耦合方案，它在一个滑动的窗口中按时序排列邻近帧的相机状态量，共同建立约束处理位姿优化[11]. 近年来，随着研究的深入和计算机性能的提高，基于优化的算法逐渐占据 VIO 研究的主体地位. 这类算法通过求解一个非线性优化问题，实现对历史位姿更为平滑的估计.

1　半直接法和 IMU 融合的双目视觉里程计

1.1　符号说明

本文定义了 3 个坐标系：世界坐标系 F_W、相机坐标系 F_C 以及 IMU 坐标系 F_S. 其中 F_W 的原点设置为 IMU 的初始位置，z 轴的方向与重力相同. 两图像帧之间的变换用齐次矩阵 T 表示，如 T_{WS} 代表 IMU 坐标系到世界坐标系的位姿转换，其中旋转部分记作 R_{WS}，也可以用四元数表示为 q_{WS}.

1.2　算法框架

本文提出了一种基于视觉惯性融合的双目视觉里程计，该算法的主要框架如图 1 所示. 前端的跟踪线程中，将 IMU 先验位姿作为初值，使用半直接法进行相邻图像帧之间的运动估计. 系统初始化模块主要用于对齐视觉和惯性信息，并构建高精度初始地图. 后端进行局部优化，结合视觉和惯性信息构建联合优化模型，在滑动窗口中对关键帧位姿以及地图点进行优化. 本文的主要贡献如下：

1）提出了一种半直接法和 IMU 融合的双目视觉惯性里程计算法框架；

2）使用逆光流法跟踪均匀化的 FAST 关键点，提高图像处理的计算效率；

3）基于双目相机的特点，提出一种简化的双目视觉模型以减少图 1 右半部图像中的多余测量值的计算；

4）结合视觉和惯性测量信息，构建高精度的初始化地图，为之后的运动估计和位姿优化提供基础.

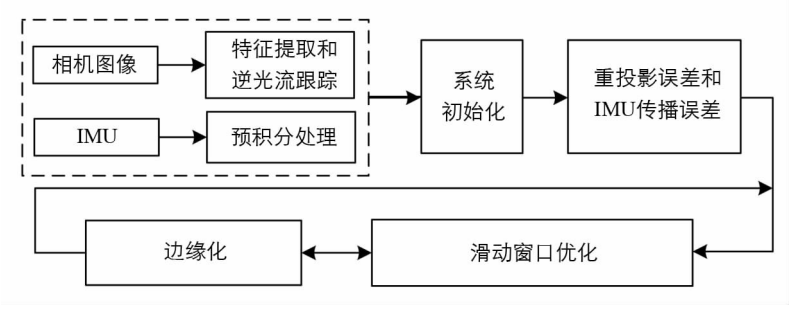

图 1　本文算法流程图

1.3　IMU 预积分

通常情况下，IMU 的采样频率远高于相机，因此在两相邻图像帧之间存在多个 IMU 测量数据，文献[12]中给出了 IMU 预积分的详细推导过程. 参考预积分理论，通过对两相邻图像帧 i 和 $i+1$ 之间的多个 IMU 测量值积分，可以得到 Δt 时间段内 IMU 的状态增量，包括位移 p，速度 v 以及旋转矩阵对应的四元数 q：

$$\begin{aligned} R_{S_iW}p_{i+1} &= R_{S_iW}\left(p_i + v_i\Delta t - \frac{1}{2}g^W\Delta t^2\right) + p^i_{i+1} \\ R_{S_iW}v_{i+1} &= R_{S_iW}(v_i - g^W\Delta t) + v^i_{i+1} \\ q_{S_iW} \otimes q_{WS_{i+1}} &= q^i_{i+1} \end{aligned} \tag{1}$$

其中，R_{SW} 表示从世界坐标系到 IMU 坐标系的变换矩阵，g^W 是世界坐标系下的重力分量.

需要注意的是，传感器零偏 b_g 和 b_a 在两图像帧之间也在缓慢地发生变化. 为了保持较快的处理速度，在下个时刻对其求一阶泰勒展开得到一个近似偏差值，则传感器偏差改变时测量值的更新式为

$$\begin{aligned}\tilde{p}^{i}_{i+1} &= p^{i}_{i+1} + \frac{p^{i}_{i+1}}{\partial b_g}\delta b_{gi} + \frac{p^{i}_{i+1}}{\partial b_a}\delta b_{ai}\\ \tilde{v}^{i}_{i+1} &= v^{i}_{i+1} + \frac{v^{i}_{i+1}}{\partial b_g}\delta b_{gi} + \frac{v^{i}_{i+1}}{\partial b_a}\delta b_{ai}\\ \tilde{q}^{i}_{i+1} &= q^{i}_{i+1} \otimes Q\left(\frac{\partial \alpha^{i}_{i+1}}{\partial b_g}\delta b_{gi}\right)\end{aligned} \tag{2}$$

其中，$\partial\alpha_i$ 是 q_i 对应的欧拉角，$\left\{\frac{\partial p}{\partial b}, \frac{\partial v}{\partial b}, \frac{\partial \alpha}{\partial b}\right\}$是相关状态量的变化梯度，可由 Δt 之间的 IMU 测量值计算求出.

