光电技术及应用

DOI：10.16818/j.issn1001－5868.2023110804

动态场景下基于视觉的SLAM技术研究

刘砚菊，晏佳华，冯迎宾*

（沈阳理工大学 自动化与电气工程学院，沈阳 110159）

摘　要： 针对视觉同时定位与地图构建(SLAM)技术在动态环境中存在定位精度低、地图虚影等问题，提出了一种基于深度学习的动态SLAM算法。该算法利用网络参数少且目标识别率高的YOLOv8n改善系统的视觉前端，为视觉前端增加语义信息，提取动态区域特征点。然后采用LK光流法识别动态区域的动态特征点，剔除动态特征点并保留动态区域内的静态特征点，提高特征点利用率。此外，该算法通过增加地图构建线程，剔除YOLOv8n提取的动态物体点云，接收前端提取的语义信息，实现静态语义地图构建，消除由动态物体产生的虚影。实验结果显示，在动态环境下该算法与ORB-SLAM3相比，定位精度提升92.71%，与其他动态视觉SLAM算法相比，也有小幅度改善。

关键词： 深度学习；动态同时定位与地图构建；YOLOv8n；静态语义地图；光流法

中图分类号： TP391.4　**文章编号：** 1001－5868(2024)02－0327－09

SLAM Based on Deep Learning and Optical Flow Constraints in Dynamic Scenes

LIU Yanju，YAN Jiahua，FENG Yingbin

（College of Automation and Electrical Engineering，Shenyang Ligong University，Shenyang 110159，CHN）

Abstract： To address the problems of low localization accuracy and map vignetting in Visual Simultaneous Localization and Mapping（VSLAM）technology in dynamic environments，this paper proposes a dynamic SLAM algorithm based on deep learning. The proposed algorithm utilizes YOLOv8n，which has few network parameters and a high target recognition rate，to improve the visual front end of the system，add semantic information to the visual front end，and extract the dynamic region feature points. The LK optical flow method is then used to identify the dynamic feature points in the dynamic region，eliminate these dynamic feature points，and retain the static feature points in the dynamic region so as to improve the utilization rate of feature points. In addition，the proposed algorithm increases the map construction thread，eliminates the dynamic object point cloud extracted by YOLOv8n，receives the semantic information extracted by the front end，constructs a static semantic map，and eliminates the virtual shadow produced by dynamic objects. Experimental verification indicates that the proposed algorithm improves the localization accuracy in dynamic environments by 92.71% as compared to that of ORB-SLAM3. Further，it achieves a small improvement compared with other dynamic vision SLAM algorithms.

Key words： deep learning；dynamic SLAM；YOLOv8n；static semantic map；optical flow

收稿日期：2023－11－08.

基金项目：辽宁省教育厅基础科研项目(LJKMZ20220614).

* **通信作者：**冯迎宾

0 引言

目前基于视觉同时定位与地图构建(Visual Simultaneous Localization and Mapping,V-SLAM)技术[1]以静态场景假设为前提,但多数实际场景存在动态物体,比如行驶的汽车、移动的行人等。在动态场景下,动态物体会使 V-SLAM 的定位精度严重下降和地图构建产生虚影[2]。

近年来,动态场景下 V-SLAM 算法研究主要集中在两个方向[3]。一是使用传统的多视图几何原理,对动态场景下的物体进行几何约束从而分离动态与静态物体。Kerl 等[4]基于光度误差,建立融合运动像素点函数,优化相机的位置和姿态,但系统鲁棒性低。Alcantarilla 等[5]基于图像运动的稠密场景流构建运动残差提取动态物体,提高了 V-SLAM 在动态场景下的性能,但该系统对场景成像纹理要求较高。另一种利用深度学习方法对图像进行语义信息提取,从而实现对动静态物体的分离。Bescos 等[6]使用基于深度神经网络的 Mask R-CNN[7]目标检测算法对先验动态信息进行分割,再根据多视图几何对提取出的先验区域进行动态特征点检测,将判定为动态特征点剔除。然后利用剩余的静态特征点进行跟踪和重建,但系统实时性差。Su 等[8]提出了基于 YOLOv5s[9]的 RTD-SLAM,该系统利用语义信息和光流提取动态特征点,保证系统的实时性。冯一博等[10]基于 YOLOv3s[11]的动态场景 V-SLAM 算法,剔除检测框中所有的特征点,降低了检测框中特征点的利用率。地图构建部分,DynaSLAM 构建了稠密的点云地图,但不利于机器人的存储与使用。Zhao 等[12]基于深度图在全局范围内构建空间语义地图。

综上所述,动态环境 SLAM 技术存在定位精度低、特征点浪费,以及地图构建存在虚影等问题。针对以上问题,本文首先采用轻量化的 YOLOv8n[13]分割动态场景的动态区域,并筛选候选动态特征点;提出基于 LK 光流法[14]的动态特征点剔除策略,利用动态与静态光流差异检测动态特征点,解决将先验动态语义信息范围内的特征点全部剔除产生的特征点浪费问题。此外,该算法增加地图构建线程,基于 RGB-D 的深度信息构建点云,利用 YOLOv8n 提出的动态区域剔除深度图动态物体,消除动态物体导致的虚影,实现静态地图构建。最后,将前端提取的语义信息传递到地图构建线程建立静态语义地图。

