全球定位系统
GNSS World of China
• 中国科学引文数据库（CSCD）
• 中文科技期刊数据库
• 中国核心期刊（遴选）数据库
• 日本科学技术振兴机构数据库（JST）
• 中国学术期刊（网络版）（CNKI）
• 中国学术期刊综合评价数据库（CAJCED）
• 中国超星期刊域出版平台

联合目标检测与深度信息的动态特征点去除方法

叶睿馨，张令文，陈　佳，乔尚兵，朱　颖

Dynamic feature point removal method with joint target detection and depth information

YE Ruixin, ZHANG Lingwen, CHEN Jia, QIAO Shangbing, and ZHU Ying

引用本文:
叶睿馨，张令文，陈佳，等．联合目标检测与深度信息的动态特征点去除方法[J]．全球定位系统，2024，49(3)：1-7．DOI：10.12265/j.gnss.2024015
YE Ruixin, ZHANG Lingwen, CHEN Jia, et al. Dynamic feature point removal method with joint target detection and depth information[J]. Gnss World of China, 2024, 49(3): 1-7. DOI: 10.12265/j.gnss.2024015

在线阅读 View online: https://doi.org/10.12265/j.gnss.2024015

DOI: 10.12265/j.gnss.2024015

联合目标检测与深度信息的动态特征点去除方法

叶睿馨[1], 张令文[1], 陈佳[1,2], 乔尚兵[3], 朱颖[3]

(1. 北京交通大学电子信息工程学院, 北京 100044; 2. 鹏城实验室, 深圳 518000; 3. 中国信息通信研究院, 北京 100191)

摘　要: 针对在动态环境中, 视觉定位系统的定位精度和鲁棒性容易受到动态特征点影响的问题, 提出了一种联合目标检测与深度信息的动态特征点去除方法. 引入 YOLOv7 目标检测网络快速获得当前图像帧的目标类别及位置信息, 加入坐标注意力 (coordinate attention, CA) 机制优化深度学习模型, 提升网络目标检测精度. 此外, 提出了一种利用深度信息和对极几何约束的动态特征点优化策略. 有效剔除了动态特征点, 同时保留了尽量多的静态点, 从而降低了动态点对系统定位精度和鲁棒性的影响. 在公开的数据集 TUM 上进行实验验证. 结果表明: 与 ORB-SLAM2 (oriented fast and rotated brief-SLAM) 相比, 所提方案在定位精度和鲁棒性上有明显优势. 同时与动态同步定位和地图构建 (dyna simultaneous localization and mapping, DynaSLAM) 相比, 定位精度基本持平, 但在运行速度上实现了显著提升.

关键词: 动态特征点剔除; 目标检测; 深度学习; 动态场景; 视觉定位

中图分类号: P228.4　　**文献标志码:** A　　**文章编号:** 1008-9268(2024)03-0001-07

0　引　言

视觉定位是通过分析相机捕捉的图像来推算物体当前位置的一种技术, 主要有视觉里程计 (visual odometry, VO) 和同步定位地图构建 (simultaneous localization and mapping, SLAM) 两种实现方法. VO 是一个根据图像之间的变换关系估算出相机运动的过程, 也是基于视觉同步定位和建图 (visual simultaneous localization and mapping, VSLAM) 系统中承担前端定位估计的模块. MUR-Artal[1] 等在 2015 年提出了一个经典的 VSLAM 算法框架, 称为 ORB-SLAM (oriented fast and rotated brief-SLAM). 该视觉定位系统同时具备单目、双目、RGB-D 相机的接口, 并创新性地使用了跟踪、局部建图和回环检测三个线程, 在处理速度和建图精度上取得了较好的效果. 该团队之后提出了 ORB-SLAM2[2] 算法, 因其实时的 CPU 性能和鲁棒性, 成为目前应用最为广泛的 VSLAM 解决方案之一. 然而 ORB-SLAM 系列严重依赖于环境特征, 在没有纹理特征和动态场景中难以获得足够多的特征点, 以致系统难以进行初始化.

