文章编号:1674-2974(2026)04-0010-09 DOI:10.16339/j.cnki.hdxbzkb.2026262

基于语义拓扑信息的物体级SLAM回环检测算法

徐彪[1,2],曾聪磊[1],张润邦[1],黄圣杰[1],刘硕[1],秦晓辉[1,2†],王若钦[3]
(1. 湖南大学 机械与运载工程学院,湖南 长沙 410082;
2. 湖南大学无锡智能控制研究院,江苏 无锡 214115;
3. 中车唐山机车车辆有限公司,河北 唐山 063011)

摘 要:回环检测是SLAM系统的重要组成部分,可以消除里程计的累积误差.传统基于图像外观进行回环检测的方法难以应对大视角变化情况,而显著的物体路标具有视角不变性,能够较好地抵抗视角变化带来的影响.本文提出了一种基于语义拓扑信息的视觉回环检测方法.该方法利用显著物体路标的视角不变性,借助物体地图的语义拓扑信息来编码场景,从而显著提升大视角变化下的系统鲁棒性.具体而言,该方法维护一个分层语义路标数据库,采用"从粗到精"的检测策略.首先,利用宏观高层物体路标提取拓扑图,通过局部拓扑描述符进行粗略匹配,并将局部与全局拓扑图统一在极坐标系下评估空间分布相似性,有效剔除误匹配;随后,在准确物体匹配的引导下,利用微观低层点路标进行精细配准,优化位姿估计.在TUM和USTC数据集上的实验结果表明,该方法在精确率和召回率上表现优异,平均精度超过80%.特别是在大视差回环场景下,定位精度提升了40%以上.

关键词:机器人技术;物体级SLAM;回环检测;语义分割;语义地图
中图分类号:TP391.4;TP183;TP242 **文献标志码**:A

Loop closure detection method with semantic topology information for object SLAM

XU Biao[1,2],ZENG Conglei[1],ZHANG Runbang[1],HUANG Shengjie[1],
LIU Shuo[1],QIN Xiaohui[1,2†],WANG Ruoqin[3]
(1. College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China;
2. Wuxi Intelligent Control Research Institute of Hunan University, Wuxi 214115, China;
3. CRRC Tangshan Co., Ltd., Tangshan 063011, China)

Abstract: Loop closure detection is a crucial component of SLAM systems, enabling the elimination of accumulated odometry errors. Traditional appearance-based methods face challenges in handling large viewpoint changes. This paper proposes a visual loop closure detection method based on semantic topological information. By leveraging the viewpoint invariance of object landmarks and encoding scenes through an object-oriented semantic topological map, the method significantly enhances system robustness under large viewpoint changes. Specifically,

* **收稿日期**:2024-12-10
基金项目:长三角科技创新共同体联合攻关计划项目(2023CSJGG0801), Yangtze River Delta Science and Technology Innovation Joint Force (2023CSJGG0801)
作者简介:徐彪(1992—),男,江苏无锡人,湖南大学副研究员,博士
† 通信联系人,E-mail:qxh880507@163.com

the method maintains a hierarchical semantic landmark database and adopts a "coarse-to-fine" detection strategy. First, high-level macroscopic object landmarks are utilized to extract topological graphs for coarse matching via local topological descriptors; to effectively eliminate mismatches, local and global topological graphs are unified in a polar coordinate system to evaluate spatial distribution similarity. Subsequently, guided by accurate object matches, fine registration is performed using low-level microscopic point landmarks to optimize pose estimation. Experimental results on the TUM and USTC datasets demonstrate that the proposed method exhibits superior performance in both precision and recall, achieving an average precision of over 80%. Notably, in large-disparity loop closure scenarios, positioning accuracy is improved by more than 40%.

Key words: robotics; object SLAM; loop closure detection; semantic segmentation; semantic mapping

视觉同时定位与建图是机器人实现自主定位与导航的关键技术之一.目前视觉SLAM系统的里程计模块在局部定位上已经有较高的精度,但在长期运行过程中仍旧不可避免地产生累积误差,难以得到全局一致的轨迹.为了矫正轨迹的累积漂移,SLAM系统需要通过回环检测模块感知当前的环境,并识别是否访问过当前环境.

传统的回环检测方法基于图像外观,利用图像的局部特征(SIFT[1]、SURF[2]、ORB[3])或者全局特征[4-7]提取视觉特征描述符编码图像,构建图像数据库,通过从数据库中检索图像计算和当前捕获图像的视觉描述符的相似度进行场景匹配.这些传统方法[8-10]对场景的表示停留在低级几何特征上,主要提取图像中的点和线特征,它们在场景外观差异较小的情况下能达到较高精度[11].BoW[12]是传统基于图像外观最有效的方法之一,许多SLAM算法,如ORB-SLAM[13],VINS-mono[14]等,使用BoW实现回环检测,并表现出出色的性能.但在场景纹理缺失或者图像外观发生显著变化的情况,例如视角、光照剧烈变化时,这类传统方法往往会失效[15].

随着深度学习在计算机视觉领域的广泛应用,从图像中提取语义信息变得十分简单,语义信息能够为场景中的物体赋予属性,使其成为高级语义路标.相比于传统的低级几何特征(点、线、面),高级物体语义特征具有视角不变性,能够提供更高层次的环境表征与理解,为场景识别提供更加丰富与具有鲁棒性的信息.因此,出现了许多优秀的物体级SLAM算法和语义回环检测方法[16-17].物体级SLAM是指构建物体级路标的语义SLAM算法,其核心思想是通过检测、识别和跟踪场景中的物体实例,结合物体的语义信息,构建物体级语义地图,并基于物体级语义地图进行回环检测.