1.4 特征提取与跟踪

本文使用 FAST 角点快速检测算法对左右相机采集到的图像提取关键点，为了避免提取的关键点过于集中，除了非极大值抑制以外，还采用均匀化提取的策略. 首先，设置阈值 Td 对视觉帧图像粗提取一批角点，然后将该图像分割成固定大小的 G 个网格，并设置每个网格中的最大角点数量 M，使用 Harris 角点评价准则计算每个网格中角点的评价得分，若某网格中的角点数量大于 M，则选取评分高的前 M 个角点作为特征关键点，若网格内的角点数量极小，则忽略这片特征.

实验对比结果如图 2 所示. 左图是传统 FAST 特征检测算法提取结果，右图是结合均匀化策略的特征提取结果. 通过对比可以看出，右图中的角点提取数量明显少于传统 FAST 特征检测算法，且角点分布较为均匀；左图中门框、三脚架以及桌子上的角点提取过于密集，而在右图中得到了明显改善，地面和墙上的部分零星角点被舍弃. 表 1 记录了两种角点检测的耗时情况，从表中数据来看，改进后算法的角点检测数量远低于传统 FAST 检测算法，虽然平均耗时有所增加，但总耗时与传统 FAST 算法差距不大，这种策略可以在很大程度上避免由于特征提取过于集中而造成的误匹配，还能减少特征提取的数量，提高特征匹配的计算效率.

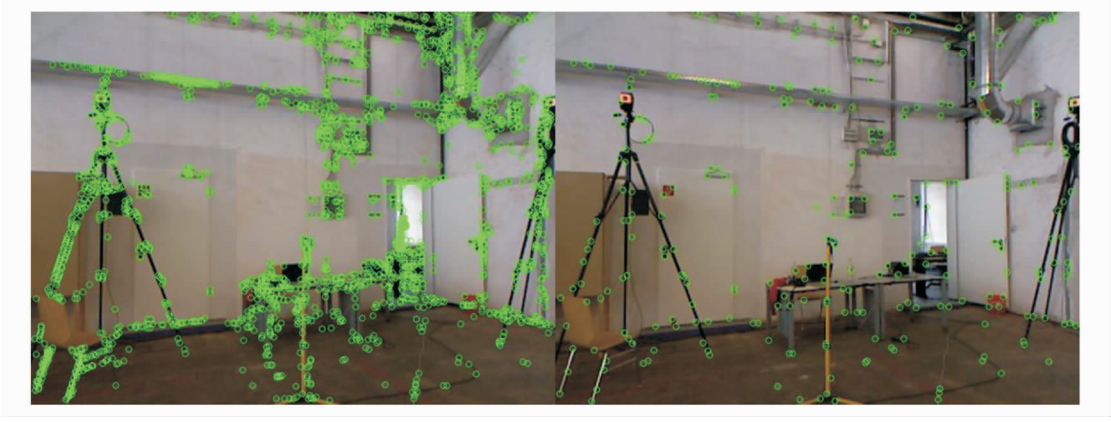

图 2 特征提取结果

表 1 角点检测耗时(ms)

检测算法	角点数量	总耗时	平均耗时
传统 FAST	1 989	73.18	0.037
FAST 角点均匀化处理	237	68.52	0.289

与采用单一传感器的纯视觉里程计不同，本文算法结合了两种传感器的信息来实现对图像特征的跟踪. 针对两相邻图像帧，通过 IMU 预积分可以为特征跟踪提供一个良好的先验位姿，如下式：

$$T_{WC} = T_{WS} \cdot T_{SC} \tag{3}$$

其中，T_{WS} 表示通过 IMU 预积分求得的相邻图像帧之间的变换矩阵，T_{SC} 表示相机与 IMU 之间的外参矩阵.