传统的视觉定位主要依赖点、线、面等基础几何信息, 在实时性方面可以达到较高的水平. 但是, 当面临光照变化、纹理缺失或动态物体繁多的场景时, 其定位精度和鲁棒性会急剧下降. 于是, 相关研究将深度学习引入视觉定位系统, 借助深度学习强大的图像处理能力进行语义分割、目标检测或实例分割以获取更高层次的特征信息, 利用检测结果剔除场景中动态特征的干扰, 提升系统对周围环境的感知能力、有效改善视觉定位中的不确定性和累积漂移, 进而在不同程度上提高定位精度和鲁棒性. 文献 [3] 分析了常用于视觉定位的语义分割、目标检测和实例分割网络, 结果表明目标检测虽在精度上稍弱, 但具有更高的实时性和更低的硬件依赖, 满足视觉定位系统的实际应用需求. 文献 [4-6] 分别提出了 DynaSLAM、DS-SLAM 和 RS-SLAM 动态 VSLAM 方案. 三者均基于 ORB-SLAM2, 结合 Mask-RCNN、SegNet、PsPNet 语义分割网络来滤除动态物体. 尽管其准确率较高, 但这些方案由于使用了语义分割网络, 耗时巨大, 实时性难以满足实际需求. 张梦珠等[7] 将 YOLOv3 目标检测结果用于建图, 利用语义点云信息辅助位姿估计. 刘胤真等[8] 利用 GhostNet 目标检测网络和对极几何约束来剔除动态特征点. 高逸等[9] 则在 YOLOv5

收稿日期: 2024-01-22

资助项目: 国家重点研发计划项目 (2022YFB2902400); 鹏城实验室重大攻关项目 (PCL2023A06)

通信作者: 陈佳 E-mail: chenjia@bjtu.edu.cn

目标检测的基础上添加了动态物体漏检方法，进一步改善系统在极端情况下的表现. 潘海鹏等[10]提出了一种多尺度的特征融合网络 MulAttenNet 进行语义分割，然后计算从参考帧到当前帧的运动差和深度差，以预测特征点的动态概率进行动态点剔除. 付豪等[11]将语义和光流法结合进一步判断物体的运动特性. 视觉定位研究虽发展迅速，但平衡系统的实时性、精度和鲁棒性仍是目前最大的挑战，尤其是在有动态物体干扰的场景下. 因此，本文提出了一种联合目标检测和深度信息的动态特征点去除方法，以有效平衡系统的实时性、定位精度和鲁棒性. 首先使用引入坐标注意力 (coordinate attention, CA) 机制的 YOLOv7 目标检测网络获取图像的目标检测结果；然后根据物体类别得到先验的动态性划分，对动态、潜在动态的物体框进行特征点检测；最后利用深度信息和对极几何约束剔除其中的动态特征点、保留尽量多的静态点，仅利用静态特征点进行相机位姿估计.

1 改进的视觉定位系统

1.1 系统整体结构

为了提升系统的定位效率和鲁棒性，提出了一种联合目标检测和深度信息的动态特征点去除方法，整体系统流程如图 1 所示. 首先，使用 RGB-D 相机获取 RGB 图像及相应的深度图像. 然后，从 RGB 图像中提取当前帧的 ORB 特征点. 同时，为了充分利用特征的位置信息，并提高目标检测的精度，引入 CA 机制改进 YOLOv7 网络. 根据目标检测结果检测 ORB 动态特征点，即判断每个矩形框的类别是否为动态. 考虑到在动态的矩形框中不仅包含了动态物体上的动态特征点，还可能包含静态特征点. 而动态特征点会给相机观测数据带来误差，进而影响定位的精度和系统的鲁棒性. 因此，提出一种利用深度信息和对极几何约束的动态特征点剔除优化算法. 从深度图像中获取深度信息，利用深度信息的随机抽样一致 (random sample consensus, RANSAC) 方法和自适应阈值的 RANSAC(optimized random sample consensus, ORSA) 算法分离出动态矩形框中的动态点和静态点，从而有效剔除动态特征点，最大化地保留静态特征点，以提高后续相机位姿估计的准确度. 总之，该视觉定位系统不仅通过引入 CA机制改进 YOLOv7 目标检测网络提升了系统的定位效率. 而且利用深度信息进一步剔除动态特征点，提高了系统的定位精度和鲁棒性.