1 动态 V-SLAM 算法框架

本文算法以 ORB-SLAM3[15]算法为基础框架,将 YOLOv8n 目标检测算法应用在前端提取 ORB 特征点部分,通过 YOLOv8n 提取潜在的动态区域。然后,利用 LK 光流法检测动态区域的动态特征点并剔除,以提高 SLAM 系统在动态环境下的定位精度和鲁棒性。本文算法利用保留的静态特征点进行位姿估计与后端优化,避免环境中动态物体对 SLAM 系统的干扰。此外,本文算法增加地图构建线程,基于 RGB-D 的深度信息构建点云,剔除 YOLOv8n 提取的动态物体点云,实现静态点云地图构建。图 1 给出了本文算法的整体框架。

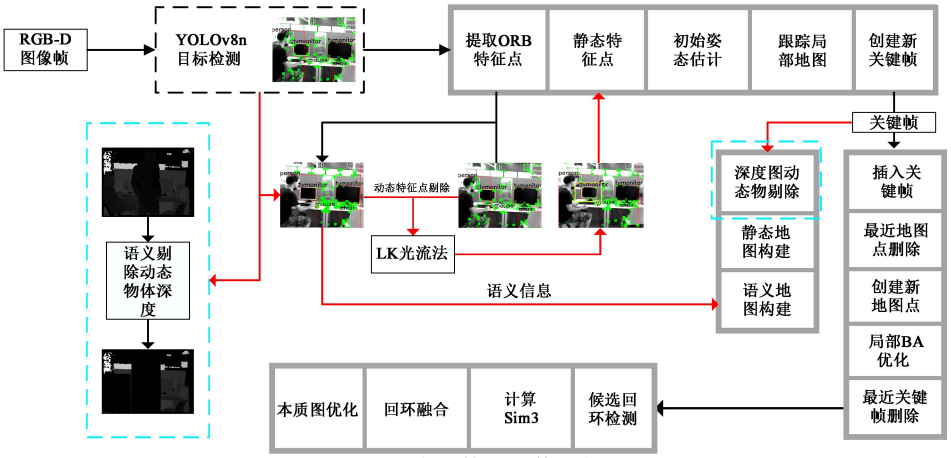


图 1 本文算法整体框架

1.1 目标检测原理

YOLOv8 是目前最新的目标检测算法,为适应不同的任务,YOLOv8 官方发布 YOLOv8n,YOLOv8s,YOLOv8m 等多种模型。由于本文算法

对实时性要求较高，根据表 1 可知，YOLOv8n 实时性最高，参数量最少，故选 YOLOv8n 作为动态目标识别模型。YOLOv8n 网络结构如图 2 所示。

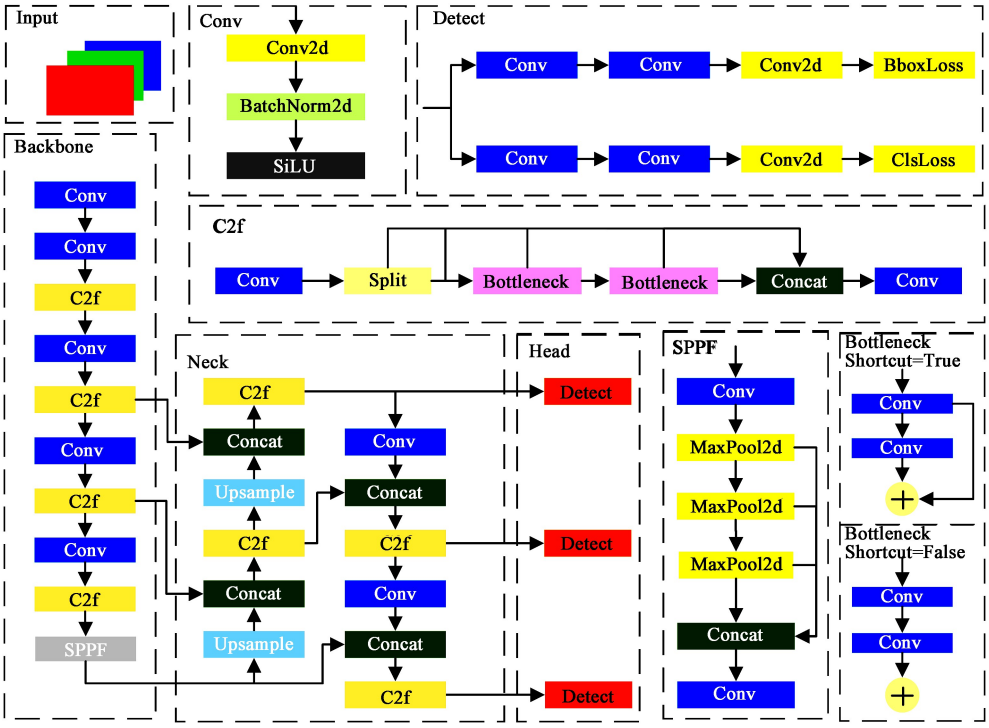


图 2　YOLOv8n 网络结构图

表 1　YOLOv8 系列性能对比

模型	Size/pd	Speed/ms	Params/M
YOLOv8n	640	0.99	3.2
YOLOv8s	640	1.20	11.2
YOLOv8m	640	1.83	25.9

