

控制与决策

Control and Decision

基于语义位置验证的鲁棒闭环检测算法

赵一夫, 苗津毓, 郭方洪, 吴祥, 董辉, 俞立

引用本文:

赵一夫, 苗津毓, 郭方洪, 吴祥, 董辉, 俞立. 基于语义位置验证的鲁棒闭环检测算法[J]. *控制与决策*, 2023, 38(9): 2511–2519.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.2277>

您可能感兴趣的其他文章

Articles you may be interested in

一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

复杂背景下全景视频运动小目标检测算法

Panoramic video motion small target detection algorithm in complex background

控制与决策. 2021, 36(1): 249–256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

多目标小尺度车辆目标检测方法

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

认知智能电网中基于能效优化的频谱分配策略

Spectrum allocation strategy based on energy efficiency optimization in cognitive smart grid

控制与决策. 2021, 36(8): 1901–1910 <https://doi.org/10.13195/j.kzyjc.2019.1448>

基于姿态估计的实时跌倒检测算法

Real-time fall detection algorithm based on pose estimation

控制与决策. 2020, 35(11): 2761–2766 <https://doi.org/10.13195/j.kzyjc.2019.0382>

基于语义位置验证的鲁棒闭环检测算法

赵一夫¹, 苗津毓², 郭方洪^{1†}, 吴祥¹, 董辉¹, 俞立¹

(1. 浙江工业大学信息工程学院, 杭州 310014;

2. 北京航空航天大学自动化科学与电气工程学院, 北京 100191)

摘要: 针对现有闭环检测算法因视觉干扰而产生的闭环误判问题, 提出一种利用场景语义信息进行验证的闭环检测算法. 该算法通过视觉词袋模型检索闭环候选帧, 用后验方法剔除算法可能产生的错误闭环. 后验方法首先将场景中的语义信息抽象为语义节点; 然后提取包含邻域信息的节点特征, 并以此匹配图像间的语义节点, 降低算法对于动态物体干扰的敏感性. 随后, 算法基于图像间的匹配语义节点构建相对位置网络, 根据网络相似度验证闭环帧, 提高闭环检测算法在具有相似纹理的场景中应对感知混淆的鲁棒性. 实验结果表明, 语义位置验证算法能够显著提升视觉词袋模型的检测准确率. 与其他经典算法相比, 所提出的闭环检测算法具备更快速、更准确的检测性能.

关键词: 闭环检测; 语义信息; 特征描述; 节点匹配; 相对位置网络

中图分类号: TP242

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.2277

引用格式: 赵一夫, 苗津毓, 郭方洪, 等. 基于语义位置验证的鲁棒闭环检测算法[J]. 控制与决策, 2023, 38(9): 2511-2519.

Robust loop closure detection algorithm based on semantic position verification

ZHAO Yi-fu¹, MIAO Jin-yu², GUO Fang-hong^{1†}, WU Xiang¹, DONG Hui¹, YU Li¹

(1. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014, China; 2. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China)

Abstract: To address the loop closure misjudgment problem due to visual interference, this paper proposes a loop closure detection algorithm that uses semantic information to verify candidates. The algorithm retrieves loop closure candidates using a visual bag-of-words model, and applies a verification method to eliminate the mismatches. The proposed verification method first extracts semantic nodes on the basis of semantic information in the scene, and then calculates the node descriptors containing neighboring information, so that we can accurately match the semantic nodes across images, reducing the sensitivity of the algorithm to dynamic objects. Subsequently, the algorithm constructs a relative position network based on matched semantic nodes from two views, and verifies loop closure candidates based on the similarity of network, which improves the robustness of the algorithm against perceptual aliasing. Experimental results show that the semantic position verification method significantly improves the detection accuracy of the visual bag-of-words model. Compared with other classic algorithms, the overall loop closure detection algorithm has achieved leading performances in both detection accuracy and computational efficiency.

Keywords: loop closure detection; semantic information; local feature descriptor; node matching; relative position network

0 引言

以视觉传感器为主要输入的视觉 SLAM (visual SLAM, V-SLAM) 系统因其价格低廉和感知信息丰富等优点被广泛应用于移动机器人等领域^[1]. V-SLAM 系统框架主要包括前端视觉里程计、后端非线性

性优化、闭环检测和地图构建等模块^[2]. 视觉里程计的任务是估计相邻图像间相机的运动, 此过程不可避免地会产生估计误差; 后端非线性优化模块则根据不同时刻相机对同一路标的观测来优化误差. 在长期运行过程中, 前端产生的估计误差会逐渐累积, 导

收稿日期: 2021-12-30; 录用日期: 2022-04-08.