在此基础上，使用逆光流法对相邻图像帧中的关键点进行跟踪，如图 3 所示. 假设像素 a 是上一图像帧中的稳定角点，利用光流跟踪可以获得下一图像帧上对应的像素 b，然后再反向追踪像素 b 在上一帧中的对应像素 c，由于噪声的影响，a 和 c 的位置不可能完全重合，通过比较两者之间的像素距离 d 与设定阈值

的大小，判断此次跟踪的正确性. 然而在多数的场景中，载体相机采集的图像无法完全满足光流法的条件，因此会出现像素点跟踪错误甚至跟踪失败的情况，本文使用 RANSAC(随机抽样一致) 算法对光流跟踪结果进行过滤[13]，以消除异常点.

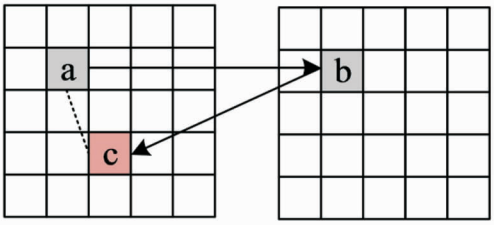

图 3　逆光流跟踪算法原理

1.5　系统初始化

系统初始化是视觉和惯性信息之间松耦合的过程. 首先需要对视觉部分进行初始化，得到三维空间中一系列的相机位姿，同时 IMU 持续进行预积分，获取图像帧之间的预积分量，之后把视觉初始化得到的相机位姿和预积分结果进行对齐，估计出初始化过程中图像帧速度、陀螺仪的零偏以及重力矢量等.

本文系统中，IMU 与双目相机可以看作一个整体，所以相机从 i 到 $i+1$ 帧图像的旋转变换应与 IMU 在对应时间内的旋转增量相同，忽略传感器的观测噪声将两者的结果对齐，通过最小化相对旋转误差得到陀螺仪零偏的初始值：

$$b_g = \arg\min_{b_g} \sum_{i=1}^{N-1} \left\| \log\left(\left(\Delta R_{i,\,i+1} \exp\left(\frac{\partial R_{i,\,i+1}}{\partial b_g} b_g \right) \right)^T R_W^{S(i+1)} R_{Si}^W \right) \right\|^2 \tag{4}$$

其中，N 是图像帧数目，$\Delta R_{i,\,i+1}$ 为相邻图像帧之间的陀螺仪积分，R_S^W 为 IMU 坐标系到世界坐标系的旋转矩阵.

针对要求解重力矢量以及各图像帧时刻的速度，设目标状态量为 $\chi = [v_1, \cdots, v_i \cdots, v_n, g]$，其中 v_i 代表第 i 帧图像对应的速度，g 代表重力矢量. 由预积分公式可得：

$$\begin{bmatrix} 1 & -1 & -\Delta t_{i,\,i+1} \\ -\Delta t_{i,\,i+1} & 0 & -\frac{1}{2}\Delta t_{i,\,i+1}^2 \end{bmatrix} \begin{bmatrix} v_{i+1} \\ v_i \\ g \end{bmatrix} = \begin{bmatrix} R_i v_{i+1}^i \\ R_i p_{i+1}^i + p_i - p_{i+1} \end{bmatrix} \tag{5}$$

其中，v_{i+1}^i 和 p_{i+1}^i 可以通过预积分求出，p_i 和 p_{i+1} 可以由视觉初始化求出.

根据多帧图像间的上述约束关系联立方程组，通过 SVD 分解即可得到初始化过程中的速度和重力矢量的估计值. 此外，根据重力矢量相对世界坐标系 z 轴的旋转，可以将系统初始位姿与世界坐标系对齐，并以此位姿作为参考构建初始地图，为之后的运动估计提供精确的初值. 为了提高初始地图的精度，采用以下 3 个步骤构造和维护地图点：

1) 构造地图点

地图点的位置可以用关键点的归一化坐标 z 和深度 d 表示为向量 $d \cdot z$，其计算过程如下所示：

$$\begin{bmatrix} z_1^T z_1 & z_1^T R_{12} z_2 \\ -z_1^T R_{12} z_2 & -z_2^T R_{12}^T R_{12} z_2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} z_1^T t_{12} \\ z_2^T R_{12}^T t_{12} \end{bmatrix} \tag{6}$$

整理后可得：

$$d \cdot z = \frac{d_1 \cdot z_1 + R_{12}(d_2 \cdot z_2) + t_{12}}{2} \tag{7}$$

其中，z_1 和 z_2 表示两个相匹配关键点的归一化坐标，R_{12} 和 t_{12} 表示两个图像帧之间的旋转矩阵和平移向量.