Backbone 部分使用 Darknet53 作为主干网络提取图片特征信息，借鉴了 ELAN[16] 的设计思路，设计了全新的 C2f 模块，替换过去使用的 C3 模块，实现算法的轻量化。Neck 部分使用 PAN-FPN 方式，FPN 采用自顶向下的方式融合不同层次的特征，PAN 采用自底向上的结构保留位置信息，将 Backbone 部分输出的特征图进行多尺度特征融合，获取浅、中、深层的特征信息。Head 部分使用 Decoupled-Head 将分类任务和定位任务分开预测，两条并行的分支分别提取类别特征和位置特征，然后各用一层 1×1 卷积完成分类和定位任务。Head 部分借鉴了 DFL loss 的思想，因此定位任务的通道数 C_{reg} 为 4 * reg_max 的形式，其通道数：

$$C_{reg}=\max(16,256/4,4*\text{reg_max}) \tag{1}$$

式中，reg_max 是一个超参数，默认为 16。即每个特征点会输出 4 * reg_max 个回归预测值，分别对应框的中心点(x,y)和预测框的宽高信息(w,h)。

分类任务输出通道数为 C_{cls}，表示被检测缺陷的类别数，每一个特征点用于判断缺陷的类别信息。其通道数为：

$$C_{cls}=\max(4,\text{cls_mum}) \tag{2}$$

式中，cls_num 为类别数。

YOLOv8 的损失函数由类别损失和边界框回归损失两部分组成，使用 VFL Loss 或 BCE Loss 作为类别损失：

$$VFL(p,q)=\begin{cases}-q(q(\log(p)+(1-q)\log(1-p))) & q>0\\ -\alpha p^{\gamma}\log(1-p) & q=0\end{cases} \tag{3}$$

式中，当正样本 $q>0$ 时，q 为预测框和真实框的交并比，负样本 $q=0$ 时，p 为概率。α，γ 为超参数。

$$BCE(p,y)=-y\times\log(p)-(1-y)\log(1-p) \tag{4}$$

式中，p 为概率，y 为真实值。

选用 DFL Loss＋CIOU Loss 作为边界框损失，有：

$$DFL(S_i,S_{i+1})=-((y_{i+1}-y)\log(S_i)+(y-y_i)\log(S_{i+1})) \tag{5}$$

$$S_i = \frac{y_{i+1} - y}{y_{i+1} - y_i} \tag{6}$$

$$S_{i+1} = \frac{y - y_i}{y_{i+1} - y_i} \tag{7}$$

式中，y 为 target_bbox 网格点上的真实值，y_i 为真实值向下取整，y_{i+1} 为真实值向上取整，S_i，S_{i+1} 为输出分布。

$$I_{\mathrm{CIoU}} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{8}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{9}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{10}$$

式中，IoU 是预测框与真实框的交并比，b 和 b^{gt} 表示两个矩形框的中心点，ρ 表示两个矩形框之间的欧式距离，c 表示两个矩形框闭合区域的对角线距离，v表示两个矩形框相对比例的一致性，α表示权重系数。w，h 为矩形框宽、高。w^{gt}，h^{gt} 为矩形框真实宽、高。

YOLOv8 使用 Task-Aligned Assigner 样本匹配方式，增强了检测学习比对能力。该方式通过排序加权分数选择 TopK 个正样本，加权分数计算公式为：

$$t = s^{\alpha} \times u^{\beta} \tag{11}$$

式中，s 是 GT 的预测分值，u 是预测框 GT Box 的 IoU，α 和 β 为权重超参数。

本文算法采用 COCO 数据集训练 YOLOv8，利用 YOLOv8 提取图像的语义信息生成图像的像素掩膜，如图 3 所示。首先，将人、扫地机器人和汽车等大概率处于移动状态的物体设为先验动态物体，根据先验信息利用 YOLOv8 提取出动态区域。然后，将动态区域传递到追踪线程进行特征点筛选。最后，地图构建线程接收语义信息和像素掩膜进行静态语义地图构建。

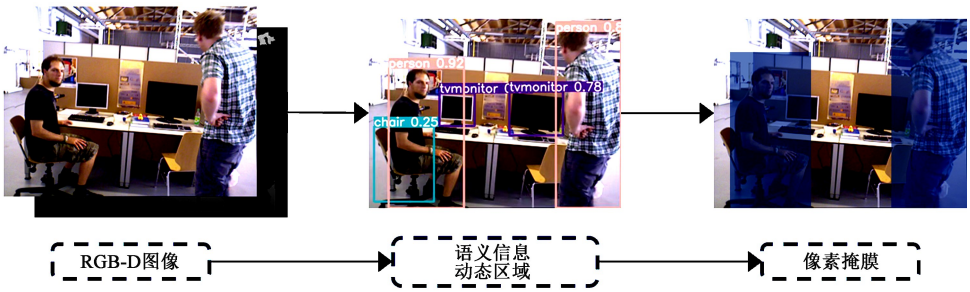


图 3 YOLOv8 提取语义信息形成动态区域像素掩膜

1.2 光流法

光流是描述像素在连续图像之间运动的方法，如图 4 所示。同一个像素在连续图像中不同位置变化称为像素运动。计算图像中部分像素运动称为稀疏光流。计算图像中所有像素运动称为稠密光流。由于本文算法仅需要对动态区域的特征点进行光流计算，因此选用稀疏光流法，同时降低了计算量，保证了系统实时性。

