

引文格式: 丁晓, 晏玉祥, 张永建, 等. 一种改进点线特征融合的双目视觉惯性定位算法[J]. 航空制造技术, 2023, 66(10): 85–92.

DING Xiao, YAN Yuxiang, ZHANG Yongjian, et al. An improved binocular visual inertial navigation and positioning algorithm based on point–line fusion[J]. Aeronautical Manufacturing Technology, 2023, 66(10): 85–92.

# 一种改进点线特征融合的双目视觉惯性定位算法\*

丁晓<sup>1</sup>, 晏玉祥<sup>2</sup>, 张永建<sup>1</sup>, 兰卫旗<sup>2</sup>, 白晓亮<sup>2</sup>

(1. 航空工业成都飞机工业(集团)有限责任公司, 成都 610092;

2. 西北工业大学, 西安 710072)

**[摘要]** 在复杂产品装配过程中, 对增强现实设备定位是实现虚拟引导信息与装配现场实时融合的核心。传统基于标签或预先构建离线装配基体模型的定位方式, 存在装配任务与视觉定位兼容性低, 单一视觉定位鲁棒性和稳定性差的问题。利用多传感器融合的视觉惯性导航方式进行定位, 可提高定位的精度和鲁棒性, 有效提高复杂产品装配质量和效率。本文提出了一种基于点线特征融合的双目视觉惯性定位算法, 对双目相机和惯性测量单元(Inertial measurement unit, IMU)进行联合标定, 并通过 ORB 和 LSD 线端融合的点线特征提取匹配, 采取视觉和 IMU 紧耦合的方式, 建立基于点线特征的视觉惯性融合的位姿误差融合模型。对比 VINS-Fusion、PL-VIO 和本文改进 IPL-VIO 算法, 结果表明 IPL-VIO 算法在结构化场景下的绝对位移和旋转误差比原算法误差更小, 同时结构化场景信息更加丰富, 能够应用在弱纹理的 AR 装配现场中, 为增强现实辅助装配平台提供稳定可靠的位姿数据。

**关键词:** 增强现实(AR); 双目视觉; 惯性测量单元(IMU); 点线特征; 定位

## An Improved Binocular Visual Inertial Navigation and Positioning Algorithm Based on Point–Line Fusion

DING Xiao<sup>1</sup>, YAN Yuxiang<sup>2</sup>, ZHANG Yongjian<sup>1</sup>, LAN Weiqi<sup>2</sup>, BAI Xiaoliang<sup>2</sup>

(1. AVIC Chengdu Aircraft Industrial (Group) Co., Ltd., Chengdu 610092, China;

2. Northwestern Polytechnical University, Xi'an 710072, China)

**[ABSTRACT]** In the assembly process of complex products based on augmented reality, the location of augmented reality equipment is the core of the real-time integration of virtual guidance information and assembly site. The traditional positioning method based on label or pre-built offline assembly matrix model has the problems of low compatibility between assembly task and visual positioning, and poor robustness and stability of single visual positioning. The fusion of vision and inertial measurement unit (IMU) can improve the positioning accuracy and robustness, and effectively improve the assembly quality and efficiency of complex products. In this paper, a positioning algorithm based on the fusion of binocular vision and IMU is proposed. This algorithm extracts image features through point and line features. The binocular camera and IMU are jointly calibrated, and the point and line features are extracted and matched by ORB and LSD line end fusion. The pose error fusion model of visual-inertial fusion based on point and line features is established by means of tight coupling of vision and IMU. By comparing the experiments of VINS-Fusion, PL-VIO algorithms and IPL-VIO algorithm improved in this paper, the absolute displacement and rotation error of the IPL-VIO algorithm in structured scene is smaller than that of the original algorithm, and the structured scene information is more abundant, which can be applied in AR assembly field with weak texture, and provide stable and reliable pose data for augmented reality assembly platform.

**Keywords:** Augmented reality (AR); Binocular vision; Inertial measurement unit (IMU); Point and line features; Positioning

**DOI:** 10.16080/j.issn1671-833x.2023.10.085

\* 基金项目: 国防基础科研计划(×××2018205B021, ×××2018213A001)。

在航空航天等大型复杂产品装配过程中,装配工艺复杂<sup>[1]</sup>,由于时间和空间上的限制,目前装配工作大部分依赖操作人员对照工艺文件和现场环境展开装配,信息不集中且可视化程度不高,装配工作量约占整个产品研发和生产时间的50%<sup>[2]</sup>。利用增强现实(Augmented reality, AR)辅助装配<sup>[3-4]</sup>可以在一定程度上解决采用二维纸质工艺文件带来的劣势问题,给操作工人提供一种真实与虚拟引导信息无缝衔接的混合现实操作场景,能提高操作的代入感和真实感,满足产品装配质量的同时也可以节省时间和成本。