此外，为了验证初步构建地图点的有效性，本文使用卡方检验标准来评估其位置向量. 位置向量的偏差定义为：$e = dz - d_1 z_1$，测量的距离定义为：$D = dz - t_{12}/2$，地图点的位置偏差应限制在测量距离的范围之内，其卡方标准可以定义如下：

$$e_{chi2} = \frac{e_L^T e_L}{(D\sigma)^T D\sigma} \tag{8}$$

其中，σ 是对应关键点的标准方差. 如果 e_{chi2} 大于预定阈值，则表示估计偏差很大或该地标点非常接近相机，这两种情况下都会导致初始地图一定程度的失真.

2) 深度更新

本文使用多对具有丰富特征的连续图像帧来构造初始地图，因此在初始化的过程中可以得到同一地图点的多个深度值，每个地图点的深度应不断进行更新，使得估计结果的不确定性逐渐变小. 根据文献[14]中的理论，本文使用高斯+均匀混合分布建模，当出现新的深度估计时，使用递归贝叶斯方法更新此混合分布中的参数[15]，使得期望值更接近真实深度，如公式(9)所示：

$$p(d)=\rho_N(d)N(d\mid d_{true},\sigma^2)+\rho_Z(d)Z(d\mid d_{\min},d_{\max}) \tag{9}$$

3) *BA* 优化

在创建足够多的地图点之后，先对初始地图中的所有图像帧进行一次分类，若某图像帧中的可视地图点数量达到设定阈值，便将其标记为有效图像帧，反之，则标记为无效图像帧. 然后，联合初始地图中所有有效图像帧的位姿及其可见地图点的三维坐标构建重投影误差项，其目标函数可以定义为

$$e_y=\sum_{i=1}^{M}\sum_{l=1}^{L}\frac{1}{2}\left\|(u_{il}-h(T_i,z_l))^T\Omega_l(u_{il}-h(T_i,z_l))\right\|^2 \tag{10}$$

其中，$e_x=u_{ij}-h(T_i,z_j)$ 表示相机在位姿 T_i 处对地图点 l 的观测误差，h 是相机投影模型，Ω 是相关地图点的信息矩阵.

通过最小化此目标函数可以获得优化的图像帧位姿以及地图点坐标，由于先对参与优化的图像帧进行了一次筛选，提高了优化的效率和精度，为后续运动估计提供了良好的初值.

1.6 视觉惯性联合优化

根据相机投影模型可以将任何可见的3D地标点投影到像素坐标系中. 左侧相机的重投影误差可表示为

$$e_l=\begin{bmatrix}u_l\\v_l\end{bmatrix}-\begin{bmatrix}f\cdot x_C/z_C+c_x\\f\cdot y_C/z_C+c_y\end{bmatrix} \tag{11}$$

其中，u_l 和 v_l 是左图像帧中的观测坐标，f,c_x 和 c_y 是相机内参.

可以看到，如果按照传统的方式构建左右相机的重投影误差，会产生一个四维的误差向量，这无疑增加了计算的复杂度. 本文使用一种简化的视觉模型来解决这个问题：假定双目相机已经严格校正，则左右两幅图像中的投影点的纵坐标是相同的，某一时刻下的左右图像匹配后，可获取右图像中对应特征点横坐标的观测值[16]，根据双目相机投影模型的特点，可以计算得到右图像中地标点投影位置的横坐标，然后将其与观测值相比较，便得到右侧相机的重投影误差：

$$e_r=u_r-(f\cdot(x_C-b)/z_C+c_x) \tag{12}$$

其中，u_r 是右图像中的横坐标观测值，b 是双目相机的基线. 此时，可将双目重投影误差用一个三维向量表示：

$$e_c=\begin{bmatrix}u_l\\v_l\\u_r\end{bmatrix}-\begin{bmatrix}f\cdot x_C/z_C+c_x\\f\cdot y_C/z_C+c_y\\f\cdot(x_C-b)/z_C+c_x\end{bmatrix} \tag{13}$$

IMU的测量误差主要分为两部分，包括预积分项的 R,v,p 的误差以及 b_g 和 b_a 的误差. 根据预积分模型，将各误差分量表示为

$$e_p=R_{S_iW}\left(p_i-p_j+v_i\Delta t-\frac{1}{2}g^W\Delta t^2\right)+p_j^i+\frac{\partial p}{\partial b_g}\delta b_{gi}+\frac{\partial p}{\partial b_a}\delta b_{ai} \tag{14}$$

$$e_v=R_{S_iW}(v_i-v_j-g^W\Delta t)+v_j^i+\frac{\partial p}{\partial b_g}\delta b_{gi}+\frac{\partial p}{\partial b_a}\delta b_{ai} \tag{15}$$