基金项目: 浙江省重点研发计划项目 (2020C01109); 金华市重大(重点)科学技术研究计划项目 (2021-1-012).

责任编辑: 方勇纯.

[†]通讯作者. E-mail: fhguo@zjut.edu.cn.

致V-SLAM算法构建的地图无法全局一致.为此,闭环检测通过判断相机是否回到一个曾经访问过的地点(闭环位置),使得后端非线性优化模块可以获得闭环位置这一长期约束来修正累计误差^[3].因此,闭环检测算法的性能对于大规模场景下V-SLAM系统的精度十分重要,近年来一直是研究的热点.

在V-SLAM算法中,闭环检测通常根据图像间相似度判定闭环的发生,因此,其核心问题是如何紧凑、鲁棒地描述图像^[4].提取图像特征是经典的图像描述方法,通过特征匹配可以较为便捷地得到图像相似度,但是,对所有历史数据进行遍历匹配和搜索会消耗大量时间和内存,严重影响V-SLAM系统的实时性^[5].为了解决这一问题,将图像特征聚类成视觉单词,并使用视觉词向量来表示图像的词典(bag-of-words, BoW)模型^[6]被广泛应用于闭环检测算法中.但是,BoW模型忽略了图像中视觉单词的位置信息,因此,在具有相似纹理的场景下容易受到感知混淆问题的干扰.此外,位于动态物体上的特征会对视觉词向量的表征能力产生干扰,导致BoW模型产生错误的检测结果.除惯用的BoW算法外,文献[7]通过改进FAST算法获得鲁棒描述子,并进一步将闭环检测问题转化为最大值优化问题,获得了不错的闭环检测结果.近年来,深度学习为鲁棒图像表征提供了新的方向,文献[8]改进了Darknet,提取了更为鲁棒的深度学习特征.但上述方法依旧无法很好地应对场景中动态物体的干扰问题.

在V-SLAM算法中增加场景的语义信息,是解决动态物体干扰和感知混淆的常见手段^[9].目前,目标检测和语义分割是获取场景语义信息的两种主要途径.目标检测可以检测出图像中目标物体的类别和位置,获取场景的稀疏信息;语义分割则通过对图像的每一个像素点进行分割,获取场景的稠密信息^[10].语义分割相对于目标检测可以获得更加完整的场景信息,但计算量也更大.为了满足视觉系统中语义分割的实时性要求,MiniNet^[11]和BiSeNet V2^[12]等轻量化语义分割网络取得了显著的效果,这使得语义信息能够被有效地应用于V-SLAM系统中.

为解决动态物体干扰对V-SLAM系统的影响,很多研究者根据语义区域的运动一致性来判定其动静特性,剔除动态区域的图像特征或图像信息,以提高V-SLAM系统的鲁棒性.文献[13]提出的DS-SLAM算法将语义信息与运动一致性检查相结合,剔除了图像特征中属于动态物体的部分.文献[14]提出的SLCD算法使用Fast YOLO模型对道路场景进行目

标提取,通过语义不一致性检测恢复场景中的不变部分,提升了闭环检测算法应对道路场景中动态实例干扰的鲁棒性.在闭环检测算法中引入物体或特征间的位置信息可以高效地解决感知混淆问题^[15],而基于语义信息构建的语义拓扑图是记录位置信息的一种惯用手段.在室外场景的研究中:文献[16]对图像逐帧进行语义分割,将局部子图与视觉信息相结合构建出语义拓扑图,解决了室外大范围场景下的轻量级语义地图构建问题;文献[17]提出了X-View算法,对场景构建语义拓扑图,基于随机游走算法提取图描述子以进行场景识别,实现了多视图全局定位.在室内场景的研究中:文献[18]通过目标聚类构建局部语义拓扑图,有效地提高了室内场景下的闭环检测准确性;文献[19]提出了一种仅依赖语义和几何信息的闭环检测方法,通过结合目标识别和深度信息构建基于欧氏距离的三维拓扑图,利用Kuhn-Munkres算法对应语义节点,进而度量拓扑图间的相似度.