物体级语义路标是场景的判别性标志,许多方法利用外观和方向等属性计算物体路标间相似度,进而判断回环,但由于物体属性的不确定性,往往会出现回环的误检和漏检.针对这个问题,出现了许多结合场景语义拓扑信息的回环检测算法[18-22],这类算法将物体级语义地图提取为拓扑图结构,用节点与边表示场景中的物体路标及拓扑关系,将回环检测表述为两个语义拓扑图的相似度计算问题.Lin等[18]提出了一种基于物体地图和语义图匹配的闭环方法,利用长方体模型构建物体地图并将其表示为包含拓扑信息的语义图,同时提出一种基于编辑距离的图匹配方法实现回环检测.Zhao等[21]提出了一种基于物体级语义地图的图匹配全局定位方法,通过三维OBB建模生成语义拓扑图,利用图核特征矩阵和投票机制实现全局图匹配.这两种方法主要关注语义图的全局拓扑匹配,计算复杂度较高,而本文引入显式的局部语义拓扑描述符与极坐标验证机制,减少了描述符模糊性导致的误匹配,有效降低了匹配计算开销.Qian等[19]提出了一种称为SmSLAM+LCD的方法,将语义回环方法集成到语义SLAM系统中,利用物体级语义信息和低级几何特征信息,实现具有鲁棒性的回环检测和有效的漂移消除.本文与其相似之处在于均融合语义与几何特征,不同之处在于本文采用分层语义路标维护策略,实现从粗到精的匹配,从而在保证效率的同时提高检测和定位精度.此外,Yu等[20]提出了SemanticLoop,将物体建模为TSDF,同时将环境表示为具有语义和拓扑的3D图,通过3D语义图的图匹配进行回环检测.该

方法依赖体素建模，计算复杂度较高.本文在保持拓扑语义表达能力的同时，使用椭球模型进行物体建模和轻量的局部拓扑描述符保证实时性与高效性.Xu等[22]提出混合动态点剔除与椭球建模算法构建语义地图，利用语义地图全局匹配实现回环检测.本文采用类似的椭球模型物体表示，但针对回环检测采用局部语义拓扑描述符，避免了复杂的全局匹配代价，使语义拓扑匹配更高效.基于游走描述符的方法[23-25]使用游走描述符对图中每个节点进行编码.然而，游走描述符仅考虑基于距离的相邻顶点的分布，当具有相同语义的对象较多时，容易导致两个语义图的误匹配.此外，当随机游走的数量很大时，查询图和目标图之间的描述符匹配效率很低.

近期研究进一步拓展了语义回环的表达形式.Wang等[26]提出了一种结合语义关键帧筛选与图熵度量实现复杂动态场景下的鲁棒回环检测，该方法通过计算关键帧间语义与结构信息熵的相似度实现全局匹配.Kim等[27]引入了语义一致性判别网络，通过隐式建模语义关系约束来减少假回环的发生.Arshad等[28]提出SLGD-Loop算法，将局部与全局语义描述符结合，用于长期SLAM回环检测，通过多层语义特征融合提升了长时间运行下的识别鲁棒性.这些工作强调了语义特征间的关联建模，本文则利用局部语义拓扑描述符进行物体路标间关系的显式建模，同时采用极坐标空间下的局部验证策略，直接利用物体路标的空间分布特征来抑制错误匹配.Chen等[29]提出的TextGeo-SLAM系统利用文本语义与几何约束联合进行回环检测，在相似结构场景中减少误检.相比之下，本文方法并非依赖特定的语义文本或几何约束，而是通过通用的椭球物体模型构建语义地图，同时利用多层语义路标的分层匹配策略实现语义与几何特征的融合.

总体来看，近年来的研究普遍采用物体级语义拓扑图作为环境表示形式，通过提取物体间的语义与空间关系进行回环检测.这类方法充分利用了高层语义特征的判别性，在复杂场景中表现出较好的鲁棒性.在继承已有研究思想的基础上，本文提出了一种用于物体级SLAM系统的回环检测方法.首先，设计了一种基于局部空间关系的拓扑描述符，能够实现快速的图匹配，同时图提取的时间复杂度也不会随地图规模增加.其次，针对描述符的模糊性，本文添加了误匹配的验证策略，通过将物体统一置于同一极坐标系下，剔除误匹配，提高节点匹配的准确性.最后设计了融合高级语义和低级几何特征的策略，通过分层的特征，实现鲁棒回环检测.本文贡献如下：

1)提供一种分层的语义路标维护策略，融合高级物体级路标的拓扑信息和低级点路标的几何信息，实现从粗到精的鲁棒回环检测.

2)提供一种高效的局部语义拓扑图提取和匹配方法.该描述符编码物体的局部场景信息实现节点间的快速匹配.

3)提供一种准确的图匹配的异常值去除方法.通过将物体节点统一在同一极坐标系下，评估节点的空间分布相似性，有效剔除节点误匹配.

1 语义拓扑回环系统设计

1.1 系统概览

语义回环系统框架如图1所示，该系统建立在物体SLAM算法基础之上，主要实现以下功能：在SLAM系统构建物体级地图的同时，按照分层策略维护融合高层语义信息和低层几何信息的语义拓扑数据库；给定RGB-D语义图像作为源帧(source frame)输入，提取局部拓扑描述符构成源拓扑图G_{source}，在语义拓扑数据库构成的目标拓扑图G_{target}中进行图匹配，得到候选的物体关联；针对物体关联结果，统一在同一极坐标下验证，评估空间相似性并滤除异常值；最后综合评估相似度判断回环，如果检测到回环，则利用物体级信息和特征点信息进行位姿矫正.