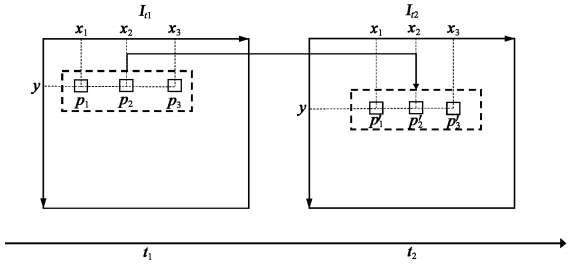


图 4 光流示意图

LK 光流法基于两种假设：(1)像素位置随时间连续变化；(2)灰度不变假设，即空间中的同一点(两张图像表示同一空间点的像素点)，在系列图像中的灰度是不变的。

由 LK 光流法假设可知：

$$I(x, y, t) = I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t) \tag{12}$$

式中，$I(x, y, t)$和 $I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t)$为同一像素在相邻两帧图像的灰度值，$t$ 和 $t + \mathrm{d}t$ 为相邻图像对应时间。

将式(12)右端泰勒展开得：

$$I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t) \approx I(x, y, t) + \frac{\partial I}{\partial x}\mathrm{d}x + \frac{\partial I}{\partial y}\mathrm{d}y + \frac{\partial I}{\partial t}\mathrm{d}t \tag{13}$$

由公式(12)可知，$I(x, y, t)$ 和 $I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t)$相等，等号两边相消再除以 $\mathrm{d}t$，可得：

$$\frac{\partial I}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial I}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial I}{\partial t} = 0 \tag{14}$$

式中，$\mathrm{d}x/\mathrm{d}t$，$\mathrm{d}y/\mathrm{d}t$ 分别表示沿 x 轴，y 轴方向的速度，记为 u，v。$\partial I/\partial x$，$\partial I/\partial y$ 分别表示沿 x 轴，y

轴方向的梯度，记为 I_x，I_y。$\partial I/\partial t$ 为灰度对时间的变化量，记为 I_t。代入式(14)可得：

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -I_t \tag{15}$$

式(15)包含两个未知量 u，v 不能直接解算。为求解式(15)，需添加额外约束。本文假设以特征点为中心的 3×3 正方形窗口的像素运动相同，可得到 9 个方程。

$$\begin{bmatrix} I_x & I_y \end{bmatrix}_k \begin{bmatrix} u \\ v \end{bmatrix} = -I_{tk}, \quad k = 1, \cdots, 3^2 \tag{16}$$

此时方程(15)转变为关于 u，v 的超定线性方程(16)，可通过最小二乘法求解。在 SLAM 系统工作的动态场景中，相机运动使像素产生光流矢量。静态物体光流矢量由相机运动产生，动态物体光流矢量由动态物体自身运动与相机运动共同产生，导致动态场景下静态物体和动态物体的光流矢量不同，如图 5。p_1，p_2 的运动仅由相机运动产生，p_3 的运动由自身运动和相机运动共同组成。故 p_1，p_2 与 p_3 产生的光流不一致。

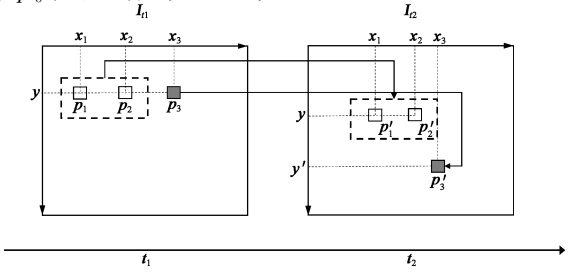


图 5　动态特征点光流示意图

首先依据式(17)计算出图像静态区域特征点的平均光流矢量(U,V)。

$$\begin{bmatrix} U \\ V \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \tag{17}$$

式中，N 为静态特征点个数。

根据静态光流平均矢量与动态特征光流矢量不同，将特征点光流矢量与静态平均光流矢量进行二范数判断是否为静态特征点。

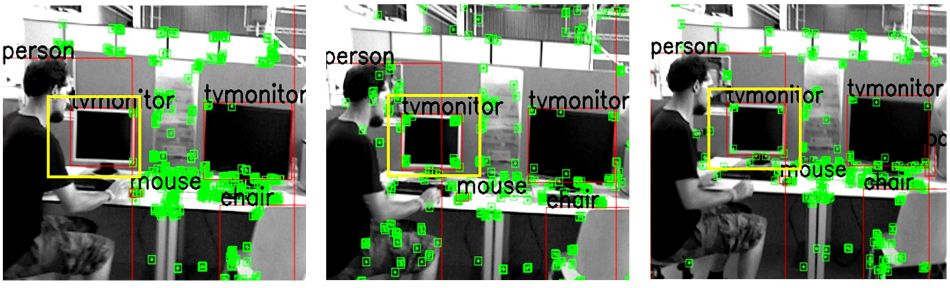
$$\sqrt{(u-U)^2 + (v-V)^2} > l \tag{18}$$