$$e_\alpha=\left(2q_j^i\otimes Q\left(\frac{\partial\alpha}{\partial b_g}\delta b_{gi}\right)\otimes(q_{WS_j}^{-1}\otimes q_{WS_i})\right) \tag{16}$$

$$e_b=[b_{gi},b_{ai}]^T-[b_{gj},b_{aj}]^T \tag{17}$$

因此，IMU误差项可表示为

$$e_S=[e_p^T,e_v^T,e_a^T,e_b^T]^T \tag{18}$$

基于上述的视觉重投影误差与IMU传播误差，构建联合优化的模型，以求解更准确的相机位姿和地图

点三维信息，构建视觉惯性联合优化的目标函数：

$$E(x)=\sum_{i\in M}\sum_{l\in L(i)}e_C^{l,\,iT}\Omega_C^{l,\,i}e_C^{l,\,i}+\sum_{i\in M}e_S^{iT}\Omega_S^{i}e_S^{i} \tag{19}$$

其中，M 是当前窗口中的关键帧集合，$L(i)$ 是两个相机在第 i 帧中同时观察到的地图点集合，Ω_C 与 Ω_S 分别代表视觉信息矩阵和惯性信息矩阵，$e_C^T\Omega_C e_C$ 和 $e_S^T\Omega_S e_S$ 表示窗口中对应时刻的视觉误差和 IMU 误差的二次形式.

1.7　边缘化

在不改变一致性的前提下，对滑动窗口内最旧的关键帧状态边缘化处理，可以避免丢失该关键帧包含的关联信息，有效地保留历史信息对窗口内的其他状态的影响，该过程的本质是一个最小二乘问题，本文用高斯牛顿迭代法求解，定义如下：

$$H\delta x=b \tag{20}$$

在边缘化过程中，不直接删除被边缘化的状态量以及与其相关的地标点，否则会减少优化过程中的约束. 假定 δx_a 是要被边缘化的状态量，δx_b 是需要被保留的约束，根据舒尔补公式对优化量进行边缘化操作，该过程可简化为

$$\begin{bmatrix}H_{aa}&H_{ab}\\H_{ba}&H_{bb}\end{bmatrix}\begin{bmatrix}\delta x_a\\\delta x_b\end{bmatrix}=\begin{bmatrix}b_a\\b_b\end{bmatrix} \tag{21}$$

$$\begin{bmatrix}I&0\\-H_{ba}H_{aa}^{-1}&I\end{bmatrix}\begin{bmatrix}H_{aa}&H_{ab}\\H_{ba}&H_{bb}\end{bmatrix}\begin{bmatrix}\delta x_a\\\delta x_b\end{bmatrix}=\begin{bmatrix}I&0\\-H_{ba}H_{aa}^{-1}&I\end{bmatrix}\begin{bmatrix}b_a\\b_b\end{bmatrix} \tag{22}$$

整理得：

$$(H_{bb}-H_{ba}H_{aa}^{-1}H_{ab})\delta x_b=b_b-H_{ba}H_{aa}^{-1}b_a \tag{23}$$

其中，$H_{bb}-H_{ba}H_{aa}^{-1}H_{ab}$ 和 $b_b-H_{ba}H_{aa}^{-1}b_a$ 表示被边缘化的 H 矩阵，状态量 δx_a 被边缘化.

2　实验验证及分析

本文通过以下实验来评估所提出双目视觉惯性里程计的性能. 首先，使用 EuRoC MAV 数据集对算法性能进行评估，然后将所提出算法与 ORB-SLAM2 以及 VINS-Fusion 算法进行比较分析. 实验所用笔记本配置为六核 i5-8400 处理器，满载频率为 3.8 GHz.

EuRoC MAV 数据集由安装在微型飞行器上的各种传感器收集而得到，该数据集分为 3 个系列的飞行场景，包含一系列由简单到复杂的运动序列. 选取 EuRoC MAV 数据集中较有代表性的 MH_01，MH_04，V2_01 以及 V2_03 运动序列作为实验场景，本文算法的运行结果如图 4—图 7 所示：