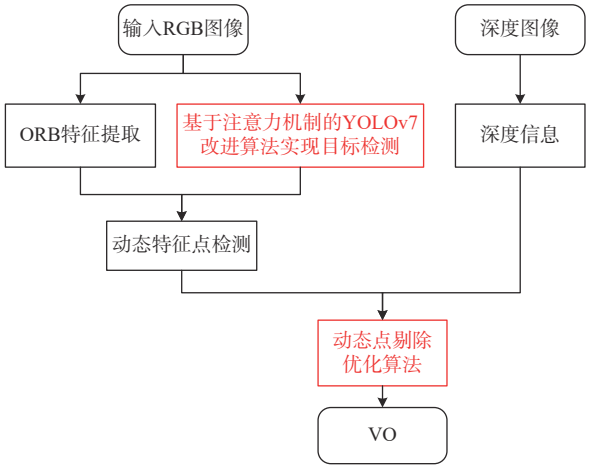


图 1 系统整体流程图

1.2 基于注意力机制的 YOLOv7 目标检测网络

目标检测实现了对图像中的物体进行定位和分类. 早期出现的方法多为两阶段检测，以 R-CNN、Fast R-CNN 方法为代表. YOLO 的出现一步到位，完成了物体的分类和定位，极大地提高了目标检测的精度和速度. 其核心思想是将图像划分为若干个网格，每个网格对区域内的目标进行分类和定位，预测多个边界框，并计算目标中心点的坐标位置(x,y)、矩形框的宽度w和高度h，以及置信度分数. YOLOv7 通过对算法的改进，进一步提高了检测的精度，因此将其引入当前的视觉定位系统. 然而网络在提取图像局部特征信息时，缺乏对特征重要性的理解，且全局池化会丢失特征的空间关联信息. 为解决上述问题，提高 YOLOv7 的目标检测性能，本文在主干网络中加入了 CA 机制[12]，以实现更加精准的目标检测.

CA 是一种轻量化的注意力机制，通过将位置信息嵌入到通道注意中，并将二维的特征编码过程分解为水平和垂直方向的两个一维过程，得到了具有精确位置信息的通道关系和长距离依赖关系[12]，从而提升了网络在目标检测任务中的性能. 如图 2 所示，YOLOv7 网络主要由 Backbone、Neck、Prediction 三部分组成，在 Backbone 部分使用不同大小的卷积核和池化采样对尺寸为 640×640×3 的图像进行特征提取，得到三种尺寸的特征，分别为 80×80×128, 40×40×256, 20×20×1 024. 在此阶段加入 CA 机制以增强特征表达.

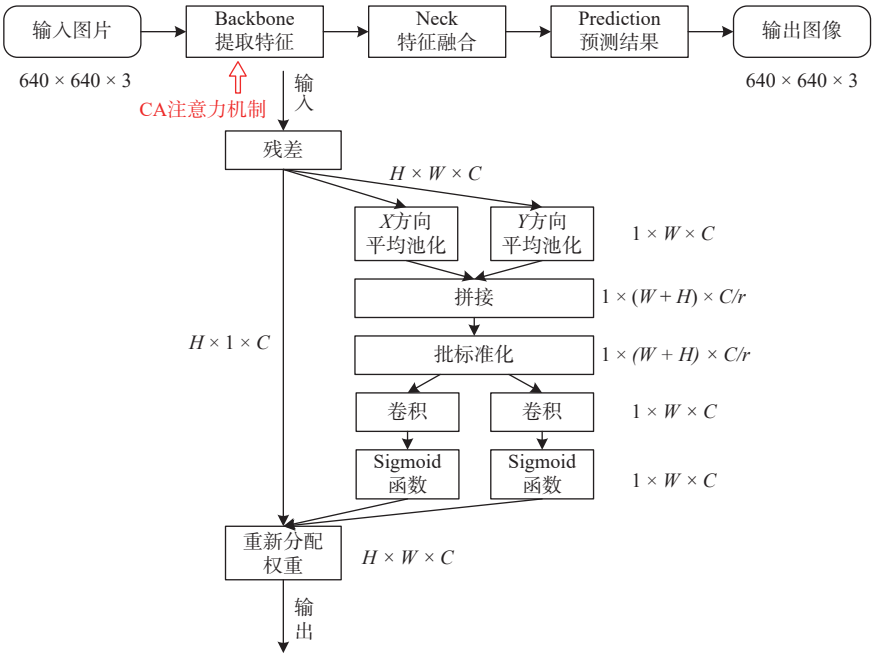


图 2　引入 CA 机制的 YOLOv7 目标检测流程

输入特征向量 $\boldsymbol{X}=[x_1,x_2,\cdots,x_C]\in\mathbb{R}^{C\times W\times H}$ 由 C 个通道值组成，W、H 分别表示宽和高. 经过两个方向的池化分解嵌入坐标信息，得到输出：

