DOI:10.19651/j.cnki.emt.2417269

基于残差的场景流动态目标跟踪视觉 SLAM 算法*

刘泽峰 冉 腾 肖文东 袁 亮
(新疆大学智能制造现代产业学院(机械工程学院) 乌鲁木齐 830017)

摘 要: 大多数现有的动态同时定位和地图构建(SLAM)算法简单地移除动态对象,导致帮助系统自身定位和导航的动态对象运动信息的丢失,对于复杂和不断变化的工业环境具有局限性。本研究提出了一种改进的目标跟踪的视觉 SLAM 算法,在进行定位的同时,获得更准确的目标位姿估计。该算法使用背景点进行自身定位,利用细化的光流信息,减少噪点的影响,进行准确的定位,然后结合多项式残差的场景流信息,获得准确的动态目标感知结果,降低算法对目标位姿估计的误差。最后,在公开的 KITTI Tracking 数据集和真实场景上对所提算法进行了评估。实验结果显示,在公共数据集上,所提算法定位效果平均旋转误差(RPER)为 0.027°,平均位移误差(RPET)为 0.069 m。目标位姿估计平均旋转误差为 0.686 97°,平均位移误差 0.103 50 m,具有更好的自定位和动态目标跟踪性能。在真实场景中,所提算法也表现出良好的定位与跟踪性能。

关键词: 同时定位和地图构建;目标跟踪;光流;多项式残差
中图分类号: TP391.4;TN-9 **文献标识码**: A **国家标准学科分类代码**: 520.6030

Residual-based visual SLAM algorithm for dynamic target tracking in scene flow

Liu Zefeng Ran Teng Xiao Wendong Yuan Liang
(College of Intelligent Manufacturing Modern Industry (School of Mechanical Engineering), Xinjiang University, Urumqi 830017, China)

Abstract: Most existing dynamic simultaneous localization and mapping (SLAM) algorithms simply remove dynamic objects, resulting in the loss of dynamic object motion information that aids in the system's own localization and navigation, and have limitations for complex and ever-changing industrial environments. In this paper, we propose an improved visual SLAM algorithm for target tracking that performs localization while obtaining a more accurate estimate of the object's pose. The algorithm uses background points for its own localization, uses refined optical flow information to reduce the effect of noise for accurate localization, and then combines the scene flow information with polynomial residuals to obtain accurate dynamic object sensing results and to reduce the algorithm's error in estimating the object's pose. Finally, the proposed algorithm is evaluated on the publicly available KITTI Tracking dataset and real scenes. The experimental results show that on the public dataset, the proposed algorithm has an average rotation error (RPER) of 0.027° and an average displacement error (RPET) of 0.069 m. The average rotation error of object pose estimation is 0.686 97°, and the average displacement error is 0.103 50 m. The proposed algorithm is able to have a better performance of self-localization and dynamic object tracking. The proposed algorithm also shows excellent localization and tracking performance in real scenarios.

Keywords: simultaneous localization and maping;object tracking;optical flow;polynomial residuals

0 引 言

视觉传感器的同时定位和地图构建(visual simultaneous localization and mapping,VSLAM)能够利用相机捕获的图像进行自身定位和场景重建[1-2]。VSLAM 广泛应用于机器人,自动驾驶和虚拟现实(VR)/增强现实(AR)。ORB-SLAM2[3]、ORB-SLAM3[4] 等传统的 VSLAM 算法假设环境结构是静态的,但是并不符合现实情况。VSLAM 系统在动态环境中的应用已经成为一个重要的研究领域。动态 SLAM 系统利用两种不同的策略来

收稿日期:2024-11-04
* 基金项目:新疆维吾尔自治区自然科学基金(2022D01C673)项目资助

处理环境中的动态对象。首先,采用消除策略将动态对象的特征点作为离群点去除[5]。其次,利用动态对象的信息估计动态对象的位姿。DS-SLAM[6]使用 Segnet[7]实时提供基于 caffe 的像素级语义分割,提取场景中对象的语义信息,然后将其与特征上的运动信息相结合,以过滤掉每帧中的动态对象,从而提高位姿估计的准确性。RDS-SLAM[8]算法利用各种语义分割技术来提取动态对象,并实现关键帧选择策略以获得可用的最新语义信息。DP-SLAM[9]使用采用了一个移动的概率传播模型来检测动态特征点,在贝叶斯概率估计框架下将几何方法和语义分割方法信息融合在一起。DynaSLAM[10]在单目和双目的情况下,使用 CNN 来分割图像中的先验动态对象而不提取它们的特征。在 RGB-D 的情况下,将多视图几何模型和基于深度学习的算法相结合来检测动态目标。PSPNet-SLAM[11]将金字塔结构的 PSPNet 作为语义线程,完成对动态对象的分割和利用最优误差补偿单应性矩阵来提高动态点检测的精度。RS-SLAM[12]针对 PSPNet 分割的初始结果,利用基于环境信息的贝叶斯更新方法进行精炼,再结合深度信息对移动目标进行状态估计。SG-SLAM[13]通过经验阈值提出新的动态目标剔除策略并进行可视化点云与语义地图,提高了动态环境下 SLAM 的鲁棒性。DRV-SLAM[14]结合全局下采样和目标对象数据增强的高密度映射方法,有效减少了密集点云地图的内存占用,从而在不同场景中高效构建大规模密集点云地图,并通过动态剔除不可靠的动态特征点来提高复杂动态环境中的定位精度和鲁棒性。孙龙龙等[15]计算图像特征点的运动矢量,并使用期望最大化方法求解运动矢量角度的高斯混合模型参数,基于前一帧的运动点检测结果剔除当前图像中的运动特征点。然而,同时定位和目标跟踪对智能控制和路线规划等高级任务具有决定作用[16-17]。因此,动态目标感知对于动态场景中的定位和跟踪至关重要。