1.2 语义拓扑图提取

1.2.1 目标拓扑图提取

通过目前最先进的物体级SLAM算法VOOM[30]构建物体级地图.在构建物体级地图的过程中，结合物体路标的语义、拓扑、几何信息，按照分层策略维护包含高层语义路标和低层地图点路标的分层路标数据库.

如图2所示，分层路标数据库共分为三层：语义层、物体层、地图点层.语义层涵盖了当前物体地图的所有语义标签，每个标签关联对应标签的物体实例，该层可以表述为物体地图中的标签集合$L_{\text{T}}=\left\{l_j\right\}_j^N$.在输入源图$G_{\text{source}}$时，通过$G_{\text{source}}$物体标签快速查询目标图中所有可能匹配的物体.物体层包含了

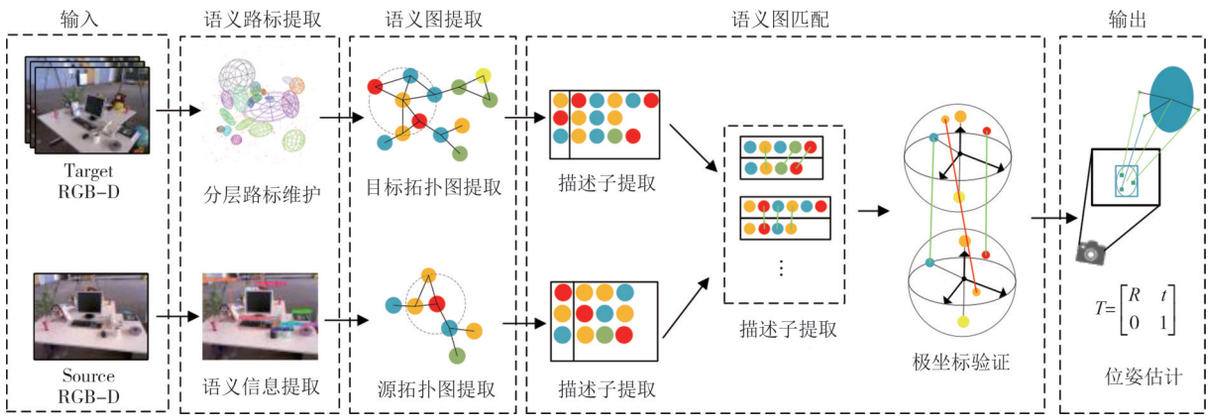


图 1　系统概览

Fig.1　System overview

每个语义标签对应的物体实例，为增强物体的可识别性，利用物体间的局部共视关系，构造物体的局部拓扑描述符，通过局部拓扑描述符再次筛选匹配物体，找到最可能匹配的两个物体，构造极坐标系，将源图 G_{source} 和目标图 G_{target} 统一在同一极坐标系下，评估空间分布相似性，剔除误匹配，得到最终的匹配结果，基于物体匹配结果获取粗矫正位姿.地图点层包含了单个物体实例关联的地图点，在物体关联的结果下，匹配源帧特征点，利用特征点投影约束细化位姿，得到最终的矫正位姿.

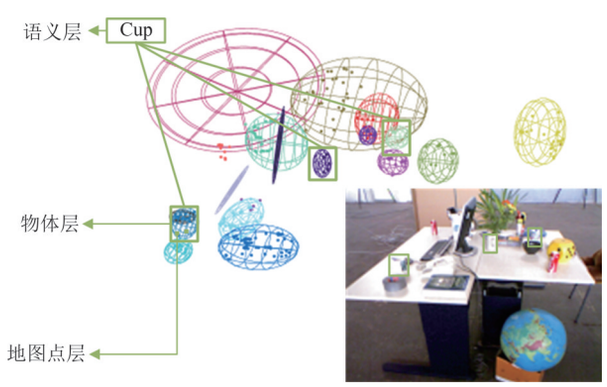


图 2　分层路标维护策略

Fig.2　Maintenance strategy of hierarchical landmarks

1.2.2　源拓扑图提取

源拓扑图 G_{source} 基于语义 RGB-D 图像构建，语义信息通过 YOLOv10[31] 获取，按照置信度阈值过滤后，得到当前帧的物体检测结果，同时将其与特征点关联.然后，将源帧中的二维物体初始化为三维物体路标 $\boldsymbol{n}_i=[l_i,\boldsymbol{m}_i,\boldsymbol{p}_i]$，其中 l_i 为物体的语义标签信息，$\boldsymbol{m}_i$ 为图像语义掩码信息，$\boldsymbol{p}_i$ 为初始化的质心，由对应语义掩码的深度图随机采样获取.

由源帧初始化物体地图后，与目标拓扑图提取方式相同，将每个物体提取为节点，同时按照距离阈值 $d_{\text{threshold}}$ 连接最近的邻居节点以创建边，进而完成源拓扑图 G_{source} 提取.上述过程也可以表述为物体的描述符 $\boldsymbol{d}_i$，描述符形式与 G_{target} 相同，将其添加到物体属性 $\boldsymbol{n}_i=[l_i,\boldsymbol{m}_i,\boldsymbol{p}_i,\boldsymbol{d}_i]$.