$$z_c^h(h)=\frac{1}{W}\sum_{0\leqslant i<W}x_c(h,i) \tag{1}$$

$$z_c^w(w)=\frac{1}{H}\sum_{0,j<H}x_c(j,w) \tag{2}$$

式中，$x_c(h,i)$、$x_c(j,w)$ 分别表示高度、宽度为 h、w 的特征分量，其第 c 个通道在水平、竖直方向上第 i,j 个位置的特征值. 然后进行特征级联、批标准化和卷积变换 F_1，得到中间特征向量 $\boldsymbol{f}\in\mathbb{R}^{C/r\times(H+W)}$，如式 (3) 所示. r 为下采样比例，用于压缩通道数、降低计算量. δ 为非线性激活函数 ReLU.

$$\boldsymbol{f}=\delta(F_1([\boldsymbol{z}_h,\boldsymbol{z}_w])) \tag{3}$$

为了恢复被压缩的通道数，将 $\boldsymbol{f}$ 拆分为 $\boldsymbol{f}^h\in\mathbb{R}^{C/r\times H}$ 和 $\boldsymbol{f}^w\in\mathbb{R}^{C/r+W}$，并进行卷积 F_h 和 F_w，如式 (4) 所示，σ 表示 sigmoid 激活函数，

$$\begin{aligned}\boldsymbol{g}^h&=\sigma(F_h(\boldsymbol{f}^h))\\ \boldsymbol{g}^w&=\sigma(F_w(\boldsymbol{f}^w))\end{aligned} \tag{4}$$

扩展 $\boldsymbol{g}^h,\boldsymbol{g}^w$ 作为注意力权重，最终输出特征向量 $\boldsymbol{Y}=\left[y_1,y_2,\cdots,y_C\right]$，可以表述如式 (5) 所示：

$$y_c(i,j)=x_c(i,j)\times g_c^h(i)\times g_c^w(j) \tag{5}$$

在 CA 机制的作用下，网络在特征提取过程中全局池化丢失位置空间信息的问题得到改善，得到了具有方向感知和位置敏感的特征图 [12]，有助于后续的特征表达，提升系统的目标检测性能.

1.3　动态特征点优化策略

对于相机位姿估计而言，特征点的质量至关重要. 动态特征点会引入误差，并且影响对静态点的选择. 对于动态和潜在动态的矩形框，若直接根据目标检测结果剔除矩形框内的特征点则会误删其中的静态点，从而降低系统的定位精度和稳定性. 更严重的情况则是当动态物体框占比较大时，可能会导致跟踪丢失. 因此，提出了一种特征点筛选方法，利用深度信息和对极几何约束，结合 RANSAC 和 ORSA 算法分离出矩形框中的静态特征和动态特征，最大化地剔除动态点并保留静态点.

RANSAC 算法在 1981 年由 Fishler 和 Bolles 提出，是计算机视觉中常用的从一组数据中剔除异常值的方法. 每次使用 N 点来计算 RANSAC 拟合模型时，用 P 表示在 k 次迭代过程中，从点集中随机选择的点都是外点的概率. 则有

$$1-P=\left(1-\theta^N\right)^k \tag{6}$$

式中：θ 为内点的概率；k 为迭代次数，对式 (6) 的两边取对数，得到

$$k=\frac{\log(1-P)}{\log(1-\theta^N)} \tag{7}$$

式中：θ 是一个先验值；P 是预先定义的概率；k 是一个自适应值，在开始时被设定为无穷大，在每次更新模型参数估计时使用当前的内点概率 θ 来更新迭代

次数k. 在本文中, 首先选取矩形框内的部分特征点放入一个集合中. 然后进入一个循环, 直到迭代次数不小于k. 在循环中, 任意选取矩形框中的两个特征点, 计算两点之间的方差作为判定指标. 遍历剩余的所有点, 找到方差在阈值内的点. 随后, 可以得到一组特征点. 遍历结束时, 包含最多内点的集合将作为动态物体轮廓内出现的特征点. 最后, 从所有提取的特征点中筛选出这些被选中的动态点, 则所有剩余的点都是静态点. 算法中使用的方差评判指标为

$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N} \tag{8}$$

式中: s^2为方差; N为每次选择估计模型的点数; x_i为所选点的深度; $\bar{x}$为点集的平均深度值.