为满足 AR/VR 等高级任务需求,同时估计相机运动与移动物体成为研究热点[18]。MaskFusion[19]使用实例分割来检测动态区域,并利用直接方法来跟踪对象,可以有效地跟踪场景中的动态对象。然而,传统的多目标跟踪严重依赖于相机定位的准确性。Dynaslam Ⅱ[20]计算像素级语义分割,提取匹配 ORB 特征,然后使用恒速模型通过最小化重投影误差跟踪优化相机和物体位姿。然而,在真实场景下,物体保持恒速比较困难,并且运动物体在相机图像中占据部分较小,提取目标 ORB 特征数量较少,在一定程度上会影响系统性能。VDO-SLAM[21]通过光流信息与场景流来进行目标感知与位姿估计,位姿估计受感知结果影响。Cluster SLAM[22]构建了一个噪声感知的运动亲和矩阵,该矩阵采用聚类来区分这些刚体。然而,Cluster SLAM 的性能取决于前端感知和数据关联的质量。由于感知能力的限制,感知噪声和异常值是不可避免的,对于不同的环境,使用固定阈值或经验阈值会具有局限性。

针对存在的感知噪声以及不同的环境下固定阈值或经验阈值的局限问题,本文提出一种改进的动态目标跟踪视觉 SLAM 算法,使用更为准确的动态目标感知结果,利用光流进行准确的匹配,并结合光流信息进行自身位姿优化,获得准确的定位结果。在此基础上,计算光流的多项式残差,结合场景流进行自适应阈值的目标感知,提高运动目标状态估计准确度,更准确地估计运动目标位姿,提高了算法的定位与跟踪性能。

1 算法介绍

本文提出的基于残差的场景流目标跟踪视觉 SLAM 算法框架如图 1 所示,系统以图像与深度作为输入,经过预处理感知光流信息和掩码信息,通过跟踪线程进行特征提取与目标分类,经过建图线程进行优化,最后输出轨迹图像。

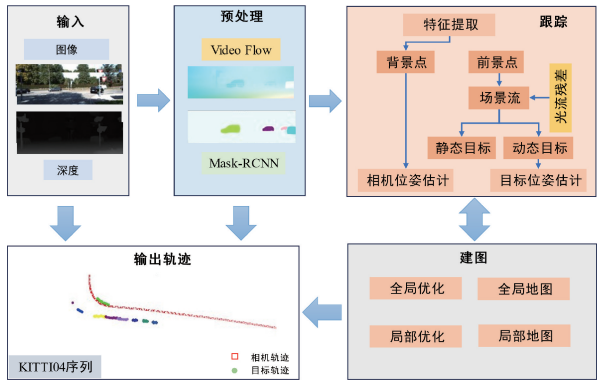

图 1 算法整体框架

Fig. 1 Overall framework of the algorithm

1.1 输入

系统以双目相机或 RGB-D 相机获取图像与深度作为输入。

1.2 预处理

实例分割在语义分割的基础上进一步完善,分离对象的前景和背景,以实现像素级的对象分离。语义分割只会对不同类别的物体进行分割,而实例分割则会进一步分割同一类别中的不同对象实例。本研究使用 Mask R-CNN[23],它是经典的实例级语义分割法来区分场景中的物体。在检测到的图像中,物体只占一小部分区域。因此稀疏光流误差较大,无法保证长期特征跟踪。此外,稀疏光流流可能会降低系统性能,甚至导致跟踪失败。所以,本研究使用密集光流进行模式估计。相比于 PWC-Net[24]光流处理,Video Flow[25]光流不只以两帧图片作为输入,而是充分挖掘和利用多帧数据的线索,显著提升了光流估计的性能。具体地如图 2 所示,Video Flow 光流以三帧相邻的图片作为输入时,采用共享权重的特征编码器获得对应特征图,然后分别建立中间帧与前后两帧的 Cost Volume,采用类似 RAFT[26]的结构,迭代优化光流估计。