1.3　语义拓扑图匹配

1.3.1　语义标签匹配

当给定源图 G_{source} 时，需要在目标图 G_{target} 中寻找匹配物体.首先依据 G_{source} 中的节点标签信息 $L_{\text{S}}=\{l_i\}_i^M$ 在维护的分层路标数据库的标签层 L_{T} 中匹配标签：

$$L_{\text{match}}=\{l_{\text{a}},l_{\text{b}},\cdots|l_{\text{a}},l_{\text{b}},\cdots\in L_{\text{S}}\cap L_{\text{T}}\}\tag{1}$$

在分层路标数据库中，标签层每个标签索引了对应标签的物体集合，因此 G_{target} 中每个节点在 L_{match} 中匹配对应标签，可以获得可能关联的所有物体集合 $O_{\text{match}}=\bigcup_{l_i\in L_{\text{match}}}O(l_i)$. $O(l_i)$ 表示分层路标数据库中由标签 l_i 索引到的物体实例集合.

1.3.2　拓扑描述符匹配

在标签层快速搜索，得到 G_{source} 的节点集合 N_{s} 中节点 $\boldsymbol{n}_i$ 可能匹配的 Object 集合 O_{match}.此时，单个节点与对应标签的物体集合是一对多的匹配 $\text{pair}_i^j=\{\boldsymbol{n}_i,\boldsymbol{o}_j|\boldsymbol{n}_i\in N_{\text{s}},\ \boldsymbol{o}_j\in O_{\text{match}}\}$，需要从中找到真实匹配.这里采用本文设计的局部语义拓扑描述符计算每个节点-物体匹配间的相似度，相似度评分最高的匹配被认为是真实匹配.对于 pair_i^j 中每一对节点-物体匹配，相似度得分计算方法如下：

$$S_{i,j}=\frac{c(\boldsymbol{d}_i^{\text{n}},\boldsymbol{d}_j^{\text{o}})}{\min(|\boldsymbol{d}_i^{\text{n}}|,|\boldsymbol{d}_j^{\text{o}}|)}\tag{2}$$

$c(\boldsymbol{d}_i^{\text{n}},\boldsymbol{d}_j^{\text{o}})$ 是描述符 $\boldsymbol{d}_i^{\text{n}}$ 和 $\boldsymbol{d}_j^{\text{o}}$ 之间相同边的数量，根据边的标签与距离权重判断是否匹配，$|\boldsymbol{d}|$ 表示描述符的边数.相似度评分最高的节点-物体匹配被认

为是真实匹配，最后可以得到节点-物体匹配集合 $P_{\text{match}}=\{\text{pair}_i^j|\boldsymbol{n}_i\in N_s,\boldsymbol{o}_j\in O_{\text{match}}\}$.

1.3.3 极坐标验证

由于描述符只是隐式地编码了节点局部范围内的拓扑关系，利用描述符相似度评估得到节点-物体匹配对 P_{match} 可能存在误匹配.为了进一步剔除异常值，将节点与物体统一置于同一极坐标系下直接评估空间分布相似度.节点极坐标构造过程如图3所示.

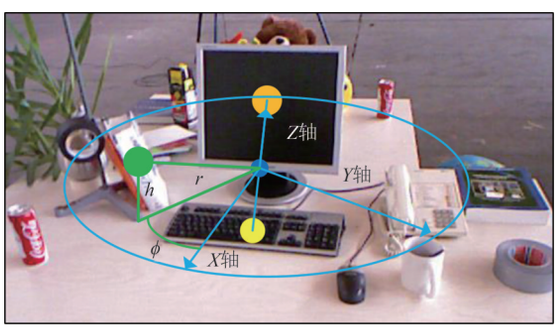


图3 节点极坐标示意图

Fig.3 Diagram of nodes in polar coordinates

在 P_{match} 中选择描述符相似度最高的两对节点-物体匹配对，分别构建源图 G_{source} 和目标图 G_{target} 的 Z 轴方向的向量，同时，分别选择这两对节点-物体匹配对的中点作为源图 G_{source} 和目标图 G_{target} 原点.源图 G_{source} 的原点 $\boldsymbol{O}_s$，Z 轴 $\boldsymbol{Z}_s$ 和目标图 G_{target} 的原点 $\boldsymbol{O}_t$，Z 轴 $\boldsymbol{Z}_t$ 表示如下：

$$\boldsymbol{Z}_s=\frac{\text{pair}\left(\boldsymbol{n}_1(\boldsymbol{p})\right)-\text{pair}\left(\boldsymbol{n}_2(\boldsymbol{p})\right)}{\left|\text{pair}\left(\boldsymbol{n}_1(\boldsymbol{p})\right)-\text{pair}\left(\boldsymbol{n}_2(\boldsymbol{p})\right)\right|},\quad \boldsymbol{Z}_t=\frac{\text{pair}\left(\boldsymbol{o}_1(\boldsymbol{p})\right)-\text{pair}\left(\boldsymbol{o}_2(\boldsymbol{p})\right)}{\left|\text{pair}\left(\boldsymbol{o}_1(\boldsymbol{p})\right)-\text{pair}\left(\boldsymbol{o}_2(\boldsymbol{p})\right)\right|} \tag{3}$$

$$\boldsymbol{O}_s=\frac{\text{pair}\left(\boldsymbol{n}_1(\boldsymbol{p})\right)-\text{pair}\left(\boldsymbol{n}_2(\boldsymbol{p})\right)}{2},\quad \boldsymbol{O}_t=\frac{\text{pair}\left(\boldsymbol{o}_1(\boldsymbol{p})\right)-\text{pair}\left(\boldsymbol{o}_2(\boldsymbol{p})\right)}{2} \tag{4}$$