利用对极几何约束的基础矩阵进行 RANSAC 误匹配剔除时, 阈值常设为固定值, 且数值大小依赖于经验, 难以完成复杂场景和数据噪声严重时的任务. 本研究提出了 ORSA 算法, 这是一种自适应阈值的 RANSAC 算法, 通过 T 参数平衡阈值和内点数量, T 的定义如式 (9) 所示:

$$T\left(\{\varepsilon_j : j = 1,\cdots,n\}, i\right) = (n-4)\binom{n}{i}\binom{i}{4}(\varepsilon_i)^{i-4} \tag{9}$$

$$\boldsymbol{F} = \underset{\boldsymbol{F}}{\operatorname{argmin}}\left(\min_{i=5,\cdots,n}\left(\{\varepsilon_j(\boldsymbol{F})\}, i\right)\right) \tag{10}$$

式中: T 表示如果存在i个内点, 则内点和外点的阈值为ε_i, 即在大小为n的数据采样点中, 得到的第i个最小的误差被设定为阈值; ε_j为基础矩阵对应关系的误差.

2 实验与分析

在公开数据集 TUM[13] 上进行了实验, 与 ORB-SLAM2 和 DynaSLAM 进行了比较, 评估所提方案的性能. 实验平台为 12th Gen Intel(R) Core(TM) i5-1240P CPU, 16.0 G 内存和 Ubuntu16.04 操作系统, 未使用 GPU 加速.

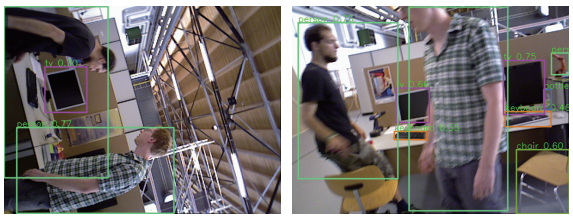
在本次实验中, 选取了 TUM 数据集中的一个动态场景 (fr3/walking) 作为测试数据集. fr3/walking 是一个包含高动态范围的数据集, 相机一共有四种运动状态: 静止 (static); 沿着 x、y、z 轴移动; 在滚动轴、俯仰轴和偏航轴上旋转 (rpy); 沿着直径为 1 的半球移动 (halfsphere). 本文重点研究动态场景, 因此选择了 xyz、rpy 和 halfsphere 三组数据进行性能测试. 分别进行了目标检测、动态特征点剔除和最终的定位实验结果分析.

在准确度方面, 使用绝对轨迹误差 (absolute trajectory error, ATE) 和相对位姿误差 (relative pose error, RPE) 进行评估, ATE 描述估计位姿和真实位姿的直接差值, RPE 描述相隔固定时间差两帧位姿差的精度, 而在速度方面使用平均时间进行度量.

2.1 基于 CA 机制的 YOLOv7 目标检测结果分析

目标检测结果如图 3(a)、(b)、(c) 所示, 分别是 fr3/walking 数据集中 xyz、rpy、halfsphere 序列的结果. 经过 CA 机制改进的 YOLOv7 目标检测网络处理, 得到了带有目标类别及置信度的矩形框, 以及相应的 .txt 格式文件, 记录了矩形框的顶点位置、框的高度和宽度、识别目标的类别以及置信度概率.

(a) fr3/walking_xyz序列检测结果



(b) fr3/walking_rpy序列检测结果

(c) fr3/walking_halfsphere序列检测结果

图 3 改进的目标检测结果

同时, 本文对加入 CA 机制和不加 CA 机制的网络性能表现进行了对比. 根据深度学习的相关理论, 对比准确率、召回率和均值平均精度三个指标, 对模型的性能进行分析. 其中准确率表征在全部预测中正确预测的比例, 召回率表征在全部阳性结果中正确预测的比例. 由于准确率和召回率是互相矛盾的, 因此引入均值平均精度综合两者进行衡量. 结果如表 1 所示, 加入 CA 机制后, 网络的准确度、召回率以及均值平均精度 mAP@0.5 均得到了改善. 同时, 在实验中发现, CA 机制对模型性能的提升还体现在增强了对

