

## 农业场景下移动机器人的双目视觉定位与地图构建方法

余涛, 熊盛武

引用本文

余涛, 熊盛武. [农业场景下移动机器人的双目视觉定位与地图构建方法](#)[J]. 计算机科学, 2023, 50(12): 185-191.

YU Tao, XIONG Shengwu. [Stereo Visual Localization and Mapping for Mobile Robot in Agricultural Environments](#) [J]. Computer Science, 2023, 50(12): 185-191.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

#### [基于双目视觉的车辆速度测量方法](#)

Vehicle Speed Measurement Method Based on Binocular Vision

计算机科学, 2021, 48(9): 135-139. <https://doi.org/10.11896/jsjcx.201000047>

#### [一种自适应尺度与学习速率调整的背景感知相关滤波跟踪算法](#)

Background-aware Correlation Filter Tracking Algorithm with Adaptive Scaling and Learning Rate Adjustment

计算机科学, 2021, 48(5): 177-183. <https://doi.org/10.11896/jsjcx.200300109>

#### [双视系统的室内三维场景重建研究](#)

Study on Reconstruction of Indoor 3D Scene Based on Binocular Vision

计算机科学, 2020, 47(11A): 175-177. <https://doi.org/10.11896/jsjcx.200400096>

#### [基于信息熵的级联Siamese网络目标跟踪](#)

Cascaded Siamese Network Visual Tracking Based on Information Entropy

计算机科学, 2020, 47(9): 157-162. <https://doi.org/10.11896/jsjcx.190800160>

#### [基于新型语言机制的异构集群应用通信优化方法](#)

Communication Optimization Method of Heterogeneous Cluster Application Based on New Language Mechanism

计算机科学, 2020, 47(8): 17-15. <https://doi.org/10.11896/jsjcx.200100124>

# 农业场景下移动机器人的双目视觉定位与地图构建方法

余涛<sup>1</sup> 熊盛武<sup>1,2</sup>

1 武汉理工大学计算机科学与人工智能学院 武汉 430070

2 武汉理工大学三亚科教创新园 海南 三亚 572000

(tyu@whut.edu.cn)

**摘要** 视觉定位与地图构建是实现移动机器人自主导航的关键技术。针对农业场景下特征跟踪困难、场景规模大、运动不稳定引起系统精度和鲁棒性下降的问题,提出了一种适用于农业场景的双目视觉定位与地图构建方法。该方法首先利用静态立体匹配点来增加跟踪阶段地图点的数量和覆盖范围,从而增加了深度计算的准确率,同时提出一种点选择算法对密集地图点进行采样并移除离群点,进一步提高了系统的准确率和运行效率;然后通过显式尺度估计来减小大规模场景下定位与地图构建的尺度误差,并结合场景特点改进关键帧判别策略,避免了远处大目标导致关键帧稀疏的问题;最后提出新的运动假设构建位姿估计失败时的恢复策略,提高了系统在颠簸运动时的鲁棒性。在农业场景数据集上的评估结果表明,相比于当前先进的视觉定位与地图构建系统,提出的方法在困难序列上的轨迹误差降低幅度超过50%,其中3个序列上的尺度误差下降了一个数量级,取得了更高的精度和鲁棒性,能有效地应对农业场景下视觉定位与地图构建的挑战。

**关键词:** 农业场景;视觉定位与地图构建;直接法;双目视觉;尺度估计

**中图法分类号** TP391

## Stereo Visual Localization and Mapping for Mobile Robot in Agricultural Environments

YU Tao<sup>1</sup> and XIONG Shengwu<sup>1,2</sup>

1 School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China

2 Sanya Science and Education Innovation Park of Wuhan University of Technology, Sanya, Hainan 572000, China

**Abstract** Visual-based localization and mapping is the key technology for autonomous robots. Visual localization and mapping in agricultural environments faces more challenges, including few distinguishable landmarks for tracking, large-scale scene, unstable movements. To address these problems, a stereo visual localization and mapping method is proposed. Static stereo matching points are used to increase the number and coverage of map points, which improve the accuracy of depth calculation. A point selection method is proposed to further improve the accuracy and efficiency by sampling the dense map points and removing outliers. Then scale estimation is proposed to reduce the scale error of localization and mapping in large scale agricultural scenes. Keyframe criteria is adapted to avoid the impact of large far-away objects that could cause abnormal keyframe distribution. Finally, a new motion assumption is proposed to recover the system from failure tracking, which improves the system's robustness at the case of unstable movements. Experimental results show that the proposed method achieves better performance than other state-of-the-art visual localization and mapping systems. By addressing the challenges individually, the proposed visual localization and mapping system is more accurate and robust in agricultural environments.

**Keywords** Agricultural environments, Visual localization and mapping, Direct method, Stereo vision, Scale estimation

## 1 引言

自主移动机器人在精准农业领域的应用越来越广泛<sup>[1]</sup>,如除草、作物监测、农产品收割等,因此受到了越来越多的

关注。在这个背景下,即时定位与地图构建(Simultaneous Localization and Mapping, SLAM)算法成为了自主农业机器人进行作物数据采集、感知、解译工作的重要基础。视觉传感器具有纹理信息丰富、成本低、轻量级、容易部署等优点,因

到稿日期:2023-03-14 返修日期:2023-07-07

基金项目:国家自然科学基金(62176194,62101393);车联网重大项目(2020AAA001);三亚崖州湾科技城项目(SCKJ-JYRC-2022-76);武汉理工大学三亚科教创新园项目(2021KF0031);重庆市自然科学基金(cstc2021jcyj-msxmX1148)

This work was supported by the National Natural Science Foundation of China(62176194,62101393), Major Project of IoV(2020AAA001), Project of Sanya Yazhou Bay Science and Technology City(SCKJ-JYRC-2022-76), Sanya Science and Education Innovation Park of Wuhan University of Technology(2021KF0031) and Natural Science Foundation of Chongqing, China(cstc2021jcyj-msxmX1148).