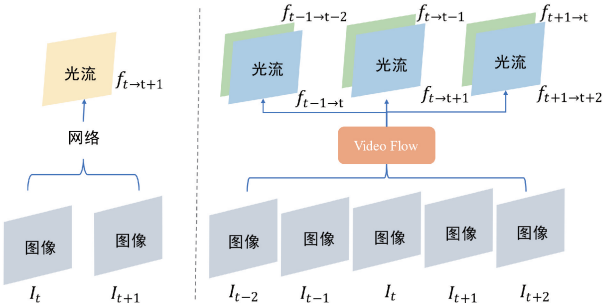

图 2　Video flow 光流获取

Fig. 2　Video flow optical flow acquisition

1.3　跟踪

相机位姿估计：跟踪线程首先进行提取图像特征，将预处理部分中实例分割得到的背景点进行相机位姿估计。对于 3D-2D 检测到的静态匹配点，通过使用运动函数和相机位姿联合方法减少重投影误差来确定相机位姿。为了实现鲁棒的位姿估计，通过比较两个模型的两个模型的内部点数量，在相机位姿初始化中使用内部点数较多的模型。其中一组内点是通过推断上一个相机位置的运动而生成的内部点。而另一组则是通过 P3P[27]算法和 RANSAC 算法计算新的运动变换后生成的内点。

目标位姿估计：在变化环境中，固定阈值的场景流对于目标感知是一个挑战。算法将预处理部分中实例分割得到的前景点利用场景流与光流残差结合的方法，设置自适应阈值，将目标分为动态目标与静态目标，对动态目标跟踪，进行位姿估计。具体地，在获得世界坐标系下相机位姿 ${}^0\boldsymbol{X}_t$ 之后，描述帧 $t-1$ 和 t 之间的 3D 点 ${}^0\boldsymbol{m}_t^i$ 的运动的场景流向量 $\boldsymbol{sf}_t^i$ 可以如文献[28]中方式计算：

$$\boldsymbol{sf}_t^i = {}^0\boldsymbol{m}_{t-1}^i - {}^0\boldsymbol{m}_t^i = {}^0\boldsymbol{m}_{t-1}^i - {}^0\boldsymbol{X}_t^X \boldsymbol{m}_t^i \tag{1}$$

得到场景流向量后，计算基于多项式光流的残差：

$$residual_t^i = \sum_{k=1}^{3} \{(|\boldsymbol{sf}_t^i|)^k/3\} \tag{2}$$

然后计算得到残差的标准差：std，最后通过构建固定阈值 threshold 与标准差之间的函数计算平均值得到自适应阈值：

$$rethreshold = (threshold + std)/2 \tag{3}$$

式中：$rethreshold$ 代表自适应阈值。通过自适应阈值判断物体的运动状态。

联合光流估计与运动估计：相机位姿和物体运动估计都依赖于良好的图像对应关系。由于遮挡、大的相对运动和大的相机对象距离，移动对象上的点的跟踪可能是非常具有挑战性的。本文提出细化光流的估计与运动估计。相机运动产生光流，像素 ${}^{I_t}p_{t-1}^i$ 从图像帧 I_t 到 I_{t-1} 的运动的位移矢量用 ${}^{I_t}\boldsymbol{\varphi}^i \in \boldsymbol{IR}^2$ 表示，并且由式(4)给出：

$${}^{I_t}\boldsymbol{\varphi}^i = {}^{I_t}\widetilde{\boldsymbol{p}}_t^i - {}^{I_{t-1}}\widetilde{\boldsymbol{p}}_{t-1}^i \tag{4}$$

式中：${}^{I_t}\widetilde{\boldsymbol{p}}_t^i$ 是 ${}^{I_t}\widetilde{\boldsymbol{p}}_{t-1}^i$ 在 I_t 中的对应。对于相机位姿，通过最小化重新投影误差来估计相机位姿：

$$\boldsymbol{e}_i({}^0\boldsymbol{X}_t) = {}^{I_t}\widetilde{\boldsymbol{p}}_t^i - \pi({}^0\boldsymbol{X}_t^{-1}{}^0\boldsymbol{m}_{t-1}^i) \tag{5}$$

本文通过李代数 $\boldsymbol{x}_t \in \mathrm{se}(3)$ 的元素来参数化 SE(3) 相机位姿：

$${}^0\boldsymbol{X}_t = \exp({}^0\boldsymbol{x}_t) \tag{6}$$

并定义 ${}^0\boldsymbol{x}_t^\nu \in IR^6$，其中 νee 算子是从 se(3) 到 $\boldsymbol{IR}^6$ 的映射。使用 SE(3) 李代数参数化，将式(6)代入式(5)，最小二乘函数的解由式(7)给出：

$${}^0\boldsymbol{x}_t^{*\nu} = \mathrm{argmin}\sum_i^{n_b} \rho_h(\boldsymbol{e}_i^T({}^0\boldsymbol{x}_t)\boldsymbol{\Sigma}_p^{-1}\boldsymbol{e}_i({}^0\boldsymbol{x}_t)) \tag{7}$$