然后选择过原点且垂直于 Z 轴的平面作为基平面($X-Y$ 平面)，选择该平面下过原点的任意单位向量作为 X 轴，并通过向量积方式获得 Y 轴：

$$\boldsymbol{Y}_s=\boldsymbol{Z}_s\times\boldsymbol{X}_s,\ \boldsymbol{Y}_t=\boldsymbol{Z}_t\times\boldsymbol{X}_t \tag{5}$$

计算其他节点和物体到各自坐标系原点的连线在基平面的投影线与正 X 轴的方位角 ϕ，节点到基平面的距离 h(即 Z 轴坐标)，节点到 Z 轴的距离 r.用 (ϕ,h,r) 描述节点在当前坐标系下的坐标.任意点 $\boldsymbol{a}$ 的坐标计算方法：

$$\phi=\arctan\left(\frac{\boldsymbol{a}\cdot\boldsymbol{Y}}{\boldsymbol{a}\cdot\boldsymbol{X}}\right) \tag{6}$$

$$h=\boldsymbol{a}\cdot\boldsymbol{Z} \tag{7}$$

$$r=\frac{|\boldsymbol{a}\cdot\boldsymbol{Z}|}{|\boldsymbol{Z}|} \tag{8}$$

源图 G_{source} 和目标图 G_{target} 的其余物体坐标和节点坐标都可以通过上述公式获取，逐一比较节点-物体匹配对的坐标 (ϕ,h,r)，对应坐标的差值超过一定阈值 $(\phi_{\text{threshold}},h_{\text{threshold}},r_{\text{threshold}})$，则认为该匹配为异常值.通过极坐标验证后剔除误匹配，得到最终的节点-物体匹配结果 P_{match}.

1.4 位姿估计

图匹配得到剔除异常值的节点-物体匹配对 P_{match} 后，通过最小化3D物体椭球路标和2D检测框拟合椭圆之间的重投影误差来粗略估计位姿.

$$\hat{\mathrm{T}}=\arg\min_{\mathrm{T}}\sum_{(\boldsymbol{n}_i^s,\boldsymbol{o}_j^t)}r(\boldsymbol{n}_i^s,\boldsymbol{o}_j^t)^2 \tag{9}$$

式中：$r(\boldsymbol{n}_i^s,\boldsymbol{o}_j^t)$ 残差由2D椭圆和3D椭球的Wasserstein距离[32]确定，它能够有效避免由物体遮挡、观测不准确导致的位姿估计恶化的情况.

在物体重投影约束给定初值后，基于维护的分层路标数据库，利用地图点层中物体关联的地图点和当前帧匹配节点关联的特征点，用于进一步的位姿细化，获得更加准确的位姿.

2 实验验证

2.1 实验设置

为了验证本文的回环检测算法，在两个公开数据集上进行了广泛的评估：TUM数据集[33]和USTC数据集[18].TUM数据集是目前SLAM算法评估中广泛使用的数据集，它包含真实室内场景的序列，场景中具有丰富的物体信息，可以用于构建物体级路标进行场景识别.USTC数据集是一个为物体级SLAM算法设计的RGB-D数据集，室内RGB-D序列数据由Azure Kinect DK传感器获取，位姿真值由VINS-Mono获得.同时它包含大视差变化下的回环，对传统的回环检测算法具有挑战性.

所有实验均在个人笔记本电脑中进行，配备AMD Ryzen 7 4800H with Radeon Graphics CPU@2.90 GHz、NVIDIA GeForce GTX 1650Ti GPU 和 16 GB RAM，操作系统为Ubuntu 20.04.

2.2　回环检测性能

回环检测被视为两个语义拓扑图的图匹配问题.本文选择经典的图匹配方法RandomWalk[16]作为实验对比方法,为了定量评估不同方法的回环检测性能,本文采用精确率-召回率曲线(precision-recall curve,简称PR曲线)进行性能分析,其中精确率(precision)表示检测出的回环中真实回环的比例,召回率(recall)表示真实回环中被正确检测出的比例.通过改变检测阈值,计算不同精确率和召回率下的平均精度(average precision, AP),以综合评价算法性能.

在TUM序列下测试了不同方法的回环检测性能.其中本文方法和随机游走方法设置的参数如表1所示.对于两种方法,设置了两套参数,其中右边的参数配置代表了更加宽松的回环阈值,宽松的阈值会带来更多误匹配,导致平均精度下降.

表1　方法参数设置

Tab.1　Parameter settings of different methods

本文		RandomWalk	
参数	值	参数	值
$d_{\text{threshold}}$	0.1 m/0.1 m	length	300/200
$h_{\text{threshold}}$	0.2 m/0.3 m	depth	4/4
$r_{\text{threshold}}$	0.2 m/0.3 m	$d_{\text{threshold}}$	0.2 m/0.3 m
$\phi_{\text{threshold}}$	3°/5°		

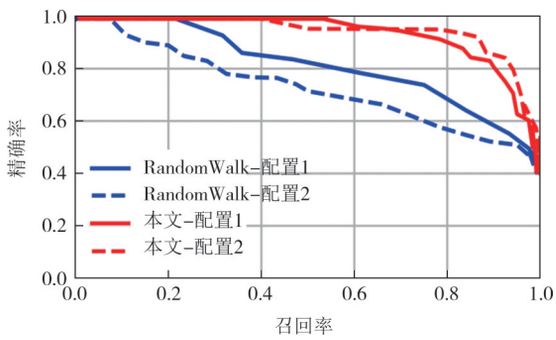
最后,通过调整阈值计算不同参数配置的精确率和召回率,得到不同配置的平均精度,并绘制了PR曲线,结果如表2和图4所示,表2中两套参数配置的最优结果分别用加粗和下划线突出显示.