通信作者:熊盛武(xiongsw@whut.edu.cn)

此,基于视觉的定位和地图构建是自主机器人领域关注的重点。然而,农业场景是一个典型的非结构化环境,具有可区分性的视觉特征少、纹理重复、光照变化、路面不平、物体呈动态等特点,该场景下的视觉定位与地图构建面临着新的挑战。

近年来,实时的视觉定位与地图构建算法取得了重大的进步,在自动驾驶、虚拟现实、机器人自主导航等领域都有广泛的应用。但是,目前大多数的研究集中在室内或者城市等具有明显结构化特征的场景。这些研究成果很难在具有挑战性的农业场景下取得较好的效果。由于农业场景具有可区分特征少、重复纹理多等特点,基于特征点的算法<sup>[2-5]</sup>在跟踪阶段面临较大的困难。另一方面,由于不需要计算复杂的特征描述子,因此,基于光流的算法<sup>[6]</sup>具有较高的效率;但是这类算法在长时间的运动中准确率较低。由于充分利用了图像的信息,直接法<sup>[7-11]</sup>可以提供准确和鲁棒的运动估计;但是基于非凸的能量函数高度依赖于位姿和深度初值的选取。与此同时,由于农业机器人偶尔不稳定的运动,特别是在与其他导致视觉变化较大的运动(转弯、快速运动)叠加的情况下,系统可能出现位姿估计失败的情况。大多数农业任务(作物监控、除草等)需要获取环境的尺度信息,所以本文使用双目相机作为视觉传感器。

为了应对农业场景下的挑战,本文提出了一个适用于农业场景的双目视觉定位与地图构建系统。图1展示了本文提出的方法输出的关键帧深度图、输入图像、机器人的轨迹以及环境点云地图。本文工作的主要贡献如下:1)针对可区分性特征少的问题,加入静态立体匹配点来增加跟踪阶段地图点的数量和深度计算的准确率,同时提出一种点选择算法对地图点进行采样并移除离群点,进一步提高了系统的准确率和效率;2)通过显式尺度估计来减小大规模场景下定位与地图构建的尺度误差,并结合场景特点改进关键帧判别策略,避免了远处大目标导致关键帧稀疏的问题;3)提出新的运动假设构建位姿估计失败恢复策略,提高了系统在颠簸运动时的鲁棒性;4)在农业数据集上对多个先进的视觉定位和地图构建系统做了全面的比较,提出的系统取得了更好的性能。

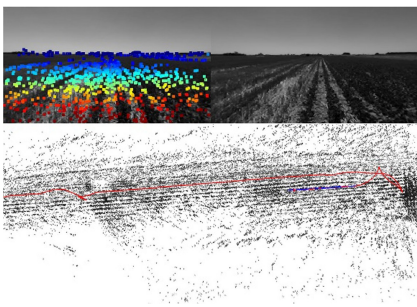


图1 系统在Rosario数据集中04序列的输出

Fig.1 Outputs of the proposed system for sequence 04 on Rosario dataset

## 2 相关工作

### 2.1 视觉定位与地图构建方法

视觉定位与地图构建是近几十年来的计算机视觉和机器人社区的热门话题。相关领域的研究者提出了大量的算法,

这些算法通常可以分为两类:间接法和直接法。

间接法依赖于可重复的视觉特征,如ORB<sup>[12]</sup>,通过对特征点的描述子匹配或光流跟踪来实现数据关联,然后通过最小化几何重投影误差来估计相机的运动和三维的地图点。近年来,为了提高视觉定位和地图构建的准确率、鲁棒性和运行效率,研究者们提出了大量优秀的基于间接法的系统。例如,PTAM<sup>[2]</sup>利用多核中央处理器(CPU)的特性,将单目视觉定位和地图构建运行在不同的线程中,实现了算法的实时运行。S-PTAM<sup>[3]</sup>对其进行了扩展,通过使用双目图像获取了环境的尺度信息,并提高了地图构建的准确率。基于ORB特征的ORB-SLAM2<sup>[4]</sup>加入了地图重用、闭环检测、重定位模块,是一个完整的视觉定位和地图构建系统。ORB-SLAM3<sup>[5]</sup>通过使用抽象相机模型增强了通用性,系统将双目相机视为两个固定变换的单目相机进行处理,不再依赖于双目立体校正。VINS<sup>[6]</sup>是一个基于多传感器融合优化的状态估计器,通过融合惯性测量单元的信息,提高了系统的准确率和鲁棒性。

间接法在具有明显特征的场景下具有稳定、鲁棒等优点,并且可以利用特征点进行回环检测,应用十分广泛。但在可区分性特征少的农业场景下,特征匹配的难度增加,重定位比较困难,而且特征点的提取会消耗大量计算资源,添加额外的处理流程难以保证系统的实时性。