式中：ρ_h 是 *Huber* 函数，而 $\boldsymbol{\Sigma}_p$ 是与重投影误差相关的协方差矩阵。n_b 是对于连续帧之间的所有可见 3D-2D 静态背景点对应个数。估计的相机位姿由 ${}^0\boldsymbol{X}_t^* = \exp({}^0\boldsymbol{x}_t^*)$ 给出，并且使用 *Levenberg-Marquardt* 算法来求解式(7)。联合光流估计与运动估计，对于相机位姿估计，考虑到式(4)和(5)中的误差项被重新公式化为：

$$\boldsymbol{e}_i({}^0\boldsymbol{X}_t, {}^{I_t}\boldsymbol{\varphi}^i) = {}^{I_{t-1}}\boldsymbol{p}_{t-1}^i + {}^{I_t}\boldsymbol{\varphi}^i - \pi({}^0\boldsymbol{X}_t^{-1}{}^0\boldsymbol{m}_{t-1}^i) \tag{8}$$

应用 SE(3) 元素的李代数参数化，通过最小化成本函数获得最优解：

$$\{{}^0\boldsymbol{x}_t^{*\nu}, {}^{I_t}\boldsymbol{\Phi}_t^*\} = \mathrm{argmin}\sum_i^{n_b}\left\{\begin{matrix}\rho_h(\boldsymbol{e}_i^T({}^{I_t}\boldsymbol{\varphi}^i)\boldsymbol{\Sigma}_\varphi^{-1}\boldsymbol{e}_i({}^{I_t}\boldsymbol{\varphi}^i)) + \\ \rho_h\left(\boldsymbol{e}_i^T({}^0\boldsymbol{x}_t, {}^{I_t}\boldsymbol{\varphi}^i)\sum\nolimits_p^{-1}\boldsymbol{e}_i({}^0\boldsymbol{x}_t, {}^{I_t}\boldsymbol{\varphi}^i)\right)\end{matrix}\right\} \tag{9}$$

式中：$\rho_h(\boldsymbol{e}_i^T({}^{I_t}\boldsymbol{\varphi}^i)\boldsymbol{\Sigma}_\varphi^{-1}\boldsymbol{e}_i({}^{I_t}\boldsymbol{\varphi}^i))$ 是正则化项。

$$\boldsymbol{e}_i({}^{I_t}\boldsymbol{\varphi}^i) = {}^{I_t}\hat{\boldsymbol{\varphi}}^i - {}^{I_t}\boldsymbol{\varphi}^i \tag{10}$$

式中：${}^{I_t}\hat{\boldsymbol{\varphi}}^i$ 是通过经典方法或基于学习的方法获得的初始光流，$\boldsymbol{\Sigma}_\varphi$ 是相关的协方差矩阵。

2　实验结果与分析

本文所提算法在 KITTI Tracking 公共数据集上验证有效性，并且与 ORBSLAM2，DynaSLAM，DynaSLAM Ⅱ，VDO-SLAM 相关工作进行对比。此外，本文在真实场景下也进行了测试。算法测试主要使用 Linux 18.04 系统，Intel Core i5-12600kf CPU，NVIDIA GeForce GTX 2080Ti GPU 和 32 G RAM 等硬件。根据先前的工作，本文主要使用相对旋转误差(RPER)，单位为°，相对旋转误差(RPET)，单位为 m，来评估定位结果和目标跟踪性能。最优结果进行加粗表示。

2.1　在 KITTI Tracking 数据集的性能

1)相机位姿估计。本文在 7 个 KITTI Tracking 公共数据集上验证算法的有效性。表 1 显示了相机定位的详细对比结果。ORB-SLAM 2 算法由于受到动态目标的干扰，在定位精度方面明显不如本文的算法。DynaSLAM 为 ORB-SLAM 2 添加了动态剔除模型。本文所提的算法利用动态目标提供的丰富线索来约束定位，因此一些序列的

定位精度略高于 DynaSLAM。与合并动态对象跟踪的其他现有技术 SLAM 系统相比,所提出的算法在所评估所有序列上获得了更低的旋转误差,平均旋转误差为 0.027°。实验结果表明,本文提出的算法在 4 个序列上的 RPET 优于现有其他算法。本文认为,差异的产生与后端优化有关。注意,由于所提出的感知模块,所提出的系统相对于 VDO-SLAM 具有显著的平均性能改进。实验结果表明,该算法在动态环境下具有良好的稳定性。