式中，l 是判断动态区域的特征点是否静态特征点的阈值。若小于 l，则判定为静态特征点，反之为动态特征点。

1.3　动态区域特征点提取方法

视觉 SLAM 技术以静态场景假设为前提，在动态场景下，动态物体产生的特征点会对 V-SLAM 的定位精度和建图效果产生严重影响。以往算法利用目标检测算法将动态区域检测出来，并将动态区域的特征点全部剔除。但目标检测算法分割的动态区域中含有部分静态特征点，将动态区域的特征点全部剔除降低了特征点利用率。如图 6(a)、(b)所示，黄色框中静态特征点被剔除，导致视觉 SLAM 后续位姿估计部分可用特征点减少而影响定位精度。

本文算法使用第 1.2 节所述光流法判断策略对动态区域特征点进行识别，保留静态特征点，剔除动态特征点，实现特征点的高利用率。动态区域的特征点剔除效果如图 6(c)所示，能够看到动态检测框中人身上的动态特征点被剔除，而检测框中的静态特征点因光流法检测策略被保留。



(a) 原始特征点　(b) 动态区域全部剔除　(c) 动态区域动态部分剔除

图 6　LK 光流动态特征点剔除

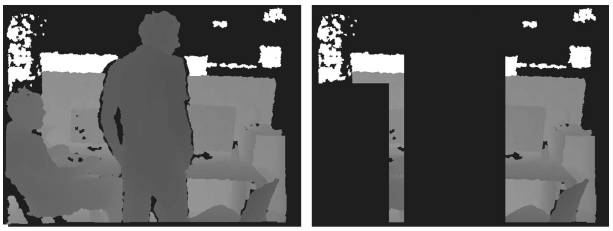
1.4　静态语义地图构建

地图构建线程利用 RGB-D 图片的深度信息和关键帧的位置信息进行地图构建。首先基于 RGB-D 的深度图进行单张图片点云构建，然后利用关键帧的位置信息，将深度图生成的点云从图像坐标系转换到世界坐标系，最后将点云拼接在一起，形成全局点云。点 p 的坐标系转换公式如下：

$$\frac{d}{\boldsymbol{K}} \begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} + \boldsymbol{T} \tag{19}$$

式中，d，$\boldsymbol{K}$ 为深度值和相机的内参矩阵；$\boldsymbol{R}$，$\boldsymbol{T}$ 为旋转矩阵和平移向量；$[u \quad v \quad 1]^{\mathrm{T}}$ 为图片坐标系下像

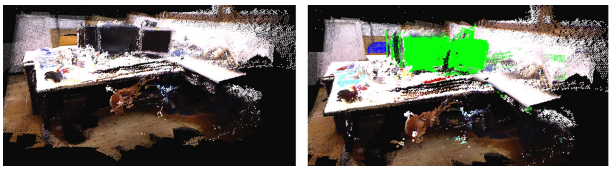
素点的位置；$[x \quad y \quad z]^T$ 为点在空间坐标系下的坐标。RGB-D 的深度图包含动态物体时，直接利用原始信息进行点云构建，会产生重叠的动态物体虚影。因此，地图构建过程中须剔除环境中的动态物体。本文算法利用前端的语义信息生成掩膜将动态物体（图中动态物体为人）深度值设置为 0，如图 7 所示，实现深度图动态物体剔除，从而构建静态点云地图。



（a）原始深度图　　（b）动态物体剔除后深度图

图 7　深度图动态物体剔除

将 YOLOv8n 产生的语义信息传递到地图构建线程并对构建的静态地图进行渲染，形成静态语义地图。图 8(a)为静态地图，图 8(b)为静态语义地图。静态语义地图中使用颜色标签，绿色点云表示显示器，蓝色点云表示椅子。



（a）静态地图　　（b）静态语义地图

图 8　点云地图

2　实验与分析

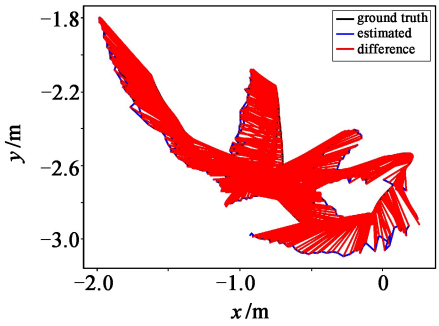
2.1　实验数据集

本文算法选用德国慕尼黑工业大学（TUM）[17] 开源的标准 RGBD 数据集。该数据集是评估 V-SLAM 系统的经典数据集之一，其内容包括深度相机 Kinect 获取的 RGB 图像和深度信息。本文算法使用其中 sitting_xyz，sitting_static，walking_halfsphere 以及 walking_xyz 四个数据序列进行测试，其中 walking_halfsphere，walking_xyz 具有大范围运动物体的高动态场景，而 sitting_xyz，sitting_static 具有小范围运动物体的低动态场景。