直接法的目标是利用图像的原始灰度值直接估计相机的运动和三维环境的结构信息,主要使用高斯-牛顿方法对光度误差函数进行优化。DTAM<sup>[7]</sup>可以同时进行稠密的三维地图重建和准确的状态估计,经过并行优化后,可以在高性能图形处理器(GPU)的设备上实时运行。半直接法SVO<sup>[8]</sup>利用高对比度的像素点进行图像对齐,可以有效地降低计算成本。DSO<sup>[9]</sup>在大型单目相机跟踪数据集上取得了最高的准确率和鲁棒性。SDSO<sup>[10]</sup>将静态立体匹配约束融入DSO<sup>[9]</sup>后端优化的能量函数中,提高了系统在大尺度数据集上的精度。SODSO<sup>[11]</sup>将尺度优化模块集成到DSO<sup>[9]</sup>中,来提高双目视觉里程计的精度和效率。

相比于间接法,直接法具有计算效率高、不要求明显纹理特征、图像利用率高等优点,能够在可区分性特征少的农业场景下工作。但是,此类方法使用像素灰度梯度进行位姿估计,容易陷入局部最优。因此,本文主要在直接法的基础上提高位姿估计的精度和鲁棒性。

### 2.2 农业场景下的定位与地图构建

为了应对农业场景下的挑战,有些研究者利用植物自身的特性来提高定位和地图构建的性能。Cheein等<sup>[13]</sup>提出了一种茎检测算法,并将其应用于橄榄林的视觉定位与地图构建中。Juman等<sup>[14]</sup>通过树干识别实现了油棕树林中的机器人自主导航。Mendes等<sup>[15]</sup>利用自然场景特征构建机器人的定位系统。另一些研究<sup>[16-18]</sup>利用作物行检测和跟踪的方法来对移动机器人进行导航。上述方法需要依赖特定植物的特征,是专用系统,应用领域受具体植物类别的限制。

近年来,为了提高视觉定位和地图构建的准确率和鲁棒性,多传感器融合和协同地图构建受到了越来越多的关注。Imperoli等<sup>[19]</sup>设计了一个基于多模态信息融合的农业机器人定位系统,主要使用了GNSS数据、相机数据、惯性测量单元

数据、数字高程模型。为了应对农业场景下的视觉变化问题, Dong 等<sup>[20]</sup>和 Chebrolu 等<sup>[21]</sup>通过几种板载传感器和航空地图来提高定位系统的性能。Potena 等<sup>[22]</sup>提出了基于无人机和无人车的空地协同三维地图构建系统。同时, Preto 等<sup>[23]</sup>设计了一个自适应的空地协同机器人系统,并将其用于精准农业领域。

上述融合系统主要关注数据融合方法,需要额外的传感器或者参考地图,而不是提高单个传感器的性能。本文主要研究双目视觉定位与地图构建方法,从而提高系统在农业场景下的精度和鲁棒性,也可以为融合系统提供更精确的位置和环境信息。

### 3 农业场景下的视觉定位和地图构建方法

本节将详细介绍适用于农业场景的双目视觉定位和地图构建方法。系统首先利用第一帧的静态立体匹配点生成半稠密地图,并从中挑选匹配得分较高的点作为初始地图点。然后,对于新加入的图像帧,采用恒速模型假设对该帧进行位姿估计,如果估计失败,则执行恢复策略。再利用当前帧对候选点进行深度优化,同时判断当前帧是否为关键帧,如果是关键帧,则进行滑动窗口优化,以保证系统的局部一致性。最后,输出当前帧的位姿和窗口内的地图点信息。

#### 3.1 双目直接法位姿估计

直接法通过图像的原始灰度信息来估计相机的相对运动。这里首先介绍基于单目直接法的位姿估计,然后再扩展到双目图像。假设第  $i$  个参考帧为  $F_i$ , 包含灰度图像  $I_i$  以及一个图像点集合  $\mathcal{P}_i$ , 且已知  $\mathcal{P}_i$  中的每个点  $\mathbf{p}$  对应的深度  $d_p$  对应的地图点可以用  $(\mathbf{p}, d_p)$  表示。如果新的一帧  $F_j$  可以观测到  $F_i$  上的地图点  $(\mathbf{p}, d_p)$ ,  $\mathbf{p}$  在  $F_j$  的投影为  $\mathbf{p}'$ , 则可以根据光度一致性建立如下光度误差  $r_{ij}(\mathbf{p})$ :

$$r_{ij}(\mathbf{p}) = I_j(\mathbf{p}') - I_i(\mathbf{p}) \quad (1)$$

其中,  $\mathbf{p}'$  的计算方式如下:

$$\mathbf{p}' = \pi(\mathbf{T}_{ji}\pi^{-1}(\mathbf{p}, d_p)) \quad (2)$$

其中,  $\pi(\cdot)$  与  $\pi^{-1}(\cdot)$  分别是相机的投影函数和反投影函数;  $\mathbf{T}_{ji} \in SE(3)$  是第  $i$  帧到第  $j$  帧的相对变换, 即两帧之间的相对运动, 也可用旋转矩阵  $\mathbf{R}_{ji}$  和平移向量  $\mathbf{t}_{ji}$  表示, 即:

$$\mathbf{T}_{ji} = \begin{Bmatrix} \mathbf{R}_{ji} & \mathbf{t}_{ji} \\ \mathbf{0}^T & 1 \end{Bmatrix} = \mathbf{T}_j^{-1}\mathbf{T}_i \quad (3)$$