表 1 KITTI Tracking 数据集相机位姿估计对比

Table 1 Comparison of camera position estimates for the KITTI Tracking dataset

序列	ORB-SLAM2		DynaSLAM		DynaSLAM Ⅱ		VDO-SLAM		本文方法	
	RPER/(°)	RPET/m	RPER/(°)	RPET/m	RPER/(°)	RPET/m	RPER/(°)	RPET/m	RPER/(°)	RPET/m
00	0.06	**0.04**	0.06	**0.04**	0.06	**0.04**	0.07	0.06	**0.04**	**0.04**
01	0.04	**0.05**	0.04	**0.05**	0.04	**0.05**	0.04	0.12	**0.03**	0.10
02	0.03	**0.04**	0.03	**0.04**	**0.02**	**0.04**	**0.02**	**0.04**	**0.02**	**0.04**
03	0.04	0.07	0.04	0.04	**0.04**	**0.04**	0.03	0.08	**0.03**	0.09
04	0.06	**0.07**	0.06	**0.07**	0.06	**0.07**	0.05	0.11	**0.04**	0.11
05	0.03	**0.06**	0.03	**0.06**	0.03	**0.06**	0.02	0.09	**0.01**	0.09
06	0.04	0.02	0.04	0.02	0.04	0.02	0.05	0.09	**0.02**	**0.01**
平均值	0.043	0.050	0.043	**0.046**	0.041	**0.046**	0.040	0.084	**0.027**	0.069

2)目标位姿估计。表 2 给出了与 VDO-SLAM 相比的目标位姿估计结果,可以清楚地看出,所提出的算法获得了优于现有技术的 VDO-SLAM 的对象跟踪性能改进。目标位姿估计的 RPER 和 RPET 在大多数序列均有提升,平均旋转误差为 0.686 97°,平均位移误差为 0.103 50 m。其原因是在筛选动态对象时,首先进行目标运动状态的决策,减少了噪点,其次是感知噪声的减少。实验结果表明,该算法能够在动态环境下获得比 VDO-SLAM 算法更准确、更有竞争力的目标位姿估计。

3)定性分析。如图 3 为在 KITTI06 数据集定性结果,依次表示原始图像、实例分割结果、细化光流结果。此外,图 4 依次为本文算法在 KITTI00、KITTI04、KITTI06 数据集运行得到的轨迹地图,更为直观证明所提算法的有效性,其中红色方块代表相机运动轨迹,彩色圆圈表示目标的运动轨迹。为了进一步证明所提出的算法在场景中的轨迹的有效性,所提出的算法的相机轨迹与 ORB-SLAM2、VDO-SLAM 对比结果如图 5 所示。由图 5 可知,ORB-SLAM2、VDO-SLAM 随着行驶距离的增加,漂移逐渐增大,所提算法在跟踪过程均与真实情况相接近。图 6 表示全局对准之后的 APE,垂直数轴表示行驶的距离,由图 5 可知,算法在经过转弯后,轨迹漂移会增加,对比算法 ORB-SLAM2、VDO-SLAM 表现尤为明显。

表 2 KITTI Tracking 数据集目标位姿估计对比

Table 2 Comparison of object pose estimation on KITTI Tracking dataset

序列	VDO-SLAM		本文方法	
	RPER/(°)	RPET/m	RPER/(°)	RPET/m
00	1.052 0	0.107 7	**0.892 2**	**0.068 8**
01	0.905 1	0.157 3	**0.638 6**	**0.092 6**
02	1.235 9	0.280 1	**0.980 8**	**0.108 5**
03	0.291 9	0.096 5	**0.237 2**	**0.095 5**
04	**0.828 8**	0.193 7	0.928 7	**0.174 5**
05	0.370 5	0.114 0	**0.307 0**	**0.095 8**
06	1.080 3	0.115 8	**0.824 3**	**0.088 8**
平均值	0.823 50	0.152 16	**0.686 97**	**0.103 50**