增强现实辅助装配依赖虚拟对象和真实场景的注册一致性,定位技术是实现虚拟引导信息与装配现场实时融合的核心。增强现实辅助装配通常通过识别场景标签来对增强现实设备进行定位,但要求保证标签实时出现在相机视野中,才能实现装配引导的效果。这种传统的视觉定位方式,在纹理少的区域无法工作,当运动过快或者剧烈晃动的时候,相机拍摄的图像由于运动而导致模糊;惯性测量单元(Inertial measurement unit, IMU)可以测量角速度和加速度,对于短时间内的快速运动具有较好的估计,但是在静止情况下测量值明显存在漂移。相比IMU,相机拍摄的照片不会存在漂移,使得相机在慢速运动之后的位姿估计依然有效<sup>[5]</sup>,但当图像发生变化之后,相机自身是难以感知运动的,而IMU能够感受到本体的运动信息,可减少动态物体的影响。由于IMU和相机具有明显的互补性,基于多传感器融合的位姿估计技术在增强现实方面得到了越来越多的应用<sup>[6]</sup>。

山东大学李民<sup>[7]</sup>采用IMU和视觉标签Apriltag融合的办法组合测量头盔的位姿,这种定位方式缺点在于光照对相机产生干扰使其检测不到标签。西安交通大学王崑等<sup>[8]</sup>在面向增强现实的算法中,提出一种基于互补滤波和无迹卡尔曼滤波的定位方案,该定位方法要求AR头盔实时处在摄像机视野中,且头盔标记点易受光照遮挡影响。融合了全球定位系统(Global positioning system, GPS)、IMU以及方向传感器的融合方法可有效降低累积误差。在AR应用中,为了与场景交互,即时定位与地图构建(Simultaneous localization and mapping, SLAM)算法必须提供鲁棒的带正确尺度信息的相机位姿估计,视觉惯性定位技术成为主流的发展趋势。传统的SLAM框架主要包括前端模块和后端模块。前端部分主要通过传感器来感知场景信息、估算位姿和速度等初始信息。根据处理方法的不同,后端接收前端的初始数据,可以分为基于滤波和基于优化两种方法<sup>[9]</sup>。

基于优化的方法使用了更多信息,随着计算机应用技术的发展,采用图优化处理SLAM的方法逐渐替代滤波的方法成为主流,这其中需要采用第三方库g2o<sup>[10]</sup>、

Ceres-Solver<sup>[11]</sup>、GTSAM<sup>[12]</sup>等来进行优化。图优化方法核心是将所有的点和边联合优化,对自身位姿进行全局最优处理。

Leutenegger等<sup>[13]</sup>提出了一种基于关键帧的非线性优化的视觉惯性融合SLAM系统,能够得到比较好的精度和鲁棒性。香港科技大学提出的VINS-Mono<sup>[6]</sup>系统在经典的视觉SLAM框架之上融合了相机和IMU,视觉前端是利用KLT<sup>[14]</sup>光流进行角点跟踪,后端是采用了滑动窗口来进行非线性优化。此后,为了避免VINS-Mono初始化不稳定的情况,提出了其双目版VINS-Fusion,根据双目立体获得深度信息的原理,可以实时提供稳定可靠的场景尺度信息。在视觉信息中,点特征是最常用的特征类型,但在弱纹理的情形下,采用点特征跟踪的方法可能会追踪不到足够的特征点而出现跟踪失败或者无法跟踪的情况<sup>[15]</sup>。沈冉<sup>[16]</sup>提出了一种点特征为主,线特征为辅的点-线融合的单目视觉惯性里程计方法,确保了系统在昏暗环境中提取跟踪的点线数据的稳定性,获得了较为鲁棒的定位效果,但其单目无法提供稳定可靠的深度场景信息。而针对结构化重复纹理场景,He等<sup>[17]</sup>提出一种使用点线特征融合的视觉前端、视觉和IMU紧耦合后端优化的PL-VIO算法,优化最小化误差函数,实现了较好的定位效果。但该算法不支持在线使用,且计算量大,不具备场景深度恢复功能。赵良玉等<sup>[18]</sup>提出了一种采用Shi-Tomasi角点提取点特征信息,同时增加快速直线检测(Fast line detector, FLD)提取线特征的点线特征融合的双目惯性SLAM算法,在保证定位精度的同时能够获得更为丰富的环境地图,具备较好的鲁棒性。其采用的Shi-Tomasi角点提取方法,可以获得较快的提取速度,但在使用KLT光流法实现特征点的跟踪时,可能会因为相机的剧烈运动而导致帧间匹配失败,追踪不够稳定。

针对复杂装配任务、弱纹理下强旋转导致特征容易丢失,增强现实设备持续稳定的跟踪定位非常困难的问题,本文提出了一种改进的PL-VIO<sup>[17]</sup>(Improved PL-VIO, IPL-VIO)算法,使用多传感器融合的位姿估计技术,通过视觉和IMU的定位方式,对相机进行内参标定并进行误差分析,提高系统采集数据的精确度,对图像进行特征分析,提取特征点和线,并采用PL-VIO算法为主体框架,针对点特征,采用ORB(Oriented fast and rotated brief)算法提取特征点,并改进光度不变假设,采用多层光流进行跟踪。与文献[18]中的Shi-Tomasi角点提取相比,ORB可以用来对图像中的关键点快速创建特征向量,并用这些特征向量来识别图像中的对象,因此帧间匹配时追踪更为稳定;针对线特征,通过改进的线段检测器(Line segment detector, LSD)提取算法,