在复杂的环境下, 相机的曝光时间是未知的, 图像的光度一致性通常难以满足。假设真实灰度值为  $I_g$ , 图像上的灰度值为  $I_c$ , 用光度仿射变换<sup>[9]</sup>建模可得:

$$I_c = e^a I_g + b \quad (4)$$

其中,  $a, b$  均为变换参数;  $\exp(\cdot)$  可以保证系数为正, 并且提高计算时的数值稳定性。通过式(5), 可以根据图像灰度值和光度参数计算真实灰度值:

$$I_g = e^{-a}(I_c - b) \quad (5)$$

将式(5)代入式(1)并整理, 可得考虑光度仿射变换的光度误差:

$$r_{ij}(\mathbf{p}) = I_j(\mathbf{p}') - b_j - \frac{e^{a_j}}{e^{a_i}}(I_i(\mathbf{p}) - b_i) \quad (6)$$

其中,  $a_i, b_i, a_j, b_j$  分别为第  $i$  帧和第  $j$  帧的光度仿射参数。

然后可以计算  $\mathcal{P}_i$  中所有点的光度误差并求和, 最终得到两帧之间的总光度误差能量函数  $E_{ij}$ :

$$E_{ij} = \sum_{\mathbf{p} \in \mathcal{P}_i} \omega_p \|r_{ij}(\mathbf{p})\|_\gamma \quad (7)$$

其中,  $\|\cdot\|_\gamma$  是 Huber 鲁棒核函数, 用于降低离群点对能量函数的影响;  $\omega_p$  是一个与点  $\mathbf{p}$  处的灰度梯度  $\nabla I(\mathbf{p})$  相关的自适应权重, 用于减小高梯度点(可能是噪声点)对能量函数的影响, 具体的表达式如下:

$$\omega_p = \frac{\epsilon^2}{\epsilon^2 + \|\nabla I(\mathbf{p})\|_2^2} \quad (8)$$

其中,  $\epsilon$  可以保证数值稳定性。通过最小化两帧之间的总光度误差, 可以求得两帧之间的相对运动和  $F_j$  的光度仿射参数, 即:

$$\mathbf{T}_{ji}^*, a_j^*, b_j^* = \arg \min_{\mathbf{T}_{ji}, a_j, b_j} E_{ij} \quad (9)$$

结合已知位姿  $\mathbf{T}_i$  的参考帧  $F_i$ , 便可以通过式(10)求得当前帧  $F_j$  的位姿  $\mathbf{T}_j$ , 将其作为后续算法的输入。

$$\mathbf{T}_j = \mathbf{T}_{ji}\mathbf{T}_i \quad (10)$$

相比于单目图像, 双目图像可以获得绝对尺度信息用于移动机器人的自主导航, 在位姿估计时也可以提供更多的信息。由于双目相机之间的变换是已知的, 为了简化表示, 本文将每一帧左影像的位姿作为该帧的位姿。类似基于单目直接法的位姿估计, 我们可以分别建立参考帧  $F_i$  的左影像与当前帧  $F_j$  左右影像 ( $I_j^l, I_j^r$ ) 的光度误差。基于双目直接法的位姿估计总光度误差如下:

$$E_{ij} = E_{ij}^l + E_{ij}^r = \sum_{\mathbf{p} \in \mathcal{P}_i} (\omega_p^l \|r_{ij}^l(\mathbf{p})\|_\gamma + \omega_p^r \|r_{ij}^r(\mathbf{p})\|_\gamma) \quad (11)$$

其中,  $r_{ij}^l(\mathbf{p})$  是  $\mathbf{p}$  点关于参考帧左影像与当前帧左影像之间的光度误差, 与式(6)相同;  $r_{ij}^r(\mathbf{p})$  表示  $\mathbf{p}$  点关于参考帧左影像与当前帧右影像之间的光度误差, 即:

$$r_{ij}^r(\mathbf{p}) = I_j^r(\mathbf{p}'') - b_j^r - \frac{e^{a_j^r}}{e^{a_i^l}}(I_i^l(\mathbf{p}) - b_i^l) \quad (12)$$

$\mathbf{p}''$  在右影像上的投影位置的计算方式如下:

$$\mathbf{p}'' = \pi(\mathbf{T}_{ji}\mathbf{T}_i\pi^{-1}(\mathbf{p}, d_p)) \quad (13)$$

$\mathbf{T}_{ri}$  是左影像到右影像的刚体变换, 可由双目相机事先标定获得。然后通过最小化式(11)的光度误差, 即可求解当前帧相对于参考帧的位姿以及左右影像的光度参数。

在大规模场景中, 随着机器人的运行, 物体的尺度会发生变化, 导致图像测量误差累积, 最终导致地图点出现尺度误差。为了缓解这个问题, 本文进行显式的尺度估计, 将尺度参数  $s$  作为待优化变量加入到式(13)中, 所以  $\mathbf{p}''$  最终的计算方式为:

$$\mathbf{p}'' = \pi(\mathbf{T}_{ri}\mathbf{T}_i s \pi^{-1}(\mathbf{p}, d_p)) \quad (14)$$

### 3.2 双目视觉定位与地图构建方法

#### 3.2.1 参考帧初始化

首先将滑动窗口中的所有地图点投影到最新的关键帧上, 其中有部分点会投影到同一个位置, 这里使用加权平均的方式对深度进行平滑操作, 权重  $\omega_p$  通过深度的不确定性计算得到。为了增加地图点的数量和分布范围, 提高参考帧初始化时的精度, 本文加入了最新关键帧上一些高质量的静态立体匹配点。点的质量指标采用该点进行双目图像极线匹配时小能量与最小能量的比值。同时, 为了避免远处目标的视