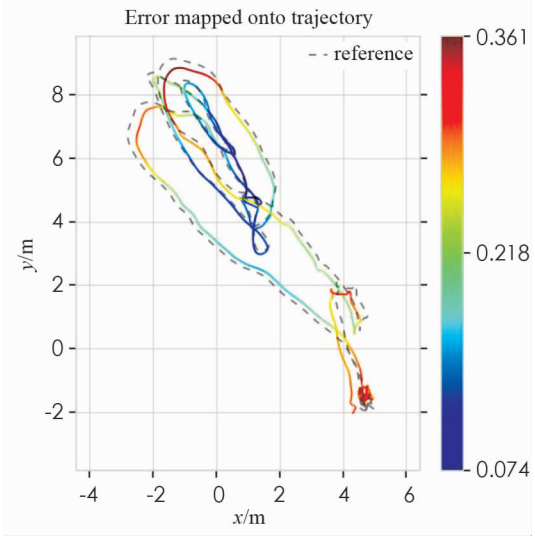

图 4　MH_01 数据集中里程计定位效果

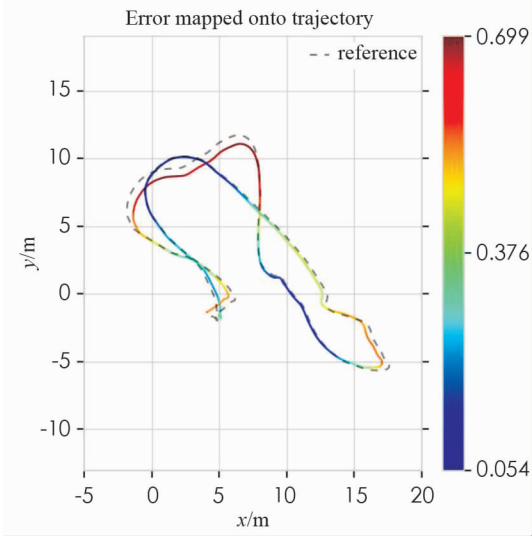

图 5　MH_04 数据集中里程计定位效果

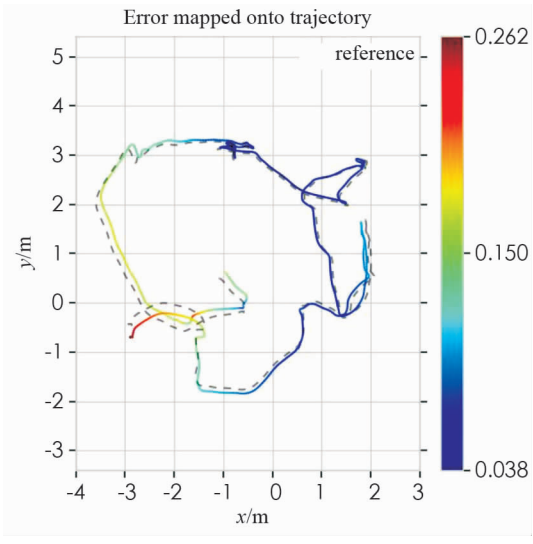

图6 V2_01数据集中里程计定位效果

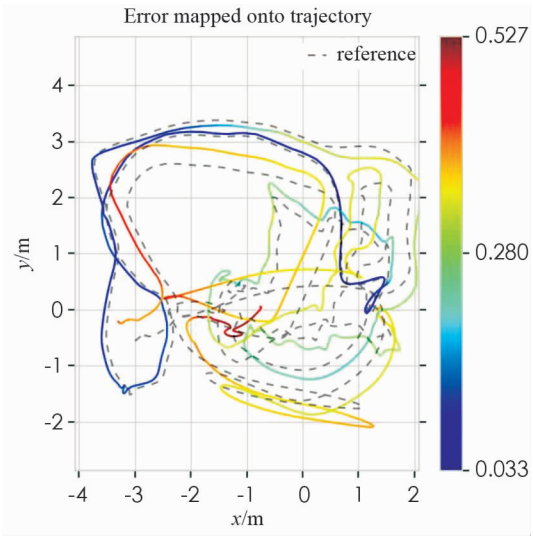

图7 V2_03数据集中算法的定位误差

图中显示了本文算法在所选数据集中的运动轨迹和定位误差. 其中，MH_01序列中的定位误差保持在0.361 m内，V2_01序列中的定位误差保持在0.262 m内，这表明本文算法在运动相对平稳，光照变化较小的场景中具有较高的定位精度. 在无人机运动较快，光照条件较差的MH_04序列与V2_03序列中，本文算法的定位误差分别保持在0.699 m和0.527 m内，且保持在一个相对平稳的状态，表明本文提出双目惯性里程计算法在恶劣场景下的鲁棒性较好.