建立合适的空间表征描述方法,同时进行双目左右相机之间的匹配计算位姿。采取视觉和IMU紧耦合的方式,建立基于点线特征的视觉和IMU融合的位姿误差融合模型,通过对比PL-VIO算法试验,本文提出的IPL-VIO算法在结构化场景下的绝对位移和旋转误差比原算法误差更小,同时结构化场景信息更加丰富,能够应用在弱纹理的AR装配现场中,为增强现实辅助装配平台提供稳定可靠的位姿数据,以提升多传感器融合的位姿估计精度,实现AR设备的实时定位。

## 1 多传感器标定

相机与IMU的标定包括在线标定和离线标定两种方式,在线标定是指在定位系统初始化过程中估计相机与IMU的外参,方便可靠,但对初始化过程要求比较高。考虑到装配场景及装配任务的复杂性,难以保证完成在线标定。因此,本文主要采取离线标定的方式,采用Rehder等<sup>[19]</sup>提出的集成多相机标定、相机和IMU标定的离线标定工具Kalibr,在估计出ZED Mini相机内参之后,可进行优化迭代操作。ZED Mini相机是立体相机公司Stereolabs研发的一款迷你版的立体声深度测绘摄像机,可以作为虚拟现实(Virtual reality, VR)头显HTC VIVE的增幅设备,为头显提供实时的拍摄视频传输、深度感知和环境映射,让VR头显转变为AR头显。在连续运动场景下标定相机系统的内外参,同时还可以标定相机和IMU的相对位姿,以及相机的重投影误差信息。

双目标定试验结果包含相机的左目和IMU单元的相对位置信息 $T_{IC}$ 和 $T_{CI}$ ,即

$$T_{IC} = \begin{bmatrix} -0.01359 & -0.99985 & 0.01059 & 0.06916 \\ 0.99961 & -0.01384 & -0.02384 & -0.12243 \\ 0.02398 & 0.01027 & 0.99965 & -0.32347 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$T_{CI} = \begin{bmatrix} -0.01359 & -0.99985 & 0.02395 & 0.13108 \\ -0.99985 & -0.01384 & 0.01027 & 0.07078 \\ 0.01060 & -0.0238 & 0.99965 & 0.319703 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

## 2 图像特征提取与跟踪

### 2.1 图像特征提取

特征是图像信息的另一种数字表达形式。在视觉SLAM中,常用的特征点提取算法有SIFT<sup>[20]</sup>、SURF<sup>[21]</sup>、FAST<sup>[22]</sup>、ORB<sup>[23]</sup>和BRIEF<sup>[24]</sup>等。SIFT特征点具有信息丰富、高速性、可扩展性强的特点,具有较强的图像特征提取鲁棒性,对旋转、尺度缩放、亮度变化保持

不变。SURF改善了SIFT计算复杂度高的缺点,对特征向量描述进行了改进。ORB是一种快速特征点提取和描述的算法,其结合了FAST特征点的检测方法和BRIEF的特征描述子的优点进行改进和优化。BRIEF特征点提取算法与SIFT的提取方法一致,不同之处在于其在特征点附近随机选取若干点,根据灰度值的大小,描述为二进制字符串,大大加快了描述子建立速度,但旋转不变性差。

检测匹配主要流程是首先对采集的照片进行特征点的提取,分别提取SIFT、SURF和BRIEF、ORB特征,并两两进行特征点的比较和匹配,同时对匹配的结果进行RANSAC<sup>[25]</sup>提纯计算,计算出其中的内点。统计匹配的特征点(即内点)所占总特征点总数的比例来计算内点率,再根据内点率及匹配耗时计算获得每个内点所需要的平均耗时(即时效比),为保证较高的内点率,调整合适的提取特征点数量。建立模型,即

$$\text{内点率} = \text{匹配内点数} / \text{特征点总数} \quad (1)$$

$$\text{时效比} = \text{内点率} / \text{耗时} \quad (2)$$

传统SLAM系统中均采用点特征来进行跟踪,而在装配环境中,由于工人头部转动导致图像模糊,无法完成定位。而装配场景存在结构化的线特征,因此采取LSD<sup>[26]</sup>直线检测方法提取装配场景中的点特征进行匹配。

5种检测匹配算法的时效比分析如图1所示。分析可知,SIFT和SURF检测的准确率高;ORB的检测匹配速度最快,时效比最高;LSD的匹配效果更好,但是检测比较耗时。单独从特征提取的角度来看,SIFT、ORB、BRIEF耗时较少。考虑到结构化的装配场景,为了保证在复杂狭小空间内工人头部转动和相机剧烈抖动的环境下追踪稳定,选用匹配效果较好的LSD算法进行直线特征提取,下文主要以ORB点特征和LSD线特征为例进行研究。

LSD算法用来表征场景的结构特征,但无参数调整的LSD算法在图像模糊或者有噪声时,易检测出过多的短线段特征,另外,检测时一条长线段容易分割成多

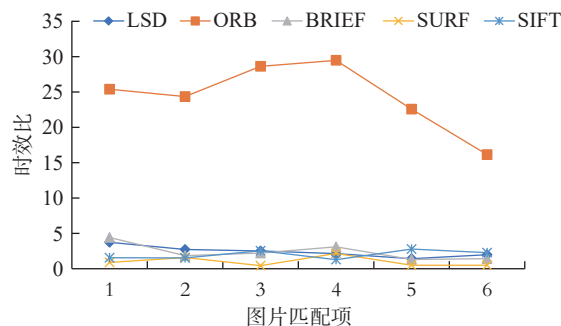


图1 5种检测匹配算法的时效比分析  
Fig.1 Time ratio analysis of five detection and matching algorithms