表2　TUM序列平均精度

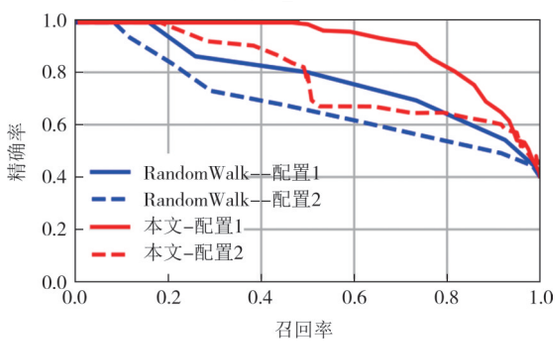
Tab.2　Average precision(AP) of TUM sequence

序列	本文		RandomWalk	
	配置1	配置2	配置1	配置2
fr1_desk	**0.928**	<u>0.905</u>	0.815	0.726
fr1_room	**0.898**	<u>0.744</u>	0.780	0.683
fr2_desk	**0.866**	<u>0.834</u>	0.849	0.679
fr2_person	**0.886**	<u>0.801</u>	0.883	0.787

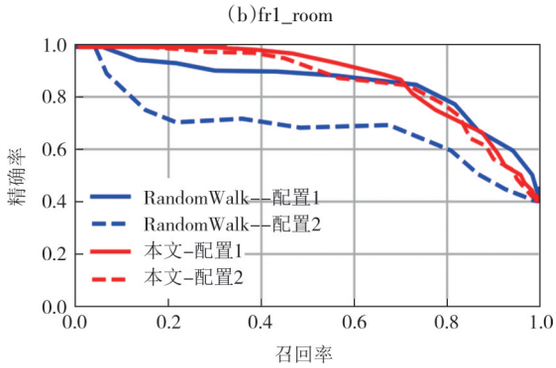
从表2可以看出,相比RandomWalk,本文方法获得了更高的平均精度,配置1的平均精度均在80%以上.这是因为本文方法通过坐标系对齐,直接比较节点坐标来评估相似性,能够更好地利用拓扑关系.图4为精确率-召回率曲线(PR曲线),从图4中也可以看出,本文方法的PR曲线在对应参数配置下基本位于RandomWalk的上方,具有更好的性能.同时,在稍微放宽阈值后,RandomWalk的回环检测性能显著降低,在fr2_desk这个序列中尤其明显,这是因为fr2_desk序列相同标签的物体较多,且位置接近,由于随机游走描述符的模糊性,无法区分这些物体,导致更多的误匹配.而本文方法充分利用了语义和拓扑信息,通过极坐标验证的方法有效抑制了回



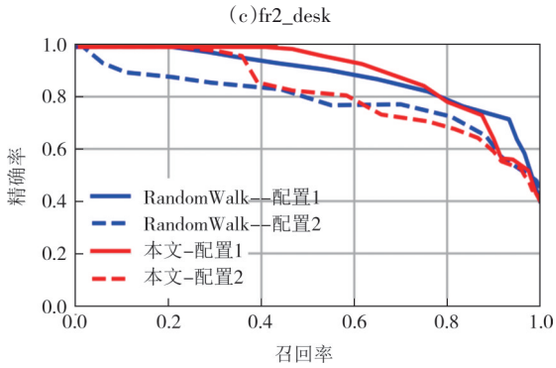
(a)fr1_desk



(b)fr1_room



(c)fr2_desk



(d)fr2_person

图4　TUM序列PR曲线

Fig.4　PR curve of TUM sequence

环的误检率，对参数的变化也更具有鲁棒性，能够在高召回率的情况下保证较高的精确率.在fr2_desk与fr2_person序列中，场景较为简单、物体种类有限，导致语义拓扑区分度较低.在这种情况下，给定严格的阈值条件，本文方法相对于RandomWalk性能提升较小，这也说明了本文方法在语义信息丰富、拓扑关系复杂的场景中有更强的实际应用潜力.

2.3 定位性能

在USTC数据集中与两种公开的算法进行了对比，以评估本文所述方法的定位性能.对比的两个算法分别是:基于传统特征的ORB-SLAM2[13]和基于语义特征的Object-Loop[17].