为了进一步验证算法性能，选择ORB-SLAM2和VINS-Fusion算法同本文算法作比较，对应的均方根误差(RMSE)如表2所示，需要注意的是，实验时关闭了VINS-Fusion算法中的回环检测功能. 从表中数据对比可以看出，ORB-SLAM2在M系列以及V1_01,V1_02和V2_01数据集中的RMSE值是最小的，这是基于特征法的纯视觉里程计的优势：适用于光照条件较好的大场景运动估计；而在光照条件差、运动较快的场景中，另外两种视觉惯性里程计的算法表现更好，这也是两种传感器互补的优势. 与VINS-Fusion算法对比，本文算法在MH_03和V1_02两个数据集中的表现更好，其运动轨迹对比如图8—图9所示，其中，虚线代表该运动序列的真实参考轨迹，橘色的线代表VINS-Fusion算法在该序列下的运动轨迹，而蓝色的线代表本文算法在该序列下的运动轨迹，对比两者的运动轨迹可以看出，本文算法在两个运动序列中定位精度的表现都优于VINS-Fusion算法，尤其是在V1_02这种较为复杂的场景中. 在V1_03,V2_03的数据集中，本文算法获得的RMSE值最小，这是由于本文算法提供了更精确的先验信息并改进了光流匹配过程，为之后的位姿优化过程提供了可靠的基础，而且以强约束关键帧作为滑动窗口的优化项，进一步提高了位姿估计的精度.

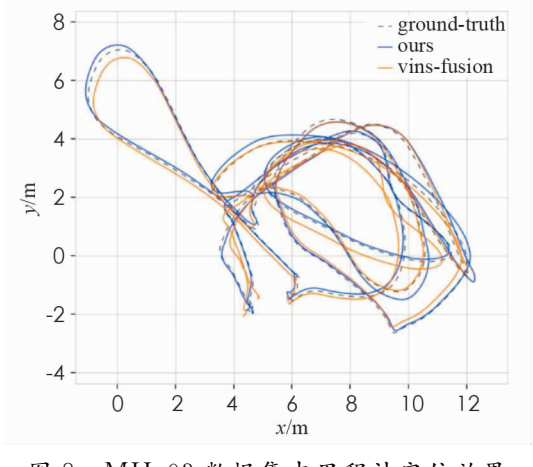

图8 MH_03数据集中里程计定位效果

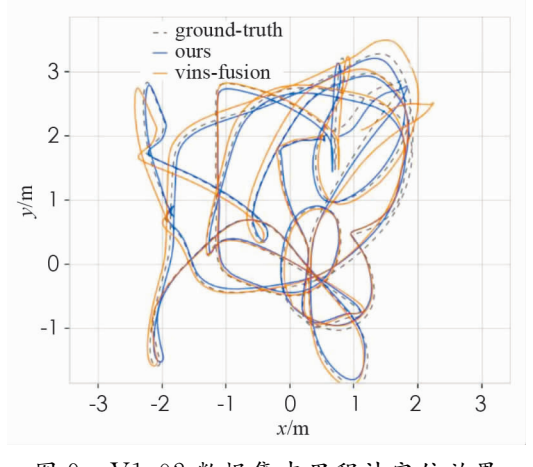

图9 V1_02数据集中里程计定位效果

表 2　各算法的 RMSE 对比

数据集	MH_01	MH_03	MH_05	V1_01	V1_02	V1_03	V2_01	V2_02	V2_03
ORB-SLAM2	0.036	0.071	0.064	0.031	0.065	0.092	0.071	0.161	0.265
VINS-Fusion	0.251	0.289	0.318	0.152	0.113	0.120	0.120	0.086	0.216
Ours	0.174	0.167	0.262	0.158	0.084	0.077	0.097	0.093	0.159

为了验证算法的实时性，表 3 记录了本文算法和 VINS-Fusion 算法在数据集上的 CPU 负载和运行帧率. EuRoC MAV 数据集中所使用相机的帧率为 20 Hz，所以若某算法的运行帧率低于 20 Hz，整个系统的数据处理会出现明显的延迟. 从表中数据对比可以看出，相比于 VINS-Fusion，本文的算法降低了约 8% CPU 负载，且运行帧率更高.

表 3　CPU 负载和运行帧率的统计

数据集	CPU 负载/%		运行帧率/Hz	
	VINS-Fusion	本文算法	VINS-Fusion	本文算法
MH_01_easy	54.14	47.25	28.62	37.02
MH_02_easy	47.78	41.72	28.13	36.83
MH_03_medium	45.84	42.52	27.14	35.31
MH_04_difficult	50.45	49.87	27.98	36.47
MH_05_difficult	53.27	50.84	26.85	35.96
V1_01_easy	55.76	48.76	26.71	34.72
V1_02_medium	49.86	45.25	27.26	34.36
V1_03_difficult	51.69	49.69	28.46	36.75
V2_01_easy	52.33	49.78	27.93	35.61
V2_02_medium	51.09	46.26	26.76	34.63
V2_03_difficult	49.84	44.65	26.84	41.38
Average	51.01	49.96	27.52	36.28