觉测量误差的影响,仅保留深度小于 $\theta$ 倍基线的点。

### 3.2.2 点选择算法

初始化完成之后,我们发现参考帧地图点分布不均匀,且存在一些离群点,这会降低优化算法的性能,因此本文提出了一个点选择算法(见算法1)来对密集点进行采样并移除部分离群点。

#### 算法1 点选择算法

输入:  $\{(p, d_p, w_p) | p \in \mathcal{P}\}, g$

输出:  $(T_{ji}, a_j^l, b_j^l, a_j^r, b_j^r)$

1. 将图像划分成  $g \times g$  的网格
2. 统计每个网格中地图点的深度均值与标准差
3. 将每个网格中的点按照权重排序
4. 只保留权重最大且符合  $3\sigma$  原则的一个点, 移除其他点
5. 构建当前帧左右影像的图像金字塔以及灰度梯度金字塔

点选择算法可以减少点的数量并移除部分离群点,更少的点有利于提高系统的运行效率,移除离群点可以提高系统的精度。

### 3.2.3 位姿估计

完成上述步骤后,使用3.1节中的双目直接法来进行位姿估计。由于图像的高度非凸性,灰度梯度只在1~2个像素范围内有效<sup>[9]</sup>,因此必须假设优化时的初始值靠近最优值,才能在一定程度上保证算法收敛。如果相机运动过快,两张图像的差异较大,优化算法容易陷入局部最优,因此使用图像金字塔进行由粗到精的位姿优化来改善这种状况。基于图像金字塔的位姿估计算法的具体步骤如算法2所示。

#### 算法2 基于图像金字塔的位姿估计算法

输入:  $(I_1^l, I_1^r, I_1^s, T_1, T_{ji}^l, \{(p, d_p) | p \in \mathcal{P}\}, a_1^l, b_1^l)$

输出:  $(T_{ji}, a_j^l, b_j^l, a_j^r, b_j^r, s, T_j)$

1. 构建参考帧左影像的图像金字塔,层数为  $L$ ,并通过线性插值将深度值传递到高层的金字塔
2. 构建当前帧左右影像的图像金字塔以及灰度梯度金字塔
3. for  $k \leftarrow L$  to 2 do

根据式(6),利用位姿和光度参数的初始值构建总光度误差  $E_{ij}^k$ ,使用高斯牛顿法优化  $T_{ji}, a_j^l, b_j^l$ ,并将其作为下一次计算的初始值。

end for

4. 根据式(11)、式(14)构建总光度误差  $E_{ji}$ ,高斯牛顿法求解  $(T_{ji}, a_j^l, b_j^l, a_j^r, b_j^r)$
5. 根据式(10)计算  $T_j$ ,输出当前帧的位姿、光度参数以及尺度

### 3.2.4 关键帧判别

在农业场景下,相机的视野较开阔,远处的大目标在每一帧中都是可见的。之前的关键帧判别方法未考虑此现象,会导致关键帧的数量很少,而且只对较新的关键帧进行边缘化,会使优化窗口内的关键帧平移量过大,最终带来不可忽略的尺度误差,甚至造成系统失败的情况。因此,本文基于DSO<sup>[9]</sup>中的关键帧判别条件,选择距离传感器较近的点进行判别计算,降低远处点的权重。权重函数  $w_p$  如下:

$$w_p = \begin{cases} 1, & d_p \leq \theta \|t_m\|_2 \\ e^{-(d_p - \theta \|t_m\|_2)^2}, & d_p > \theta \|t_m\|_2 \end{cases} \quad (15)$$

其中,  $\|t_m\|_2$  表示当前帧到参考帧之间的平移距离,  $\theta$  是一个阈值。然后可以计算光流变化的加权均方根值  $f$ ,用于后续

的关键帧判断。 $f$  的计算方式如下:

$$f = \left( \frac{1}{\sum_{i=1}^n w_{p_i}} \left( \sum_{i=1}^n w_{p_i} \|p_i - p_i'\|_2^2 \right) \right)^{\frac{1}{2}} \quad (16)$$

其中,  $p_i$  表示成功跟踪的图像点,  $n$  是图像点的数量。同时可以计算只考虑平移变换的光流变化的加权均方根值  $f_i$ ,最终通过  $w_f f + w_{f_i} f_i + w_{f_a} f_a > 1$  来判断是否需要创建新的关键帧,其中,  $w_f, w_{f_i}, w_{f_a}$  是3个指标的相对权重,  $f_a$  是当前帧与参考帧之间的相对光度变化因子<sup>[9]</sup>。

### 3.2.5 位姿估计失败恢复

典型的农业场景一般具有局部平坦的地面,机器人在竖直方向上的运动较小。本文系统使用的坐标系  $x, y, z$  对应的方向为右-前-上,这里由于初始化时机器人姿态的不确定性,整体坐标系会绕  $x$  轴有一定的旋转角度。机器人的局部轨迹在  $yz$  平面上应该是近似一条线,线的斜率可由窗口内所有帧的世界坐标集合  $T$  通过式(17)求得。

$$k = \frac{\sum_{t_x \in T} t_{ix} / \sum_{t_x \in T} t_{iy}}{\sum_{t_x \in T} t_{ix}} \quad (17)$$

其中,  $(t_x, t_y, t_z)$  表示世界坐标系中的坐标。不平的道路容易造成机器人在竖直方向的抖动,这可能导致基于恒速运动的假设初始误差太大,从而导致位姿估计失败。为了应对这种情况,当恒速模型失效时,本文假设机器人以  $x$  轴为旋转轴,做与上一时刻旋转方向相反的运动,在竖直方向上无平移。新的运动假设可以通过式(18)计算:

$$\begin{cases} (\alpha, \beta, \gamma) = f(\mathbf{R}) \\ \mathbf{R}' = f^{-1}(-\alpha, \beta, \gamma) \\ t' = (t_x, t_y, k * t_y)^T \end{cases} \quad (18)$$