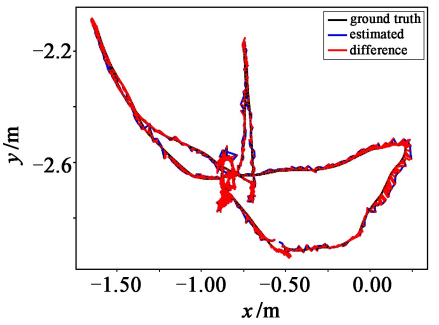
2.2　实验结果及分析

在实验过程中，将本文算法与 ORB-SLAM3 算法进行了实验对比，以量化本文算法在动态环境下对 ORB-SLAM3 算法的改进效果。

图 9，10 为使用 freiburg3_walking_halfsphere，freiburg3_walking_xyz 数据集的真实轨迹对比图。这两个系列为高动态场景下的相机位姿估计轨迹对比，其中黑色线条是真实轨迹，蓝色线条是 SLAM 系统估计的轨迹，红色线条是估计轨迹与真实轨迹之间的误差。由轨迹可知，在高动态场景下因动态目标干扰，ORB-SLAM3轨迹与真实轨

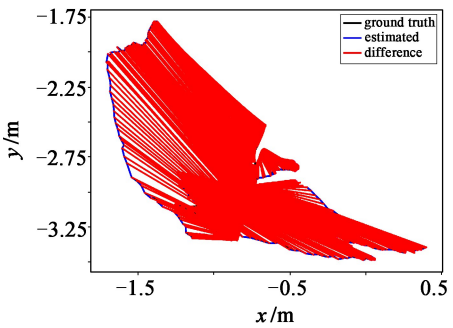


（a）ORB-SLAM3

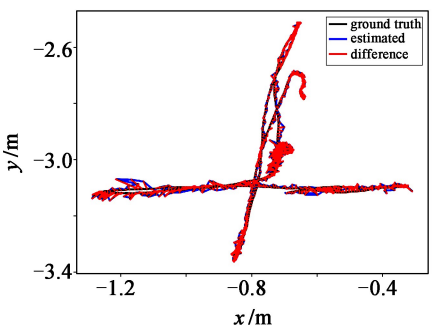


（b）本文算法

图 9　Freiburg3_walking_halfspher 序列轨迹对比



（a）ORB-SLAM3

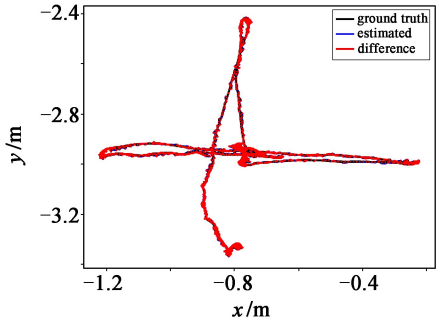


（b）本文算法

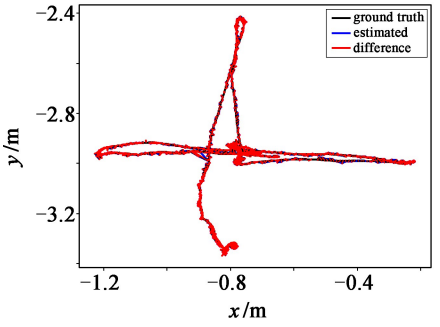
图 10　Freiburg3_walking_xyz 序列轨迹对比

迹相差较大，定位精度差。本文算法采用 YOLOv8n 和光流法将动态物体剔除，利用保留的静态特征点进行相机位姿估计，避免了动态物体干扰，所得轨迹与真实轨迹基本一致，误差较小。

图 11，12 为使用 sitting_xyz，sitting_static 数据集运行得到的真实轨迹对比图。这两个系列为低

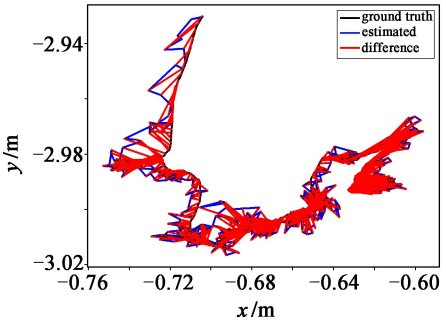


（a）ORB-SLAM3

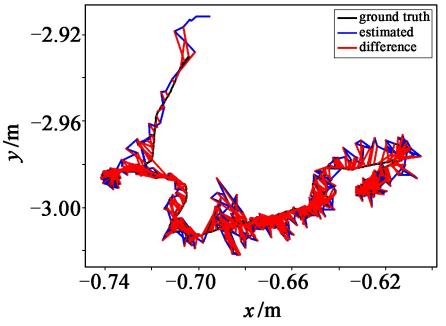


（b）本文算法

图 11　Freiburg3 _sitting_xyz 序列轨迹对比



（a）ORB-SLAM3



（b）本文算法

图 12　Freiburg3 _sitting_static 序列轨迹对比

动态场景下相机位姿估计轨迹对比。由轨迹可知，在低动态场景下，两种算法轨迹与真实轨迹基本一致，ORB-SLAM3 在弱动态场景下定位可靠，本文提出算法与 ORB-SLAM3 相差不大。