条线段,破坏了线特征的完整性,因此通过改进 LSD 算法进行线特征提取。首先通过图片大小设置最小长度阈值进行线段长度的初步筛选, LSD 检测线段的两个端点分为起点和终点,根据线段支持域的最小外接矩形的主方向、线段之间的距离、线段的最近两个端点之间的距离和线段的描述子之间的距离来判断两条线段是否在同一条直线上,如果满足阈值条件则合并两条线段,减少短线段的干扰。

### 2.2 图像特征跟踪

结构化装配场景下特征提取点的提取和描述子的计算非常耗时,为了克服这个缺陷,本研究采用直接法来进行特征点的跟踪,即根据像素灰度差异值来计算,而且装配场景存在明暗变化的情况,直接法可以直接工作<sup>[5]</sup>。本研究以文献[27]为基础,使用双目相机和 IMU 进行耦合。在装配场景中,工人头部在转动时,利用 IMU 计算头盔空间位姿,从而预测图片特征点的位姿,再利用非线性优化(如滑动窗口优化)来估计位姿。在特征点匹配完成之后,利用随机采样一致性(Random sample consensus, RANSAC)的方法进行异常匹配的移除。

本研究使用 Lucas-Kanade 光流法进行特征跟踪,由于场景中的亮度是改变的,因此使用参数  $\alpha$  和  $\beta$  来校正两个图像间的光度变化<sup>[28]</sup>,图像  $I_k$  和  $I_{k+1}$  之间的光度残差  $r_u^1(\xi)$  为

$$r_u^1(\xi) = \alpha I_k(u) + \beta - I_{k+1}(u') \quad (3)$$

$$\alpha = \frac{\sum_{u \in \Omega_L} I_1^1(u) I_2^1(u')}{\sum_{u \in \Omega_L} I_2^1(u') I_2^1(u')} \quad (4)$$

$$\beta = \frac{1}{|\Omega_L|} \sum_i (I_1^1(u') - \alpha I_2^1(u)) \quad (5)$$

式中,  $\Omega_L$  是内点的集合;  $\xi$  表示从  $I_k$  到  $I_{k+1}$  变换矩阵;  $u$  为像素坐标;  $u'$  为图像中  $u$  对应点像素坐标。

### 3 基于视觉和惯性融合的位姿误差融合模型

为了在 AR 装配过程中完成真实与虚拟装配场景的融合,需要计算 AR 设备的位姿估计,位姿主要包含相机的旋转和平移变换,其 6 维度的状态变量  $x$  可以表示为

$$x^6 = [R, p] \quad (6)$$

式中,  $R$  和  $p$  代表相机的旋转和位移。当相机运动起来后,场景中特征点比较多,特征点的管理较为复杂,为了更好实现位姿的估计,需要参考当前时刻周围的位姿,由于地图特征可以通过边缘化的方法转换为位姿间的约束,简化为对位姿序列的估计<sup>[29]</sup>。可以用位姿图的方式来直接描述。纯视觉里程计位姿估计因子图中的节点对应相机在不同时刻的位置和姿态,而边则描述了

位姿与位姿间的空间约束关系,这种约束关系通过里程计或观测信息的配准得到。

在基于点线融合的视觉里程计局部地图中,相机观测到视野中的点特征和线特征,根据投影方程建立点和线特征的误差模型,对优化的  $R$  和  $p$  进行状态约束。在对 IMU 的参数标定试验中,通过对 IMU 的零偏变换曲线分析可以得出零偏是一个动态量,其满足一阶马尔科夫的性质。本文采取一种紧耦合的方式使用相机的位姿、速度和 IMU 的零偏构建目标方程进行状态估计。系统的状态变量  $x$  共 15 维,可以表示为

$$x^{15} = [R, p, v, b_a, b_g] \quad (7)$$

式中,  $R$ 、 $p$  和  $v$  分别是相机的旋转、平移和速度;  $b_a$  和  $b_g$  分别是 IMU 的加速度计和陀螺仪的零偏。为了同时对状态变量进行优化,采用捆绑调整<sup>[30]</sup>(Bundle adjustment, BA)的方法,同时本研究根据固有的核心 BA 算法,将原有的 BA 问题分解成若干因子图,如图 2 所示,在因子图中,各个子图具有各自的局部坐标系并且可以进行优化。