其中,  $f(\cdot)$  是一个转换函数,可以把旋转矩阵转换为欧拉角;  $f^{-1}(\cdot)$  是其逆变换;  $(\alpha, \beta, \gamma)$  分别是绕  $x, y, z$  轴旋转的欧拉角。

### 3.2.6 滑动窗口优化

在这个步骤中,窗口中地图点的深度和关键帧的位姿将被联合优化,以保证帧间的几何一致性。滑动窗口的总能量函数  $E_{total}$  可以表示为:

$$E_{total} = \sum_{F_i \in F} \sum_{p \in \mathcal{P}_{F_i}} \sum_{F_j \in obs(p)} w_p \|r_{ij}^l(p)\|_2 \quad (19)$$

其中,  $obs(p)$  表示滑动窗口中所有能够观测到  $p$  点的关键帧。为了保证固定的窗口大小,控制计算复杂度,舒尔补(Schur Complement)被用于旧关键帧的边缘化,该帧对应的地图点也被舍弃以保证优化时 Hessian 矩阵的稀疏性质。

为了减小长时间运动引起的尺度漂移误差,本文首先利用窗口中的每一帧双目影像来优化对应的尺度参数和右影像的光度参数。能量函数  $E_{is}$  可以表示为:

$$r_{is}(p) = I_i^l(p') - b_i^l - \frac{e^{a_i^l}}{e^{a_i^r}} (I_i^r(p) - b_i^r) \quad (20)$$

$$E_{is} = \sum_{p \in \mathcal{P}_i} (w_p \|r_{is}(p)\|_2)$$

其中,  $r_{is}$  表示该帧左图像与右图像之间的光度误差。为了保证结果的可靠性,如果点的数量较少,则跳过此流程。优化完成之后,利用尺度参数  $s$  对地图点进行缩放,然后将所有的点投影到最新的关键帧上,使用式(20)进行优化以保证尺度的一致性。

## 4 实验与分析

### 4.1 数据集与评估指标

农业场景下用于视觉定位和地图构建的数据集较少。最新发布的 Rosario 数据集<sup>[24]</sup>包含了移动机器人在农场采集的多种传感器数据,其中包括双目图像、轮速里程计、IMU 数据、高精度 GNSS 位置信息。这个数据集总共有 6 个序列,展现了真实大豆农场环境下的挑战:高重复度的场景、大规模场景、阳光直射、崎岖地形等。场景的自相似度较高,使得特征检测和跟踪变得非常困难,导致视觉场景识别算法失效,闭环检测模块无法得到有效的结果。在这种场景下进行视觉定位和地图构建具有较高的挑战性。评估指标主要包括绝对轨迹误差和尺度误差。绝对轨迹误差 ATE 定义如下:

$$ATE = \sqrt{\frac{1}{N} \sum_{t=0}^N (\mathbf{P}_t - \tilde{\mathbf{P}}_t)^2} \quad (21)$$

其中,  $N$  表示轨迹点的数量,  $\mathbf{P}_t$  和  $\tilde{\mathbf{P}}_t$  分别表示估计的位置坐标与真实的位置坐标。先将估计的轨迹与真实轨迹通过  $Sim(3)$  变换进行对齐,得到尺度  $s$ ,即可计算尺度误差 SE:

$$SE = 100 \times |1 - s| \% \quad (22)$$

### 4.2 系统评估

系统评估实验在 Rosario 数据集<sup>[24]</sup>上进行,并和当前主流的先进视觉定位与地图构建系统进行了比较,包括 ORB-SLAM2<sup>[4]</sup>, VINS<sup>[6]</sup>, SDSO<sup>[10]</sup>, ORB-SLAM3<sup>[5]</sup> 和 SODSO<sup>[11]</sup>。所有实验均在 Ubuntu18.04 操作系统下完成,硬件设备为 Intel © Core™ i7-10700 @ 2.9 GHz CPU, 16 GB 内存,只使用 CPU 进行计算。为了降低随机性对系统造成的影响,每组实验重复 10 次并统计平均结果。

表 1 列出了各种方法的绝对轨迹误差,最好的结果使用粗体表示, X 表示该系统无法在该序列上完整运行。

表 1 各方法的轨迹误差对比

Table 1 Comparison of trajectory errors of various algorithms

(单位: m)

方法	01	02	03	04	05	06
ORB-SLAM2 <sup>[4]</sup>	2.20	2.78	4.14	2.92	4.16	6.64
VINS <sup>[6]</sup>	6.43	4.45	0.97	1.36	3.57	6.41
ORB-SLAM3 <sup>[5]</sup>	1.84	2.49	4.4	3.05	3.83	6.06
SDSO <sup>[10]</sup>	9.2	1.95	2.22	1.43	3.26	X
SODSO <sup>[11]</sup>	4.58	3.08	1.45	1.34	2.67	3.31
本文方法	<b>0.79</b>	<b>1.62</b>	<b>0.34</b>	<b>0.76</b>	<b>0.89</b>	<b>1.32</b>