定量评估 V-SLAM 需要从定位精度、运算时间、算法复杂度等多方面考虑。其中，定位精度是最重要的评估指标，本文采用绝对轨迹误差（Absolute Trajectory Error，ATE）评估定位精度。

ATE 用来评估两条运行轨迹之间的绝对姿态差，采用平均误差（Mean Error，ME）、中位误差（Median Error）、均方根误差（Root Mean Square Error，RMSE）和标准差（Standard Deviation，SD）等指标进行测量。

RMSE 是测量误差的标准差，对测量精度有很好的反映，其值越大估计轨迹越偏离真实轨迹，用 E_{RMSE} 表示：

$$E_{\mathrm{RMSE}}=\sqrt{\frac{\left(\sum_{i=1}^{n}x_{\mathrm{obs},i}-X_{\mathrm{model},i}\right)^{2}}{n}} \tag{20}$$

式中，i 为第 i 个相机姿态，n 为测量次数，$X_{\mathrm{obs},i}$ 为第 i 个相机姿态的真实值，$X_{\mathrm{model},i}$ 为第 i 个相机姿态的测量值。

ME 是估计姿态与地面真值之间误差的算术平均值，用 E_{mean} 表示：

$$E_{\mathrm{mean}}=\frac{1}{n}\sum_{i=1}^{n}\left|X_{\mathrm{obs},i}-X_{\mathrm{model},i}\right| \tag{21}$$

在相同实验条件下，本文算法相对 ORB-SLAM3 的提升率，采用指标 K 进行评估：

$$K=\left(1-\frac{\alpha}{\beta}\right)\times 100\% \tag{22}$$

式中，α 是本文算法获得的绝对轨迹误差，β 是 ORB-SLAM3 获得的绝对轨迹误差。

由表 2 数据可知，在高动态场景下，由于 ORB-SLAM3 算法利用全部特征点进行跟踪，定位精度仅为分米级，比弱动态区域降低两个数量级。本文算法剔除动态特征点，仅用静态特征点进行跟踪，定位精度为厘米级，比 ORB-SLAM3 提升一个数量级。分析实验误差对比结果可知，本文算法比 ORB-SLAM3 定位精度提高了 92.71%。

为排除硬件设备等其他因素干扰，使用提升率 K 评价本文算法和其他动态 SLAM 系统对 ORB-SLAM 算法的提升幅度。由表 3 数据可知，与 DynaSLAM，RDS-SLAM[18] 相比，本文算法平均定

位精度分别提升 1.120%和 4.100%。

表 2 ORB-SLAM3 与本文算法的 ATE 对比

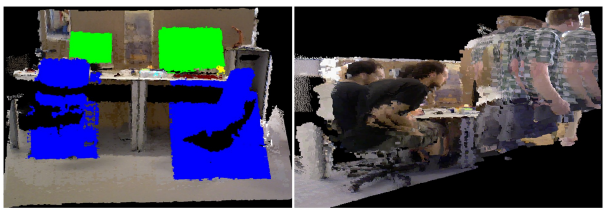
数据集	ORB-SLAM3				本文算法			
	RMSE/m	STD/m	Mean/m	Median/m	RMSE/m	STD/m	Mean/m	Median/m
Sitting_xyz	0.009 2	0.005 3	0.007 9	0.007 3	0.008 8	0.004 3	0.007 7	0.007 0
Sitting_staic	0.007 1	0.003 5	0.006 3	0.005 7	0.006 6	0.003 1	0.005 8	0.005 4
Walking_half	0.218 5	0.087 9	0.199 9	0.184 6	0.022 0	0.012 9	0.021 4	0.017 9
Walking_xyz	0.283 1	0.126 3	0.253 4	0.227 8	0.015 8	0.008 1	0.013 5	0.011 9

表 3 本文算法与其他动态 SLAM 算法的 ATE 对比

Seq	DynaSLAM	RDS-SLAM	本文算法
Walking_half	88.79%	84.46%	89.93%
Walking_xyz	93.31%	91.68%	94.41%

2.3 点云结果

直接利用 ORB-SLAM3 建立的点云地图与本文算法建立的静态点云地图对比实验结果如图 13 所示。由图 13 可知直接利用 ORB-SLAM3 建立的点云地图存在虚影，地图构建效果差。利用本文算法建立的静态点云地图剔除动态物体影响，所建地图不含虚影，更加接近真实场景，并且包含语义信息。本文算法建立的静态点云地图，不包含动态点，对静态场景有较好重建效果。在语义静态点云地图中，语义信息采用颜色标签，绿色点云代表显示器，蓝色点云代表椅子。