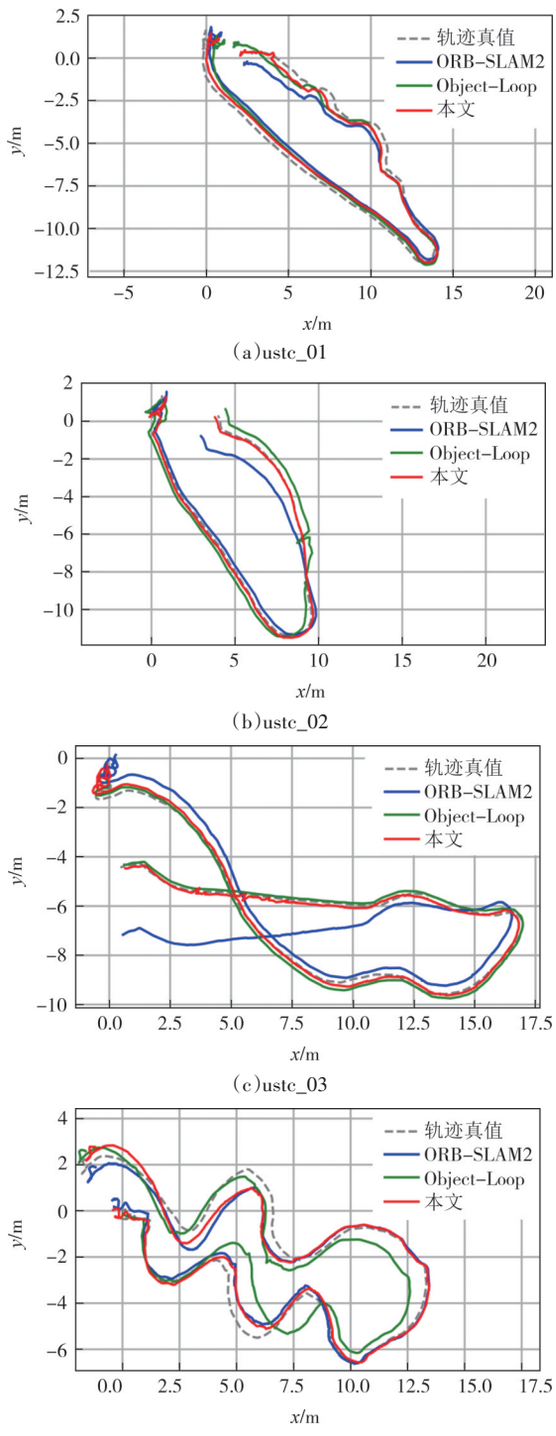
如表3所示，在USTC的四个序列中，本文方法均获得了最小的轨迹误差，与次优方法相比，平均提升超过40%.而基于传统几何特征的ORB-SLAM2在四个序列中表现最差，这是因为传统的低级视觉特征描述符难以应对USTC数据集中视角变化较大的情况，低级特征的配准将会失效.

表3 USTC序列定位误差

Tab.3 Localization errors on the USTC sequence m

序列	RMSE		
	ORB-SLAM2	Object-Loop	本文
01	0.839	0.665	0.361
02	0.609	0.386	0.175
03	0.992	0.306	0.291
05	0.742	0.720	0.272

从图5中可以看出，ORB-SLAM2在ustc_02和ustc_03序列中均没有检测出回环，因此出现了较大的轨迹漂移，这两个序列回环帧的视差均超过100°，ORB-SLAM2使用传统的词袋模型[12]很难检测出回环.而本文方法和Object-Loop利用场景中的语义路标及拓扑信息进行回环检测，因此均能够检测出大视差下的回环，并对轨迹进行矫正.Object-Loop通过将匹配物体对齐来计算位姿，由于物体级SLAM建图获得的物体路标属性不够精确，通过对齐物体计算的位姿精度较低.而本文方法通过匹配的物体路标重投影获取粗略位姿，同时利用物体匹配引导地图点的精细配准，通过精确的地图点几何约束进行更加精确的位姿估计，最后获取了更好的定位性能.



(a)ustc_01 (b)ustc_02 (c)ustc_03 (d)ustc_05

图5 USTC序列轨迹图

Fig.5 Trajectory of USTC sequence

2.4 耗时分析

在TUM序列中进行耗时分析实验，同时使用回环实验中的第一套参数配置测试了本文方法和RandomWalk.表4列出了各个方法的运行时间，其中加粗部分代表了最优结果.

从表4中可以看出，RandomWalk的耗时主要集

中在图提取部分,这是因为游走的随机性,每个节点需要耗费大量时间提取描述子,其时间复杂度随物体数量的增多呈线性增长,在fr1_room序列中物体数量最多,可以看出耗时也显著增加.而本文方法构造简单的描述符并快速匹配,图提取的时间基本控制在毫秒级,同时,较为复杂的图匹配过程也因为充分利用了局部拓扑信息,有效地控制了耗时,本文方法的平均总耗时仅为RandomWalk的22.29%,总体上获得了更高的运行效率.

表4　模块耗时分析

Tab.4　Time cost analysis　　**ms**

序列	耗时			
	方法	图提取	图匹配	总耗时
fr1_desk	本文	1.38	14.32	**15.70**
	RandomWalk	55.77	9.41	65.18
fr1_room	本文	1.87	16.82	**18.69**
	RandomWalk	64.49	8.37	72.86
fr2_desk	本文	0.72	9.18	**9.90**
	RandomWalk	39.16	5.45	44.61
fr2_person	本文	0.78	6.95	**7.73**
	RandomWalk	39.80	5.03	44.83

3　结 论

本文提出了一种用于物体级SLAM的回环检测方法.具体来说,为物体级路标构建局部语义拓扑描述符,通过描述符计算物体间的相似度,得到当前场景的物体匹配,为了解决描述符的模糊性带来的误匹配,将场景统一置于同一坐标系下,考虑匹配物体的空间分布综合评估场景相似性,最后联合物体特征与点特征的约束进行位姿估计.在公共TUM RGB-D数据集和USTC数据集上评估了本文方法,实验验证了算法的有效性和效率.当前的方法仅考虑了物体级的语义拓扑信息,场景中的背景和外观信息被忽略,这在一定程度上造成了信息损失,未来的工作计划融合场景中更丰富的语义和外观信息,实现更加准确的场景识别与回环检测.

参考文献

[1] LOWE D G. Distinctive image features from scale-invariant keypoint. International Journal of Computer Vision, 2004, 60(2): 91-110.

[2] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[3] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF [C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. IEEE, 2012: 2564-2571.