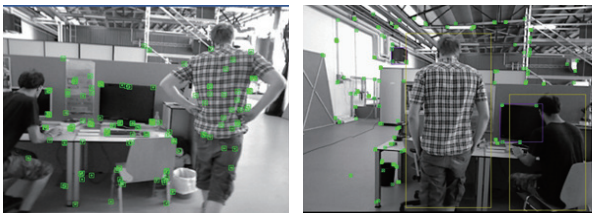
低质量图像的处理能力，主要表现在运动造成的模糊图片识别率得到了改善和提高.

表 1　改进前后网络的性能表现

模型	准确率	召回率	均值平均精度
原YOLOv7模型	0.955	0.926	0.958
加入CA注意力机制的YOLOv7模型	0.981	0.945	0.982

2.2　动态点剔结果

动态点剔除结果如图 4(a)、(b) 所示，图中的绿色点表示从图像中提取到的 ORB 特征点. 图 4(a) 为原 ORB-SLAM2 方案得到的结果，由于保留了人物身上的动态特征点，会造成大量的误匹配，进而影响最终的定位精度和鲁棒性. 而图 4(b) 则为本文所提出的联合目标检测和深度信息的动态特征点剔除优化策略得到的结果，可以看出不仅有效剔除了场景中的动态特征点，还保留了一定的静态特征点，有助于后续相机位姿估计的精度提升.



(a) ORB-SLAM2特征点　　(b) 本研究方案特征点

图 4　特征点剔除前后结果对比

2.3　视觉定位性能分析

针对所提的 CA 机制和动态特征点剔除方法，分别进行了消融实验，最后的定位结果 ATE 和 RPE 性能如表 2 和表 3 所示. 将本文方法用于视觉定位系统，得到最终的定位结果如图 5 和图 6 所示. 图 5 是 ORB-SLAM2 方案得到的轨迹图，其中黑线表示实际测量到的轨迹，蓝线表示预测得到的轨迹图，红线表示预测与实际之间的差距. 由轨迹图可以直观地看出原 ORB-SLAM2 偏离实际较多，定位精度较差. 此外，在实验过程中发现受动态特征点的影响，ORB-SLAM2 会频繁出现跟踪丢失且无法重定位的现象. 图 6 为本研究方案在 fr3/walking_xyz 序列上的轨迹估计，可以看出精度则较 ORB-SLAM2 得到显著提升.

表 2　不同研究方案的 ATE 表现 RMSE

序列	ORB-SLAM2/m	DynaSLAM/m	仅YOLOv7/m	加入CA/m	CA+动态点筛选/m	提升/%
fr3/walking_*xyz*	0.703	0.015	0.025	0.024	0.019	97.3
fr3/waling_rpy	0.590	0.035	0.048	0.046	0.036	94.0
fr3/walking_halfshpere	0.423	0.025	0.038	0.039	0.033	92.2

表 3　不同研究方案的 RPE 表现 RMSE

序列	ORB-SLAM2/m	DynaSLAM/m	仅YOLOv7/m	加入CA/m	CA+动态点筛选/m	提升/%
fr3/walking_*xyz*	0.457	-	0.023	0.022	0.021	95.4
fr3/waling_rpy	0.419	-	0.047	0.045	0.035	91.7
fr3/walking_halfshpere	0.412	-	0.039	0.041	0.032	92.2