在图 2 所示的视觉惯性里程计的位姿估计图中,首先计算出每一帧的位姿、速度和零偏。然后将位姿估计问题转化为联合优化问题,优化状态变量  $x$  主要包括相机的位姿、旋转速度以及 IMU 的零偏,通过最小化特征点和 IMU 的误差来优化当前帧的位姿,误差方程可以表示为

$$x = \arg \min_x (\sum r_x^1(\xi) + E_{IMU}(i, j)) \quad (8)$$

式中,  $\sum r_x^1(\xi)$  表示相机图像的光度残差;  $E_{IMU}(i, j)$  表示 IMU 的误差方程。

而针对线特征重投影误差<sup>[31]</sup>,图优化问题的本质是找到一个最优的状态值  $p$ ,如果利用测量协方差矩阵  $\Sigma_z$  的多元正态分布对似然分布进行建模,则状态值  $p^*$  可以表示为

$$p^* = \arg \min_p [\varepsilon_z^T \Sigma_z^{-1} \varepsilon_z] \quad (9)$$

式中,  $Z$  为图像中线特征测量值;  $\varepsilon_z$  为线特征的投影预

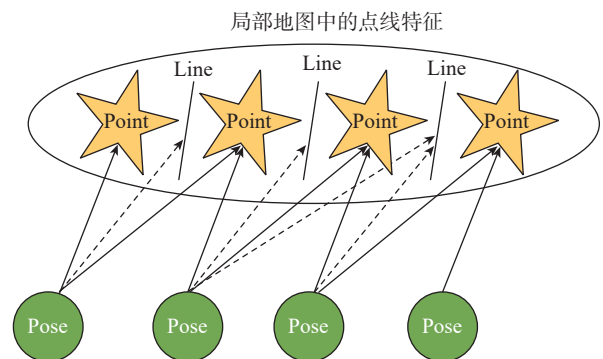


图 2 点线融合的视觉惯性里程计位姿估计因子图  
Fig.2 Position and pose estimation factor diagram of visual inertial odometer based on point and line fusion

测误差。

综上所述,最小化特征点、线特征和 IMU 的误差方程可以表示为

$$\mathbf{x} = \arg \min_{\mathbf{x}} \left( \sum r_x^l(\xi) + \sum_{\text{IMU}}(i, j) + \varepsilon_z^T \sum_z^{-1} \varepsilon_z \right) \quad (10)$$

#### 4 定位精度试验与分析

为了进一步研究 IPL-VIO 算法在复杂场景下的定位精度,选取公开数据集,通过计算公开的数据集数据,对现有的算法和 IPL-VIO 算法进行评测,公开数据集的选取要反映出线束装配场景的复杂性,存在结构化特征、重复纹理、弱纹理的特点,同时也要体现出装配任务的复杂性。本研究选取 EuRoc 数据集,它是著名的应用于视觉惯性导航的数据集之一。用数据集的好处在于数据集的采集使用高精度的设备,可以提供精准的真值进行对比,减少了因设备和客观因素导致的误差。

EuRoc 提供 20 Hz 的双目相机图像、200 Hz 的 IMU 数据和相机真实的运动位姿,选取和装配场景类似的 Machine Hall 离线测试集,图像分辨率为 752 × 480,频率为 20 Hz, IMU 频率为 200 Hz,从均方根误差、均方根旋转误差、运行时间、GPU 占用率和内存占用率这 5 个方面对 VINS-Fusion<sup>[6]</sup>、PL-VIO 和 IPL-VIO 算法进行对比。VINS-Fusion 是香港科技大学 Qin 等<sup>[6]</sup>提出

的一种基于双目视觉惯性的状态估计算法,采用紧耦合、非线性优化的方法,融合预积分后的 IMU 测量值和特征观测值,获得高精度的视觉里程计。而 PL-VIO 在 VINS-Mono 的基础上加入了线特征约束。

试验采用的计算机配置:CPU 为 i7-10750 的 12 核英特尔处理器、16 GB 运行内存、Ubuntu18.04 操作系统。为充分测试算法处理效果,数据集回放速率为正常速率的 1/2,以此减少由于显示和加载导致的数据延迟问题。

VINS-Fusion 运行测试数据集 MH\_05\_difficult 的运行效果双目特征检测与跟踪基本一致,特征点分布均匀。表 1 所示为 VINS-Fusion 运行数据集时的参数,可以看出,运行数据集 MH\_03\_medium 时位移绝对误差最小,运行 MH\_05\_difficult 时旋转绝对误差、内存占用率最小。

使用 PL-VIO 运行测试数据集 MH\_05\_difficult 的运行效果点线特征检测与跟踪效果基本一致。表 2 所示为 PL-VIO 运行数据集时的参数,可以看出,运行数据集 MH\_03\_medium 时位移绝对误差最小,运行 MH\_05\_difficult 时旋转绝对误差、运行时间和 CPU 占用率最小。和 VINS-Fusion 相比,PL-VIO 在绝对位移和旋转上有所提高,由于线特征的提取和计算耗费更多的资源,使运行时间和 CPU 占用率显著提升。

使用 IPL-VIO 算法运行测试数据集 MH\_05\_difficult,

表 1 VINS-Fusion 运行数据集的参数

Table 1 Parameters of the VINS-Fusion run data set

数据集	位移绝对误差/m	旋转绝对误差/(°)	运行时间/ms	CPU 占用率/%	内存占用率/%
MH_01_easy	0.28893	0.05354	68	16.132	15.456
MH_02_easy	0.23363	0.03802	63	15.870	14.137
MH_03_medium	0.10219	0.01927	62	13.254	13.590
MH_04_difficult	0.14326	0.01673	54	13.354	12.395
MH_05_difficult	0.20951	0.01018	58	14.623	11.472
平均值	0.19550	0.02755	61	14.647	13.410