[4] NIE J W, FENG J M, XUE D Y, et al. A training-free, lightweight global image descriptor for long-term visual place recognition toward autonomous vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(2): 1291-1302.

[5] CHEN Y Z, WANG G, ZHOU W, et al. A localization algorithm based on global descriptor and dynamic range search[J]. Remote Sensing, 2023, 15(5): 1190.

[6] ZHANG Q S, XU Z Y, KANG Y H, et al. Distilled representation using patch-based local-to-global similarity strategy for visual place recognition [J]. Knowledge-Based Systems, 2023, 280: 111015.

[7] BONIN-FONT F, BURGUERA BURGUERA A. NetHALOC: a learned global image descriptor for loop closing in underwater visual SLAM[J]. Expert Systems, 2021, 38(2): e12635.

[8] LI Y N, WEI W, ZHU H L. Incremental bag of words with gradient orientation histogram for appearance-based loop closure detection[J]. Applied Sciences, 2023, 13(11): 6481.

[9] BURGUERA A. A loop selection front-end for underwater visual GraphSLAM [C]//OCEANS 2022, Hampton Roads, October 17-20, 2022, Hampton Roads, VA, USA. IEEE, 2022: 1-6.

[10] HAN J Y, DONG R F, KAN J M. A novel loop closure detection method with the combination of points and lines based on information entropy[J]. Journal of Field Robotics, 2021, 38(3): 386-401.

[11] 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.

[12] GALVEZ-LÓPEZ D, TARDOS J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.

[13] MUR-ARTAL R, MONTIEL J M M, TARDÓS J D. ORB-SLAM: a versatile and accurate monocular SLAM system [J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.

[14] 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.

[15] KHALIQ S, ANJUM M L, HUSSAIN W, et al. Why ORB-SLAM is missing commonly occurring loop closures? [J]. Autonomous Robots, 2023, 47(8): 1519-1535.

[16] WU Y M, ZHANG Y Z, ZHU D L, et al. An object SLAM framework for association, mapping, and high-level tasks [J]. IEEE Transactions on Robotics, 2023, 39(4): 2912-2932.

[17] ZHOU B C, MENG Y Q, KAI F. Object-based loop closure with

directional histogram descriptor [C]//2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), August 20-24, 2022, Mexico City, Mexico. IEEE, 2022: 1346-1351.

[18] LIN S Q, WANG J K, XU M, et al. Topology aware object-level semantic mapping towards more robust loop closure [J]. IEEE Robotics and Automation Letters, 2021, 6(4): 7041-7048

[19] QIAN Z T, FU J, XIAO J. Towards accurate loop closure detection in semantic SLAM with 3D semantic covisibility graphs [J]. EEE Robotics and Automation Letters, 2022, 7(2): 2455-2462.

[20] YU J F, SHEN S J. SemanticLoop: loop closure with 3D semantic graph matching [J]. IEEE Robotics and Automation Letters, 2023, 8(2): 568-575.

[21] ZHAO L J, DENG X, LI R F, et al. Graph-based robust localization of object-level map for mobile robotic navigation [J]. IEEE Transactions on Industrial Electronics, 2024, 71 (1): 697-707.

[22] XU Q Y, HUANG S W, SONG Y, et al. Hybrid dynamic point removal and ellipsoid modelling of object-based semantic SLAM [J]. IET Cyber-Systems and Robotics, 2025, 7(1): e70020.

[23] GAWEL A, DEL DON C, SIEGWART R, et al. X-view: graph-based semantic multi-view localization [J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1687-1694.

[24] LIU Y, PETILLOT Y, LANE D, et al. Global localization with object-level semantics and topology [C]//2019 International Conference on Robotics and Automation (ICRA), May 20-24, 2019, Montreal, QC, Canada. IEEE, 2019: 4909-4915.

[25] GUO X Y, HU J J, CHEN J F, et al. Semantic histogram based graph matching for real-time multi-robot global localization in large scale environment [J]. IEEE Robotics and Automation Letters, 2021, 6(4): 8349-8356.

[26] WANG D Z, LUO J W. Approach to semantic visual SLAM for bionic robots based on loop closure detection with combinatorial graph entropy in complex dynamic scenes [J]. Biomimetics, 2025, 10(7): 446.

[27] KIM J, Kim J. Semantic loop closure for reducing false matches in SLAM [C]//2025 22nd International Conference on Ubiquitous Robots (UR), June 30-July 02, 2025, College Station, TX, USA. IEEE, 2025: 34-39.

[28] ARSHAD S, KIM G W. SLGD-loop: a semantic local and global descriptor-based loop closure detection for long-term autonomy [J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(12): 19714-19728.

[29] CHEN S B, LI C Y, JIANG Q, et al. TextGeo-SLAM: a LiDAR SLAM with text semantics and geometric-constraint-based loop closure [J]. IEEE Internet of Things Journal, 2025, 12 (9): 12021-12033.

[30] WANG Y T, JIANG C Y, CHEN X. VOOM: robust visual object odometry and mapping using hierarchical landmarks [C]//2024 IEEE International Conference on Robotics and Automation (ICRA), May 13-17, 2024, Yokohama, Japan. IEEE, 2024: 10298-10304.

[31] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 779-788.

[32] ZINS M, SIMON G, BERGER M O. OA-SLAM: leveraging objects for camera relocalization in visual SLAM [C]//2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), October 17-21, 2022, Singapore, Singapore. IEEE, 2022: 720-728.

[33] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems [C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. October 7-12, 2012, Vilamoura-Algarve, Portugal. IEEE, 2012: 573-580.