从表中可以看出,本文提出的方法在所有序列上都取得了最高的精度。在序列 01 上,所提方法取得了 0.79m 的绝对轨迹误差,该误差相比精度第二的 ORB-SLAM3 下降了 57.1%。这个序列最大的难点是长距离的直线运动和两个 U 型的转弯运动。相比于精度第二的 SDSO,所提方法在序列 02 上的轨迹误差降低了 0.33m。VINS 在序列 03 和序列 04 上取得了不错的精度,这两个序列有较长的直线运动,说明基于光流的方法对于纯直线运动可以得到较好的位姿估计结果。本文提出的方法在最困难的序列 05 和 06 上取得了较大的提升,相比于其他方法,轨迹误差的降低幅度超过 50%。实验过程中发现 ORB-SLAM2 和 ORB-SLAM3 在一些序列上出现了失败的情况,主要原因是农业场景不稳定的特征点增加了特征匹配的难度,且在跟踪丢失后重定位困难。根据

实验统计,本文的方法也出现了位姿估计失败的情况,但是可以通过恢复策略快速得到正确的位姿估计,说明该方法在农业场景下具有更高的鲁棒性。

表 2 比较了各种方法的尺度误差。可以看出,所有方法的尺度误差都在 10% 以内,基于直接法的 SDSO 和 SODSO 相比于基于特征点法的 ORB-SLAM 系统具有更小的尺度误差。而本文的方法在除了 02 以外的序列上都取得了最小的尺度误差,在 01, 03 和 06 序列上的尺度误差下降了一个数量级,说明该方法能够有效地抑制大规模农业场景下的尺度漂移问题,提高了系统的精度。VINS 在 02 序列上取得了较小的尺度误差,但是其轨迹误差较大,主要原因可能是该序列包含 U 形转弯运动,基于光流的方法旋转方向估计精度不高,最终导致整体轨迹偏移。

表 2 各方法的尺度误差对比

Table 2 Comparison of scale errors of various algorithms

(单位: m)

方法	01	02	03	04	05	06
ORB-SLAM2 <sup>[4]</sup>	2.79	4.60	9.54	7.03	4.20	5.23
VINS <sup>[6]</sup>	0.78	<b>0.96</b>	1.87	2.72	1.61	6.41
ORB-SLAM3 <sup>[5]</sup>	2.06	4.64	10.30	7.52	4.48	4.42
SDSO <sup>[10]</sup>	1.81	3.41	4.46	3.40	1.92	X
SODSO <sup>[11]</sup>	1.95	3.35	2.80	3.03	2.75	2.12
本文方法	<b>0.24</b>	2.05	<b>0.60</b>	<b>1.61</b>	<b>1.15</b>	<b>0.29</b>

为了进一步分析农业场景中视觉定位与地图构建的尺度变化,图 2 展示了序列 06 上的尺度估计值。图 3 展示了不同方法在序列 06 上的轨迹对比,整个序列包含 3 次直线运动和 2 次 U 型转弯,可以看到本文方法产生的轨迹更接近真实轨迹。从图 2 中可以看出,当机器人转弯的时候尺度会发生相对较大的变化,这是由于此时的图像内容变化较快,系统更高频率地插入和边缘化关键帧,此时的状态估计的精度下降。在一些纹理稀疏的场景下也同样出现了尺度变化较大的情况,主要是因为更少的地图点用于状态估计,导致产生了更大的尺度误差。

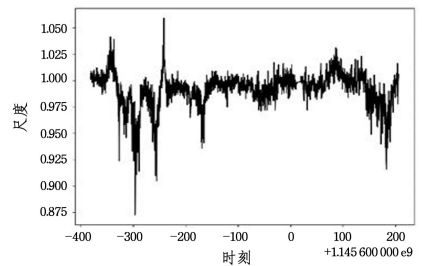


图 2 序列 06 上的尺度估计值

Fig. 2 Scale estimation on sequence 06

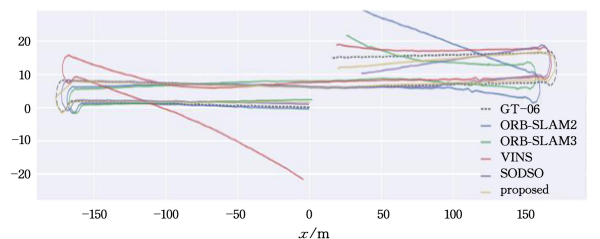


图 3 不同方法在序列 06 上的轨迹对比

Fig. 3 Trajectories comparison on sequence 06

表3中统计了各方法在6个序列上单帧图像处理时间的平均值和最大值,从表中可以看出,基于图像灰度的算法在处理效率上明显高于基于特征的方法,主要原因是其不需要提取特征点以及计算描述子。本文方法的平均效率略高于SD-SO,主要原因是通过点选择算法降低了处理点的密度,提高了位姿估计的效率。本文方法的效率与SODSO相当,两种方法都避免了在滑动窗口优化时构建静态立体匹配的光度误差,减少了计算量。

表3 各方法的单帧数据处理时间对比

Table 3 Comparison of processing time of various algorithms

(单位:ms)

	ORB-SLAM2 <sup>[4]</sup>	VINS <sup>[6]</sup>	ORB-SLAM3 <sup>[5]</sup>	S-DSO <sup>[10]</sup>	SO-DSO <sup>[11]</sup>	本文方法
平均值	26	9	25	13	<b>8</b>	<b>8</b>
最大值	58	46	54	56	32	<b>28</b>