表 2 PL-VIO 运行数据集的参数

Table 2 Parameters of the PL-VIO run data set

数据集	位移绝对误差/m	旋转绝对误差/(°)	运行时间/ms	CPU 占用率/%	内存占用率/%
MH_01_easy	0.22554	0.04697	135	20.647	18.746
MH_02_easy	0.20274	0.03204	127	18.350	18.537
MH_03_medium	0.09224	0.01516	112	17.254	17.490
MH_04_difficult	0.12942	0.01315	108	16.724	15.365
MH_05_difficult	0.18687	0.00922	102	15.353	15.572
平均值	0.16736	0.02331	116.8	17.666	17.142

运行效果如图 3 所示,点特征提取分布均匀,跟踪效果好。双目基本可以检测出场景的线特征,且双目相机的左右目匹配一致,跟踪效果良好。

使用 IPL-VIO 算法运行测试数据集 MH\_05\_difficult,位姿估计效果如图 4 所示,可以发现基于双目的线特征检测数量较多,且 IPL-VIO 算法可以较好地反映出局部线段纵横交汇处的结构信息,丰富的场景信息有利于后期场景深度恢复和地图构建。

表 3 所示为 IPL-VIO 算法运行数据集时的参数,可以看出,运行数据集 MH\_03\_medium 时位移绝对误差最小,运行 MH\_05\_difficult 时旋转绝对误差、运行时间和 CPU 占用率最小。

VINS-Fusion、PL-VIO 的平均位移绝对误差分别为 0.19550 m 和 0.16736 m, IPL-VIO 算法的平均位移绝对误差为 0.14035 m,相比于 VINS-Fusion、PL-VIO, IPL-VIO 算法平均位移绝对误差最小。VINS-Fusion、PL-VIO



图 3 IPL-VIO 算法测试数据集的点线特征跟踪

Fig.3 Point and line feature tracking of IPL-VIO algorithm test data set

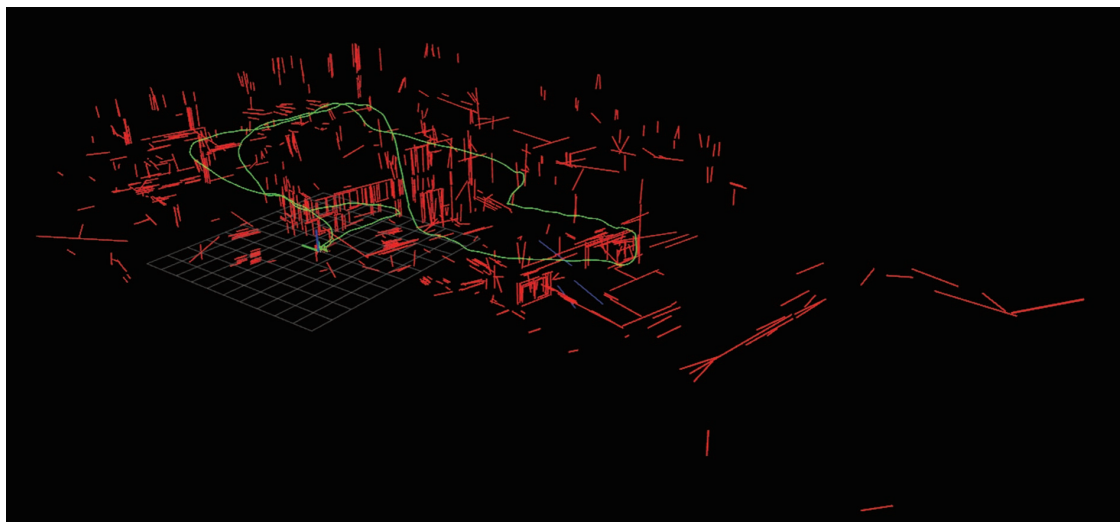


图 4 IPL-VIO 算法测试数据集点线特征位姿估计效果

Fig.4 Pose estimation effect of point and line feature in test data set by IPL-VIO algorithm

表 3 IPL-VIO 算法运行数据集的参数

Table 3 Parameters of run data set by the IPL-VIO algorithm

数据集	位移绝对误差/m	旋转绝对误差/(°)	运行时间/ms	CPU 占用率/%	内存占用率/%
MH_01_easy	0.18334	0.03627	185	30.647	35.386
MH_02_easy	0.17932	0.02734	173	28.930	33.257
MH_03_medium	0.08021	0.01402	135	27.254	30.290
MH_04_difficult	0.10532	0.01215	121	24.724	28.323
MH_05_difficult	0.15356	0.00765	110	23.283	25.692
平均值	0.14035	0.01949	144.8	26.968	30.590

的平均旋转绝对误差分别为  $0.02755^\circ$  和  $0.02331^\circ$ , IPL-VIO 算法的平均旋转绝对误差为  $0.01949^\circ$ , 相比于 VINS-Fusion、PL-VIO, IPL-VIO 算法平均旋转绝对误差最小。而在运行时间、CPU 占有率上, IPL-VIO 算法相对于 VINS-Fusion、PL-VIO 都有一定的增加。总之, 和 VINS-Fusion、PL-VIO 相比, IPL-VIO 算法在绝对位移和旋转上有明显提高, 但同时在 PL-VIO 的基础之上, 由于双目线特征余数的存在以及优化变量的增多, 使运行时间和 CPU 占用率显著提升。