(a) 本文算法 (b) ORB-SLAM3

图 13 静态语义点云地图

3 结论

针对动态物体对 SLAM 系统的影响，本文基于 ORB-SLAM3，YOLOv8n 和 LK 光流法提出一种动态 SLAM 算法。首先，本文算法利用 YOLOv8n 提取动态区域并生成像素掩码。然后，利用动态与静态光流差异筛选动态区域特征点，使用保留的静态特征点进行位姿估计与后端优化。最后，本文算法增加地图构建线程，接收前端提取的语义信息和像素掩码，剔除深度图动态物体，消除动态物体虚影，渲染静态语义地图。在公共数据集 TUM 上进行实验验证，结果表明，本文算法与 ORB-SLAM3 相比，绝对轨迹误差提升了 92.71%，比其他动态 SLAM 系统，也有小幅度改善。地图构建线程消除动态物体虚影，并且建立静态语义地图。

未来工作，一是如何更有效检测动态特征点，可使用多种数据集训练检测目标网络，提高检测算法的泛化性、精确度和召回率。二是如何更精确地渲染地图语义信息，可采用实例分割来提取语义信息，更精确分割出动态区域，使语义地图更加精确。

参考文献：

[1] 高兴波，史旭华，葛群峰，等. 面向动态物体场景的视觉 SLAM 综述[J]. 机器人，2021，43(6)：733-750.
Gao Xingbo，Shi Xuhua，Ge Qunfeng，et al. A survey of visual SLAM for scenes with dynamic objects[J]. Robot，2021，43(6)：733-750.

[2] 田 野，陈宏巍，王法胜，等. 室内移动机器人的 SLAM 算法综述[J]. 计算机科学，2021，48(9)：223-234.
Tian Ye，Chen Hongwei，Wang Fasheng，et al. Overview of SLAM algorithms for mobile robots[J]. Computer Science，2021，48(9)：223-234.

[3] 李小倩，何 伟，朱世强，等. 基于环境语义信息的同步定位与地图构建方法综述[J]. 工程科学学报，2021，43(6)：754-767.
Li Xiaoqian，He Wei，Zhu Shiqiang，et al. Survey of simultaneous localization and mapping based on environmental semantic information[J]. Chinese J. of Engineering，2021，43(6)：754-767.

[4] Kerl C，Sturm J，Cremers D. Robust odometry estimation for RGB-D cameras[C]// 2013 IEEE Inter. Conf. on Robotics and Automation，2013：3748-3754.

[5] Alcantarilla P F，Yebes J J，Almazán J，et al. On combining visual SLAM and dense scene flow to increase the robustness of localization and mapping in dynamic environments[C]// 2012 IEEE Inter. Conf. on Robotics and Automation，2012：1290-1297.

[6] Bescós B，Fácil J M，Civera J，et al. DynaSLAM：tracking，mapping，and inpainting in dynamic environment[J]. IEEE Rob. Autom. Lett.，2018，3(4)：4076-4083.

[7] He K，Gkioxari G，Dollár P，et al. Mask r-cnn[C]// Proc. of the IEEE Inter. Conf. on Computer Vision，2017：2961-2969.

[8] Su P，Luo S，Huang X. Real-time dynamic SLAM algorithm based on deep learning[J]. IEEE Access，2022，10：87754-87766.

[9] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]// Proc. of the IEEE/CVF Inter. Conf. on Computer Vision, 2021: 2778-2788.

[10] 冯一博，张小俊，王金刚. 适用于室内动态场景的视觉 SLAM 算法研究[J]. 燕山大学学报，2022，46(4)：319-326.
Feng Yibo, Zhang Xiaojun, Wang Jingang. Research on visual SLAM algorithm suitable for indoor dynamic scenes[J]. J. of Yanshan University, 2022, 46(4): 319-326.

[11] Chun L Z, Dian L, Zhi J Y, et al. YOLOv3: face detection in complex environments[J]. Int. J. Comput. Intell. Syst., 2020, 13(1): 1153-1160.

[12] Zhao Z, Chen X. Building 3D semantic maps for mobile robots using RGB-D camera[J]. Intelligent Service Robotics, 2016, 9: 297-309.

[13] Zhou Y, Zhu W, He Y, et al. YOLOv8-based spatial target part recognition[C]// 2023 IEEE 3rd Inter. Conf. on Information Technol., Big Data and Artificial Intelligence (ICIBA), 2023, 3: 1684-1687.

[14] 徐　陈，周怡君，罗　晨. 动态场景下基于光流和实例分割的视觉 SLAM 方法[J]. 光学学报，2022，42(14)：147-159.
Xu Chen, Zhou Yijun, Luo Chen. Visual SLAM method based on optical flow and instance segmentation for dynamic scenes[J]. Acta Optica Sinica, 2022, 42(14): 147-159.

[15] 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 Trans. Rob., 2021, 37(6): 1874-1890.

[16] Zhang X, Zeng H, Guo S, et al. Efficient long-range attention network for image super-resolution[C]// European Conf. on Computer Vision. Cham: Springer Nature Switzerland, 2022: 649-667.

[17] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]// 2012 IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems, 2012: 573-580.

[18] Liu Y, Miura J. RDS-SLAM: Real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access, 2021, 9: 23772-23785.

作者简介：

刘砚菊(1965—)，女，辽宁省本溪市人，博士，教授，主要研究方向为模式识别与智能系统；

冯迎宾(1986—)，男，山东省聊城市人，博士，副教授，主要研究方向为机器人智能控制与环境感知。