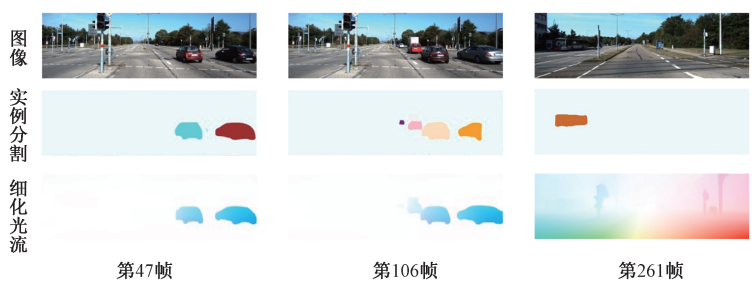

图 3 KITTI06 数据集定性结果

Fig. 3 Qualitative results for the KITTI06 dataset

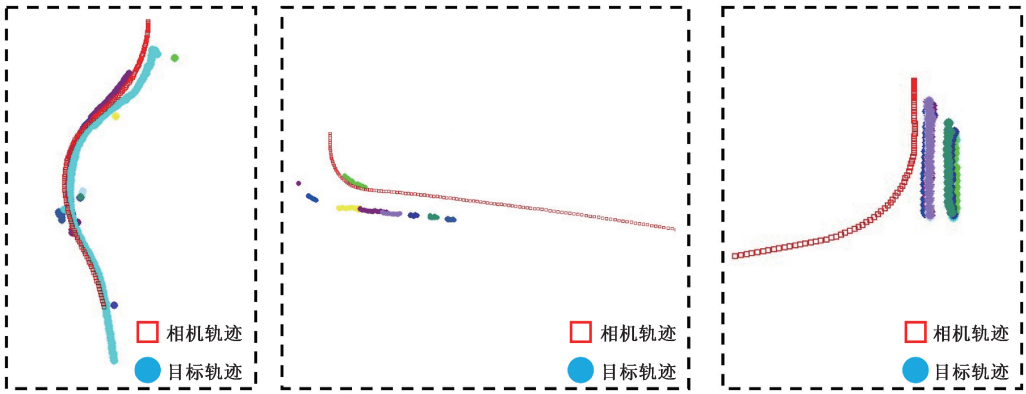

图 4　算法在数据集上运行的轨迹结果

Fig. 4　Trajectory results of the algorithm running on the dataset

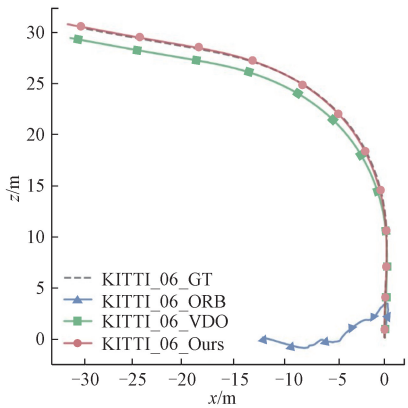

图 5　相机轨迹对比结果

Fig. 5　Camera trajectory comparison results

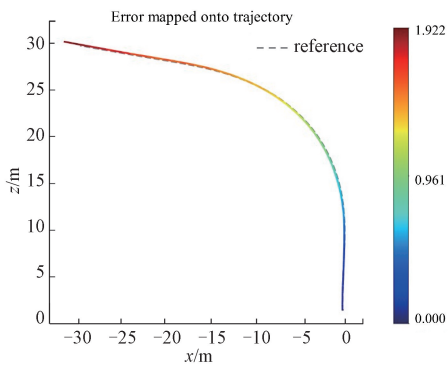

图 6　本文算法相机的绝对轨迹结果

Fig. 6　Absolute trajectory results for the cameras of the algorithms in this paper

4)时间消耗评估。表 3 为算法在 KITTI 数据集 02、03、04 序列的各模块运行时间。其中相机位姿估计平均消耗 13.940 ms，地图更新时间为 4.443 ms，Bundle Adjustment (BA)优化时间为 132.170 ms。

表 3　主要模块运行时间

Table 3　Main module runtime　　ms

序列	02	03	04	平均值
相机位姿估计	16.70	12.38	12.74	13.940
地图更新	4.33	4.67	4.33	4.443
BA 优化	142.39	99.73	154.39	132.170

2.2　真实场景测试

为了验证该算法的有效性，实验设置在 3 个道路环境中进行。涵盖了不同的光照、路况和障碍物遮挡条件，全面反映系统性能。实验使用一个配备有 ZED 相机的机器人作为主动 SLAM 探索器来捕获 RGB 图像和深度图像。该机器人还配备了一个组合的全球定位系统(GPS)/惯性测量单元(IMU)模块，来获取相机轨迹的真实值。实验装置如图 7 所示。其中 GPS/IMU 定位精度＜2.5 m，航向精度为 1°。

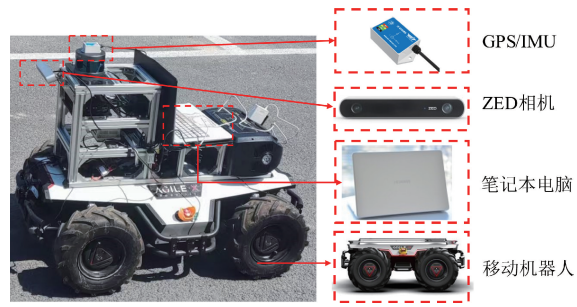

图 7　移动机器人

Fig. 7　Mobile robots

表 4 真实场景相机位姿和目标位姿结果对比是本文的算法在真实世界场景上运行与 VDO-SLAM 算法的定量比较。在定位方面，VDO-SLAM 的平均相对旋转误差为 1.149 46°，平均相对位移误差 1.225 43 m，相比于所提算法的平均相对旋转误差 0.872 30°，平均相对平移误差 1.018 40 m，该算法具有较小的定位误差，定位性能明显提升。