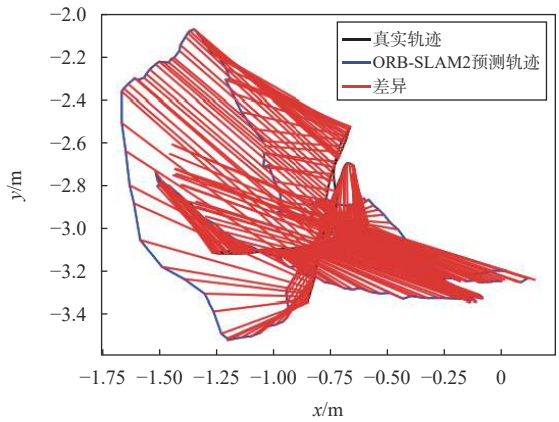


图 5　ORB-SLAM2 在 fr3/walking_*xyz* 序列上的轨迹

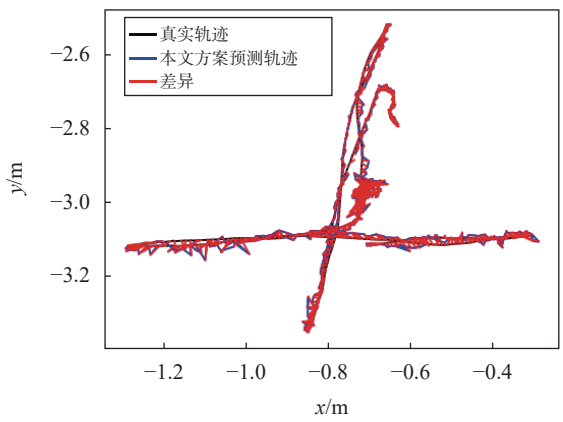


图 6　本方案在 fr3/walking_*xyz* 序列上的轨迹

数据表明目标检测网络对提升系统的定位精度和鲁棒性起主要作用，CA 机制主要提升了在 fr3/walking/rpy 序列上的定位精度和鲁棒性，动态特征点剔除优化方法则能在目标检测的基础上进一步提升定位精度和鲁棒性. 分析原因，一方面，fr3/walking/rpy 和 fr3/walking/halfsphere 序列属于高动态序列，图像成像质量有所下降，因而 CA 机制提升了 YOLOv7 网络目标检测精度，从而提升了后续的定位性能. 另一方面，在序列中的某些时间段，动态物体的占比较大，若直接根据目标检测结果进行动态点剔除，会丢失较多的静态特征点，造成跟踪丢失的现象. 而所提的动态特征点优化策略最大化地保留了静态特征点，因而提升了系统在该场景下的定位精度和鲁棒性. 总之，本文方案在系统的定位精度和稳定性上较 DynaSLAM 稍弱，但与原 ORB-SLAM2 相比，提升显著.

表 4 是 ORB-SLAM2、DynaSLAM 和本文研究方案在实时性方面的比较结果. 可以看出与原 ORB-SLAM2 方案相比，本文方案虽然在实时性上略有不足，但在定位精确度和鲁棒性方面的提升较大. 由于实验未使用 GPU 加速，DynaSLAM 在语义分割步骤会消耗大量的时间和资源，导致整体系统的实时性不能满足最基本的要求. 因此，与 DynaSLAM 相比，所提方案虽在定位精度和鲁棒性方面与其基本持平，但在实时性方面具有明显优势.

表 4　不同研究方案在 fr3 数据集上的用时表现 s/帧

序列	时间	ORB-SLAM2	Dyna-SLAM	本文方案
fr3/walking_xyz	中位数	0.064 6	-	0.298 1
	平均数	0.072 9	0.5	0.302 0
fr3/walking_rpy	中位数	0.069 8	-	0.305 9
	平均数	0.071 3	0.5	0.309 8
fr3/walking_halfsphere	中位数	0.070 5	-	0.310 8
	平均数	0.071 7	0.5	0.316 2

3　结束语

本文针对视觉定位系统在动态场景下，动态特征点造成的定位精度和鲁棒性损失的问题，提出了一种联合目标检测与深度信息相结合的动态特征点剔除方法. 首先，改进 YOLOv7 网络，加入 CA 机制实现快速准确的目标检测. 相比于使用语义分割和实例分割的视觉定位方案，系统在实时性方面得到提升. 然后，提出了一种结合深度信息和对极几何约束的特征点优化策略，利用 RANSAC 和 ORAC 算法有效剔除了动态特征点，尽可能多地保留了静态特征点，提高了系统的准确性和鲁棒性. 在公开的数据集 TUM 上测试了所提方案的性能，结果表明，该方案在定位精度和稳定性方面优于 ORB-SLAM2，实时性方面优于 DynaSLAM.