3　结　　论

本文提出了一种基于半直接法的双目视觉惯性里程计算法，在系统初始化阶段，结合惯性测量数据和双目相机的图像信息构建高精度的初始化地图，为后端的位姿优化提供良好的初值；使用简化的双目视觉模型构建重投影误差，减少了对右相机图像中多余的测量值的计算；在滑动窗口边优化的过程中选择性地剔除部分图像帧信息，确保在优化过程中拥有足够参考信息的同时尽可能地减少优化计算量；与 ORB-SLAM2 和 VINS-Fusion 算法的实验对比表明，本文算法在定位精度上已达到主流的视觉里程计的水平，并且在计算效率方面有一定程度的提高.

参考文献：

[1] 孙永全，田红丽. 视觉惯性 SLAM 综述 [J]. 计算机应用研究，2019，36(12)：3530-3533，3552.

[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] 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. IEEE，2007：225-234.

[4] MUR-ARTAL R，MONTIEL J M M，TARDOS J D. ORB-SLAM：A Versatile and Accurate Monocular SLAM System [J]. IEEE Transactions on Robotics，2015，31(5)：1147-1163.

[5] ENGLE J，SCHöPS T，CREMERS D. LSD-SLAM：Large-Scale Direct Monocular SLAM [C]//European Conference on Computer Vision. Springer，Cham，2014：834-849.

[6] ENGEL J，KOLTUN V，CREMERS D. Direct Sparse Odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2018，40(3)：611-625.

[7] 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，Hongkong，China. IEEE，2014：

15-22.
[8] QIN T, LI P L, SHEN S J. Vins-Mono: A Robust and Versatile Monocular Visual-inertial State Estimator [J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.
[9] SUN K, MOHTA K, PFROMMER B, et al. Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight [J]. IEEE Robotics and Automation Letters, 2018, 3(2): 965-972.
[10] BLOESCH M, BURRI M, OMARI S, et al. Iterated Extended Kalman Filter based Visual-Inertial Odometry using Direct Photometric Feedback [J]. The International Journal of Robotics Research, 2017, 36(10): 1053-1072.
[11] MOURIKIS A I, ROUMELIOTIS S I. A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation [C]//Proceedings 2007 IEEE International Conference on Robotics and Automation. April 10-14, 2007, Roma, Italy. IEEE, 2007: 3565-3572.
[12] LEUTENEGGER S, FURGALE P, RABAUD V, et al. Keyframe-based Visual-inertial SLAM Using Nonlinear Optimization [J]. Proceedings of Robotis Science and Systems (RSS) 2013, 2013.
[13] WANG G, SUN X, SHANG Y, et al. Two-View Geometry Estimation Using RANSAC With Locality Preserving Constraint [J]. IEEE Access, 2020, 8: 7267-7279.
[14] VOGIATZIS G, HERNáNDEZ C. Video-Based, Real-Time Multi-View Stereo [J]. Image and Vision Computing, 2011, 29(7): 434-441.
[15] 李 勇, 刘鹤飞, 王 坤, 等. 隐马尔科夫多元线性回归模型中未知隐状态个数的贝叶斯模型选择 [J]. 西南师范大学学报(自然科学版), 2020, 45(7): 11-17.
[16] 许 翊, 刘学军. 计算机双目视觉中的动态规划立体匹配算法研究 [J]. 西南师范大学学报(自然科学版), 2020, 45(9): 118-123.

A Stereo Visual Odometry Aided by IMU based on Semi-direct Method

CHONG Yi-fan, JI Jie, GONG Ming-qian, CHEN Qiong-hong

College of Engineering and Technology, Southwest University, Chongqing 400715, China

Abstract: A stereo visual inertial odometry based on semi-direct method has been proposed to improve the poor real-time performance and robustness of visual odometry based on feature-based method. The inertial measurement data is used as a priori in the tracking thread, and the reverse optical flow method is used to track the homogenized feature key points to improve the speed and robustness of feature matching. A high-precision initialization map is constructed to provide a accurate initial value for the subsequent motion estimation. The joint optimization model, which is constructed by combining the reprojection error which is constructed by a simplified stereovision model and IMU error, is solved by nonlinear-optimization in the sliding window. The experimental results show that positioning accuracy of the proposed algorithm reaches the level of the mainstream algorithm. Compared with the VINS-Fusion, our algorithm in this paper has lower CPU load and higher running frequency.

Key words: real-time performance; robustness; simplified stereo visual model; initialization map; nonlinear-optimization

责任编辑 汤振金