目标位姿估计方面，VDO-SLAM 的平均相对旋转误

表 4 真实场景相机位姿和目标位姿结果对比

Table 4 Comparison of real scenarios camera pose and object pose results

序列	VDO-SLAM				本文方法			
	Camera		Object		Camera		Object	
	RPER/(°)	RPET/m	RPER/(°)	RPET/m	RPER/(°)	RPET/m	RPER/(°)	RPET/m
0611	1.125 5	1.226 2	1.780 3	1.809 4	**0.827 8**	**0.877 3**	**0.644 6**	**0.641 9**
1025	1.587 8	1.408 6	1.967 7	2.782 4	**1.111 1**	**1.212 0**	**1.283 3**	**1.177 1**
1211	0.735 1	1.041 5	0.268 3	1.504 2	**0.678 0**	**0.965 9**	**0.211 0**	**1.486 9**
平均值	1.149 46	1.225 43	1.338 76	2.032 00	**0.872 30**	**1.018 40**	**0.7129 6**	**1.101 96**

差为 1.338 76°,平均相对平移误差为 2.032 00 m。而所提算法的平均相对旋转误差 0.712 96°,平均相对平移误差 1.101 96 m,对比 VDO-SLAM,目标位姿估计更为准确,具有更好的目标跟踪性能。实验结果表明所提出的算法在定位与跟踪方面更具有竞争力。图 8 为真实场景的定性结果,可以看出,实例分割与光流细化有着较好的结果,但是在输出轨迹方面,存在一定漂移。本文认为,这可能与传感器的精度、安装位置和道路状况密切相关。特别是由于建筑物的遮挡,GPS 信号较弱,导致位姿估计误差较大。但是总体误差在可接受范围之内。该算法估计的相机位姿总体上非常接近真实世界中的所有序列。实验结果表明,该系统可以在真实的动态环境中有效地运行。

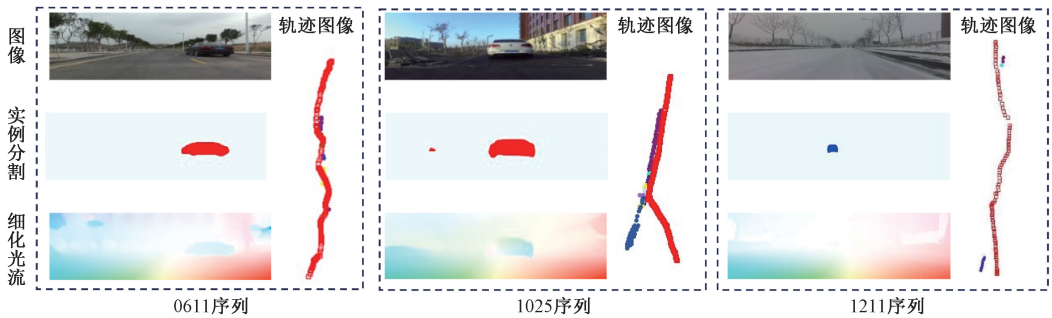

图 8 真实场景定性结果

Fig. 8 Real scenarios qualitative results

3 结 论

在本文中,本文提出一种基于残差的场景流目标运动监测的动态目标跟踪视觉 SLAM 算法,可以在动态环境中有效运行。主要思想是通过预处理得到的实例分割与细化光流的信息,利用实例分割得到背景信息用于相机定位,使用细化的结光流信息来优化相机位姿,提高了算法定位性能。同时细化的光流信息利用基于光流残差与场景流结合的方法构建函数得到自适应阈值来判断目标的运动状态,减少感知噪声的同时,减少了固定阈值或经验阈值的局限性对算法的影响,有更具竞争力的目标跟踪效果。实验结果表明,所提算法有着良好的定位效果与跟踪性能,并且可以在真实场景中有效运行。

参考文献

[1] ESPARZA D, FLORES G. The STDyn-SLAM: A stereo vision and semantic segmentation approach for VSLAM in dynamic outdoor environments[J]. IEEE Access, 2022, 10: 18201-18209.

[2] 冯洲,续欣莹,郑宇轩,等. 动态场景下基于实例分割和三维重建的多物体单目 SLAM[J]. 仪器仪表学报, 2023,44(8):51-62.
FENG ZH, XU X Y, ZHENG Y X, et. al. Multi-object monocular SLAM based on instance segmentation and 3D reconstruction in dynamic scene[J]. Chinese Journal of Scientific Instrument, 2023,44(8):51-62.

[3] 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, 2017, 33(5): 1255-1262.

[4] CAMPOS C, ELVIRA R, RODRÍGUEZ 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.

[5] 史涛, 校诺政, 丁垚, 等. 动态场景下融合改进 YOLOv7 的视觉 SLAM 算法[J]. 国外电子测量技术, 2024, 43(7):90-96.
SHI T, XIAO N ZH, DING Y, et al. Visual SLAM algorithm for fusing improved YOLOv7 in dynamic scenes [J]. Foreign Electronic Measurement Technology, 2024, 43(7):90-96.

[6] YU CH, LIU Z X, LIU X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C].

IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), 2018: 1168-1174.
[7] BADRINARAYANAN V, KENDALL A, CIPOLLA R, SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
[8] LIU Y B, MIURA J. RDS-SLAM: Real-time dynamic SLAM using semantic segmentation methods [J]. IEEE Access, 2021, 9: 23772-23785.
[9] LI AO, WANG J K, XU M, et al. DP-SLAM: A visual SLAM with moving probability towards dynamic environments [J]. Information Sciences, 2021, 556: 128-142.
[10] BESCOS B, FÁCIL J M, CIVERA J, et al. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076-4083.
[11] LONG X D, ZHANG W W, ZHAO B. PSPNet-SLAM: A semantic SLAM detect dynamic object by pyramid scene parsing network[J]. IEEE Access, 2020, 8: 214685-214695.
[12] RAN T, YUAN L, ZHANG J B, et al. RS-SLAM: A robust semantic SLAM in dynamic environments based on RGB-D sensor[J]. IEEE Sensors Journal, 2021, 21(18): 20657-20664.
[13] CHENG SH H, SUN CH H, ZHANG SH J, et al. SG-SLAM: A real-time RGB-D visual SLAM toward dynamic scenes with semantic and geometric information [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 72: 1-12.
[14] JI Q, ZHANG Z K, CHEN Y F, et al. DRV-SLAM: An adaptive real-time semantic visual SLAM based on instance segmentation toward dynamic environments [J]. IEEE Access, 2024, 12: 43827-43837.
[15] 孙龙龙，江明，焦传佳. 基于运动矢量的改进视觉 SLAM 算法[J]. 电子测量与仪器学报, 2020,34(9): 23-31.
SUN L L, JIANG M, JIAO CH J. Improved visual SLAM algorithm based on the motion vector [J]. Journal of Electronic Measurement and Instrumentation, 2020,34(9): 23-31.
[16] BALLESTER I, FONTÁN A, CIVERA J, et al. DOT: Dynamic object tracking for visual SLAM[C]. IEEE International Conference on Robotics and Automation(ICRA), 2021: 11705-11711.
[17] ZHANG J, HENEIN M, MAHONY R, et al. Robust ego and object 6-DoF motion estimation and tracking[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020: 5017-5023.
[18] ZHANG H X, WANG D Y, HUO J. A visual-inertial dynamic object tracking SLAM tightly coupled system[J]. IEEE Sensors Journal, 2023, 23(17): 19905-19917.
[19] RUNZ M, BUFFIER M, AGAPITO L. MaskFusion: Real-time recognition, tacking and reconstruction of multiple moving objects [C]. IEEE International Symposium on Mixed and Augmented Reality(ISMAR), 2018: 10-20.
[20] BESCOS B, CAMPOS C, TARDÓS J D, et al. DynaSLAM II: Tightly-coupled multi-object tracking and SLAM [J]. IEEE Robotics and Automation Letters, 2021, 6(3): 5191-5198.
[21] ZHANG J, HENEIN M, MAHONY R, et al. VDO-SLAM: A visual dynamic object-aware SLAM system[J]. ArXiv preprint ArXiv: 2005.11052, 2020.
[22] HUANG J H, YANG SH, ZHAO Z SH, et al. Cluster SLAM: A SLAM backend for simultaneous rigid body clustering and motion estimation [C]. IEEE/CVF International Conference on Computer Vision, 2019: 5875-5884.
[23] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. IEEE International Conference on Computer Vision, 2017: 2961-2969.
[24] SUN D Q, YANG X D, LIU M Y, et al. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8934-8943.
[25] SHI X Y, HUANG ZH Y, BIAN W K, et al. Videoflow: Exploiting temporal cues for multi-frame optical flow estimation[C]. IEEE/CVF International Conference on Computer Vision, 2023: 12469-12480.
[26] TEED Z, DENG J. RAFT: Recurrent all-pairs field transforms for optical flow[C]. Computer Vision-ECCV, 2020: 402-419.
[27] KE T, ROUMELIOTIS S I. An efficient algebraic solution to the perspective-three-point problem[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7225-7233.
[28] LYU ZH Y, KIM K, TROCCOLI A, et al. Learning rigidity in dynamic scenes with a moving camera for 3D motion field estimation[C]. European Conference on Computer Vision(ECCV), 2018: 468-484.

作者简介

刘泽峰，硕士研究生，主要研究方向为移动机器人定位与导航、计算机视觉。
E-mail:719120663@qq.com

冉腾(通信作者)，博士，副教授，硕士生导师，主要研究方向为移动机器人定位与导航、智能机器人技术、机器视觉与图像处理。
E-mail:rantengsky@163.com

肖文东，博士，讲师，硕士生导师，主要研究方向为移动机器人自主导航、深度强化学习、机器视觉与图像处理。
E-mail:xwendong@xju.edu.cn

袁亮，博士，教授，博士生导师，主要研究方向为智能机器人技术、机器视觉与图像处理、数字孪生、工业物联网。
E-mail:lyuan@sjtu.edu.cn