参考文献

[1] MUR-ARTAL R, MONTIEL J M, TARDÓS J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. IEEE transactions on robotics, 2015, 31(5): 1147-1163. DOI: 10.1109/TRO.2015.2463671

[2] 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. DOI: 10.1109/TRO.2017.2705103

[3] CHEN W, SHANG G, JI A, et al. An overview on visual slam: from tradition to semantic[J]. Remote sensing, 2022, 14(13): 3010. DOI: 10.3390/rs14133010

[4] BESCOS B, FACIL 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. DOI: 10.1109/LRA.2018.2860039

[5] YU C, LIU Z, 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. DOI:10.1109/IROS.2018.8593691.

[6] RAN T, YUAN L, ZHANG J, 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. DOI: 10.1109/JSEN.2021.3099511

[7] 张梦珠, 黄劲松. 基于语义特征的视觉定位研究 [J]. 测绘地理信息, 2022, 47(4): 33-37.

[8] 刘胤真, 徐向荣, 张卉, 等. 动态场景下基于语义和几何约束的视觉 SLAM 算法 [J/OL]. (2023-11-02)[2024-01-15]. 信息与控制. https://doi.org/10.13976/j.cnki.xk.2024.3089

[9] 高逸, 王庆, 杨高朝, 等. 基于几何约束和目标检测的室内动态 SLAM[J]. 全球定位系统, 2022, 47(5): 51-56.

[10] 潘海鹏, 刘培敏, 马淼. 基于语义信息与动态特征点剔除的 SLAM 算法 [J]. 浙江理工大学学报 (自然科学版), 2022, 47(5): 764-773.

[11] 付豪, 徐和根, 张志明, 等. 动态场景下基于语义和光流约束的视觉同步定位与地图构建 [J]. 计算机应用, 2021, 41(11): 3337-3344.

[12] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient

mobile network design[C]//The IEEE/CVF conference on computer vision and pattern recognition, 2021: 13713-13722. DOI:10.1109/CVPR46437.2021.01350

[13] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]// The International Conference on Intelligent Robot Systems (IROS), 2012. DOI:10.1109/IROS.2012.6385773

作者简介

叶睿馨 (1998—)，女，硕士，研究方向为视觉定位、多传感器融合定位. E-mail: 18810595205@163com

张令文 (1983—)，女，博士，副教授，硕士生导师，研究方向为多传感器数据融合（定位），包括GPS/IMU 传感器融合定位、IMU/Visual 视觉融合定位、基于机器学习的多特征指纹定位、通信网络定位. E-mail: zhanglw@bjtu.edu.cn

陈佳 (1983—)，女，博士，教授，博士生导师，研究方向为算力网络、算网融合、在网计算、全息通信、通感算一体化. E-mail: chenjia@bjtu.edu.cn

乔尚兵 (1994—)，男，硕士，研究方向为无线信道采集与信道建模、OTA 性能测试、通感一体化. E-mail: qiaoshangbing@caict.ac.cn

朱颖 (1987—)，女，硕士，高级工程师，研究方向为空天地一体化、NTN、OTA 射频测试. E-mail: zhuying@caict.ac.cn

Dynamic feature point removal method with joint target detection and depth information

YE Ruixin[1], ZHANG Lingwen[1], CHEN Jia[1,2], QIAO Shangbing[3], ZHU Ying[3]

(1. *School of Electronic and Information Engineering*, *Beijing Jiaotong University*, *Beijing* 100044, *China*; 2. *Peng Cheng Laboratory*, *Shenzhen* 518000, *China*; 3. *China Academy of Information and Communications Technology*, *Beijing* 100191, *China*)

Abstract: Aiming at the problem that the localization accuracy and robustness of visual localization systems are easily affected by dynamic feature points in dynamic environments, a dynamic feature point removal method combining target detection and depth information is proposed. The YOLOv7 target detection network is introduced to quickly obtain the target category and position information of the current image frame, and the coordinate attention (CA) mechanism is added to optimize the deep learning model and improve the target detection accuracy of the network. In addition, a dynamic feature point optimization strategy using depth information and pairwise geometric constraints is proposed. Dynamic feature points are effectively eliminated while as many static points as possible are retained, thus reducing the impact of dynamic points on the localization accuracy and robustness of the system. Experimental validation is performed on the publicly available dataset TUM. The results show that the proposed scheme has obvious advantages in terms of localization accuracy and robustness compared with ORB-SLAM2. At the same time, compared with DynaSLAM, the localization accuracy is basically the same, but the operation speed is significantly improved.

Keywords: dynamic feature point elimination; target detection; visual localization; deep learning; dynamic scene