综上所述, 通过对比 VINS-Fusion 和 PL-VIO 算法试验, 双目定位相比单目定位精度更高, 且场景恢复信息丰富; IPL-VIO 算法在结构化场景下的绝对位移和旋转误差比原算法误差更小, 同时结构化场景信息更加丰富, 但耗费计算资源较原来的算法多。

将 IPL-VIO 算法应用到座舱线束引导装配定位中, 试验采用的头戴式设备 HTC VIVE Pro 为一种虚拟现实头戴式设备, 可以通过基站进行定位, 也可用连接的摄像头观察外部真实世界实现增强现实的效果。试验采取远程协同的方式, 通过专家端和工人端来实现复杂产品装配。使用本系统装配试验的若干关键环节包括: (1) 装配工人启动装配系统, 装配系统通过相机与 HTC 坐标原点对齐等完成初始化任务; (2) 装配工人通过装配中心查看装配任务, 以座舱下某管线敷设为例, 系统提示工人佩戴挂载 ZED Mini 相机的头盔并在装配现场通过 AR 场景演示装配零件状态信息, 便于学习; (3) 当在零件区拿取零件之后, AR 相机渲染装配模型信息于座舱上, 信息中心提示装配人员进行装配操作; (4) 完成后信息中心提示进行安装检查操作, 结束后信息中心下达结束该工序命令。

试验以线束敷设安装为例, 为便于展示虚实融合效果, 本文以第一和第三视角进行混合叙述。开始后, 试验人员根据系统提示佩戴 AR 辅助终端, 在装配现场通过 AR 效果学习线束系统布局的大体空间位置信息。开始敷线操作时, 试验人员拿取束线管之后, 前往座舱装配区, 此时系统将束线管和座舱模型融合, 试验人员在现场观察虚实融合情况, 图 5 所示为试验人员第三视角和信息管理中心的第一视角下的虚拟束线管与座舱融合情况, 可以看出, 系统对试验人员头部的虚实遮挡情况处理得比较好。然后试验人员根据系统提示, 按照束线管的走向和布局进行束线操作。

装配人员可以随意走动观察零件空间位置信息, 线束模型定位稳定, 与真实场景融合良好, 且装配人员在走动时, AR 头戴式设备以及与其绑定的相机也随之运动的情况下, 虚拟线束模型在真实的座舱中虚实注册稳定, 没有出现模型剧烈移动现象。



(a) 线束引导装配定位第三视角



(b) 线束引导装配定位第一视角

图 5 座舱线束引导装配定位效果图

Fig.5 Guiding assembly and positioning rendering of cabin wiring harness

## 5 结论

提出了一种点线特征融合的双目视觉惯性定位算法, 并对该方法进行了全面的研究, 试验结果表明, 本文提出的 IPL-VIO 算法误差更小, 定位精度较高, 能够应用在弱纹理的 AR 装配现场中, 为增强现实辅助装配平台提供稳定可靠的位姿数据。近年来随着 AR 技术的发展, 特别是在大场景弱纹理的 AR 装配现场, 本方法有着较大的应用前景。下一步将考虑采取结构化的直线代替装配场景中杂乱无章的线段, 这将在线段的提取、匹配上节省一定的时间和算力, 从而减少定位的时间并提高定位的精度, 更好地实现增强现实设备的实时定位。

## 参考文献

- [1] 王峻峰, 陈亭, 李世其, 等. BRIEF 算法在基于增强现实的产品装配引导中的应用研究 [J]. 机械科学与技术, 2013, 32(2): 213-216.  
WANG Junfeng, CHEN Ting, LI Shiqi, et al. Applying BRIEF algorithm to product assembly guidance based on augmented reality[J]. Mechanical Science and Technology for Aerospace Engineering, 2013, 32(2): 213-216.
- [2] 刘检华. 军工数字化装配技术 [J]. 国防制造技术, 2011(4): 5-7, 11.  
LIU Jianhua. Digital military assembly technology[J]. Defense Manufacturing Technology, 2011(4): 5-7, 11.