**结束语** 本文主要针对农业场景下视觉定位与地图构建面临的一些挑战,提出了一个适用于该场景的双目视觉定位与地图构建系统,通过加入静态立体匹配点增加了可用的地图点,提高了位姿估计的精度。提出的点选择方法对地图点进行采样,使其分布更均匀,并且移除部分离群点,可以提高系统的运行效率和精度。利用近点进行关键帧判别,可以有效避免场景中远处的大目标对系统的影响,考虑尺度的位姿估计和滑动窗口优化有效地减小了大规模农业场景下视觉定位和地图构建的尺度误差。失败恢复策略使得系统可以在不稳定的运动中进行正确的位姿估计,提高了系统的鲁棒性。实验结果表明,在具有挑战性的农业场景下,本文提出的方法具有更高的精度和鲁棒性。目前系统的整体精度还有一定的提升空间,未来我们准备通过融合视觉和惯性测量单元的数据来进一步提高系统的精度和鲁棒性。

## 参考文献

- [1] VOUGIOUKAS S G. Agricultural Robotics [J]. Annual Review of Control, Robotics, and Autonomous Systems, 2019, 2(1): 365-392.
- [2] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces [C] // Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality. Nara: IEEE, 2007: 225-234.
- [3] PIRE T, FISCHER T, CIVERA J, et al. Stereo parallel tracking and mapping for robot localization [C] // Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg: IEEE, 2015: 1373-1378.
- [4] MUR-ARTAL R, TARDOS J D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras [J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [5] CAMPOS C, ELVIRA R, RODRIGUEZ J J G, et al. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM [J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [6] QIN T, LI P, SHEN S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator [J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.
- [7] NEWCOMBE R A, LOVEGROVE S J, DAVISON A J. DTAM: Dense tracking and mapping in real-time [C] // Proceedings of the International Conference on Computer Vision. Barcelona: IEEE, 2011: 2320-2327.
- [8] FORSTER C, ZHANG Z, GASSNER M, et al. SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems [J]. IEEE Transactions on Robotics, 2017, 33(2): 249-265.
- [9] ENGEL J, KOLTUN V, CREMERS D. Direct Sparse Odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625.
- [10] WANG R, SCHWÖRER M, CREMERS D. Stereo DSO: Large-Scale Direct Sparse Visual Odometry with Stereo Cameras [C] // Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 3923-3931.
- [11] MO J, SATTAR J. Extending Monocular Visual Odometry to Stereo Camera Systems by Scale optimization [C] // Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau: IEEE, 2019: 6921-6927.
- [12] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF [C] // Proceedings of the International Conference on Computer Vision. Barcelona: IEEE, 2011: 2564-2571.
- [13] AUAT CHEEIN F, STEINER G, PEREZ PAINA G, et al. Optimized EIF-SLAM algorithm for precision agriculture mapping based on stems detection [J]. Computers and Electronics in Agriculture, 2011, 78(2): 195-207.
- [14] JUMAN M A, WONG Y W, RAJKUMAR R K, et al. A novel tree trunk detection method for oil-palm plantation navigation [J]. Computers and Electronics in Agriculture, 2016, 128: 172-180.
- [15] MENDES J M, DOS SANTOS F N, FERRAZ N A, et al. Localization based on natural features detector for steep slope vineyards [J]. Journal of Intelligent & Robotic Systems, 2019, 93(3/4): 433-446.
- [16] UNDERWOOD J, WENDEL A, SCHOFIELD B, et al. Efficient in-field plant phenomics for row-crops with an autonomous ground vehicle [J]. Journal of Field Robotics, 2017, 34(6): 1061-1083.
- [17] SA I, LEHNERT C, ENGLISH A, et al. Peduncle detection of sweet pepper for autonomous crop harvesting—combined color and 3-D information [J]. IEEE Robotics and Automation Letters, 2017, 2(2): 765-772.
- [18] WINTERHALTER W, FLECKENSTEIN F V, DORNHEGE C, et al. Crop Row Detection on Tiny Plants With the Pattern Hough Transform [J]. IEEE Robotics and Automation Letters, 2018, 3(4): 3394-3401.
- [19] IMPEROLI M, POTENA C, NARDI D, et al. An effective multi-cue positioning system for agricultural robotics [J]. IEEE Robotics and Automation Letters, 2018, 3(4): 3685-3692.
- [20] DONG J, BURNHAM J G, BOOTS B, et al. 4D crop monitoring: Spatio-temporal reconstruction for agriculture [C] // Proceedings of the IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017: 3878-3885.

- [21] CHEBROLU N, LOTTES P, LIÁBE T, et al. Robot Localization Based on Aerial Images for Precision Agriculture Tasks in Crop Fields[C]//Proceedings of the International Conference on Robotics and Automation. Montreal:IEEE,2019:1787-1793.
- [22] POTENA C, KHANNA R, NIETO J, et al. AgriColMap: Aerial-Ground Collaborative 3D Mapping for Precision Farming [J]. IEEE Robotics and Automation Letters,2019,4(2):1085-1092.
- [23] PRETTO A, ARAVECCHIA S, BURGARD W, et al. Building an Aerial-Ground Robotics System for Precision Farming: An Adaptable Solution [J]. IEEE Robotics & Automation Magazine,2021,28(3):29-49.
- [24] PIRE T, MUJICA M, CIVERA J, et al. The Rosario dataset: Multisensor data for localization and mapping in agricultural environments [J]. The International Journal of Robotics Research,2019,38(6):633-641.



**YU Tao**, born in 1994, Ph.D candidate, is a member of China Computer Federation. His main research interests include visual SLAM and computer vision.



**XIONG Shengwu**, born in 1967, Ph. D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include artificial intelligence, machine learning and evolutionary algorithm.

(责任编辑:柯颖)