- [3] WIEDENMAIER S, OEHME O, SCHMIDT L, et al. Augmented reality (AR) for assembly processes design and experimental evaluation[J]. *International Journal of Human-Computer Interaction*, 2003, 16(3): 497–514.
- [4] RADKOWSKI R, HERREMA J, OLIVER J. Augmented reality-based manual assembly support with visual features for different degrees of difficulty[J]. *International Journal of Human-Computer Interaction*, 2015, 31(4–6): 337–349.
- [5] 高翔, 张涛, 刘毅. 视觉 SLAM 十四讲: 从理论到实践 [M]. 北京: 电子工业出版社, 2017.
- GAO Xiang, ZHANG Tao, LIU Yi. Fourteen lectures on visual SLAM: From theory to practice[M]. Beijing: Publishing House of Electronics Industry, 2017.
- [6] 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.
- [7] 李民. 基于惯性和视觉组合的头盔位姿测量系统分析与设计 [D]. 济南: 山东大学, 2019.
- LI Min. Analysis and design of helmet pose measurement system based on inertia and vision combination[D]. Jinan: Shandong University, 2019.
- [8] 王崑, 彭勃宇, 周诚, 等. 面向增强现实的多传感器组合位姿注册算法 [J]. *传感器与微系统*, 2015, 34(7): 132–136, 140.
- WANG Wei, PENG Boyu, ZHOU Cheng, et al. Multi-sensor combined posture registration algorithm for augmented reality[J]. *Transducer and Microsystem Technologies*, 2015, 34(7): 132–136, 140.
- [9] 梁明杰, 闵华清, 罗荣华. 基于图优化的同时定位与地图创建综述 [J]. *机器人*, 2013, 35(4): 500–512.
- LIANG Mingjie, MIN Huaqing, LUO Ronghua. Graph-based SLAM: A survey[J]. *Robot*, 2013, 35(4): 500–512.
- [10] KÜMMERLE R, GRISSETTI G, STRASDAT H, et al. G2o: A general framework for graph optimization[C]//2011 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2011: 3607–3613.
- [11] AGARWAL S, MIERLE K. Ceres solver: Tutorial & reference[J]. Google Inc, 2012, 2: 72.
- [12] CARLONE L, KIRA Z, BEALL C, et al. Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors[C]//2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong: IEEE, 2014: 4290–4297.
- [13] LEUTENEGGER S, LYNEN S, BOSSE M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. *International Journal of Robotics Research*, 2015, 34(3): 314–334.
- [14] BAKER S, MATTHEWS I. Lucas-kanade 20 years on: A unifying framework[J]. *International Journal of Computer Vision*, 2004, 56(3): 221–255.
- [15] 曾雅雯. 融合语义不变量的点线多特征双目 SLAM 算法 [D]. 南宁: 广西大学, 2021.
- ZENG Yawen. A multi-feature fusion stereo slam system attaching semantic invariant to points and lines[D]. Nanning: Guangxi University, 2021.
- [16] 沈冉. 点-线特征融合的单目视觉惯性里程计 [D]. 吉林: 东北电力大学, 2021.
- SHEN Ran. Monocular vision inertial odometer based on point line feature fusion[D]. Jilin: Northeast Electric Power University, 2021.
- [17] HE Y J, ZHAO J, GUO Y, et al. PL-VIO: Tightly-coupled monocular visual-inertial odometry using point and line features[J]. *Sensors (Basel, Switzerland)*, 2018, 18(4): 1159.
- [18] 赵良玉, 金瑞, 朱叶青, 等. 基于点线特征融合的双目惯性 SLAM 算法 [J]. *航空学报*, 2022, 43(3): 363–377.
- ZHAO Liangyu, JIN Rui, ZHU Yeqing, et al. Stereo visual-inertial slam with point and line features[J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(3): 363–377.
- [19] REHDER J, NIKOLIC J, SCHNEIDER T, et al. Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes[C]//2016 IEEE International Conference on Robotics and Automation (ICRA). Stockholm: IEEE, 2016: 4304–4311.
- [20] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91–110.
- [21] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346–359.
- [22] VISWANATHAN D. Features from accelerated segment test (FAST)[C]//Proceedings of the 10th workshop on image analysis for multimedia interactive services. London, 2009.
- [23] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision. Barcelona: IEEE, 2012: 2564–2571.
- [24] CALONDER M, LEPETIT V, STRECHA C, et al. BRIEF: binary robust independent elementary features[M]//Computer Vision—ECCV 2010. Berlin: Springer Berlin Heidelberg, 2010: 778–792.
- [25] FISCHLER M A, BOLLES R C. Random sample consensus[J]. *Communications of the ACM*, 1981, 24(6): 381–395.
- [26] GROMPONE VON GIOI R, JAKUBOWICZ J, MOREL J M, et al. LSD: A fast line segment detector with a false detection control[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(4): 722–732.
- [27] HUAI J Z, TOTH C, GREJNER-BRZEZINSKA D. Stereo-inertial odometry using nonlinear optimization[C]//Proceedings of the 28th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2015). Tampa, 2015.
- [28] ENGEL J, STÜCKLER J, CREMERS D. Large-scale direct SLAM with stereo cameras[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg: IEEE, 2015: 1935–1942.
- [29] THRUN S, MONTEMERLO M. The graph SLAM algorithm with applications to large-scale mapping of urban structures[J]. *International Journal of Robotics Research*, 2006, 25(5–6): 403–429.
- [30] NI K, STEEDLY D, DELLAERT F. Out-of-core bundle adjustment for large-scale 3D reconstruction[C]//2007 IEEE 11th International Conference on Computer Vision. Rio de Janeiro: IEEE, 2007: 1–8.
- [31] ZHANG G X, LEE J H, LIM J, et al. Building a 3-D line-based map using stereo SLAM[J]. *IEEE Transactions on Robotics*, 2015, 31(6): 1364–1377.

通讯作者: 白晓亮, 副教授, 主要从事增强现实、智能装备、人机交互等研究。

(责编 阳光)