综上,现有基于场景语义拓扑图的闭环检测方法大都取得了不错的闭环检测准确度,但是,拓扑图容易受到动态物体的干扰.检测运动一致性的方式会给拓扑图的构建过程带来较多的额外计算量;并且真实场景的三维拓扑图构建需要图像深度信息或者连续的图像帧,估计出语义区域在环境中的真实位置,对于只有二维图像信息的视觉系统并不友好.针对上述问题,本文结合BoW模型和二维图像中的语义信息,在保证实时性的前提下提出一种能够应对动态物体干扰和感知混淆问题的闭环检测算法.

概括而言,本文的主要创新点包括:

- 1) 提取图像中基于场景语义信息的关键节点以及表征节点邻域内语义信息的局部特征,结合位置权重进行节点匹配,可以减弱视角变化和动态干扰的影响;
- 2) 构建以图像中匹配节点为顶点的位置网络,利用场景中语义信息间的相对位置信息,高效地计算图像相似度,可以有效地应对相似纹理场景中的感知混淆问题;
- 3) 将所提出的语义验证算法与BoW模型相结合,验证闭环候选帧,提升了闭环检测算法的准确性和鲁棒性.

1 基于语义位置验证的闭环检测系统

本文所提出的闭环检测方法框架如图1所示,该框架由候选帧检索模块、语义节点模块和位置网络构建模块构成.在候选帧检索模块中,使用视觉词典模型检索若干闭环候选.在语义节点模块中,算法提

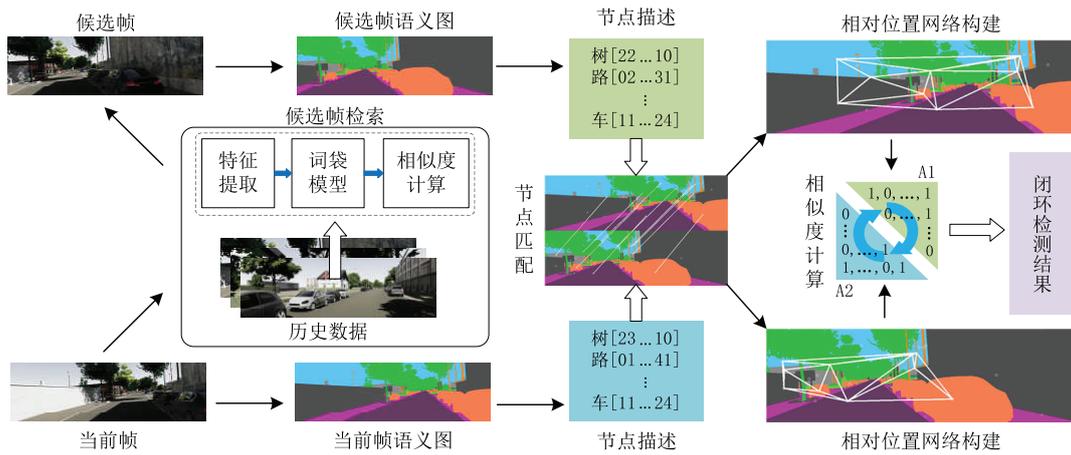


图1 闭环检测系统框架

取图像的稠密语义分割结果,进一步将语义区域抽象为语义节点,并提取每一节点窗口内的语义节点描述子. 在位置网络构建模块中,构建当前帧与候选帧中匹配语义节点的相对位置网络,并通过计算相对位置网络的相似度,判定候选帧是否为当前帧的闭环. 下面将对上述步骤进行详细介绍.

1.1 候选帧获取

为了能够从大量的历史数据中快速检索到与当前帧在视觉上相似的候选帧,本文所提出的闭环检测系统使用BoW模型作为候选帧检索模块的主要框架. BoW模型将图像中提取的若干局部特征点量化为视觉词向量,以此计算图像相似度并完成图像匹配和检索. 此步骤不限制词袋模型中词典的构建方法和局部特征的提取方法. 为了在剔除错误的闭环的同时,不影响闭环检测算法的召回率表现,本文选择与当前帧相似度最高的前 M 个历史图像作为候选帧. 图像相似度由图像词向量间的 L_1 距离得到,有

$$s(w_a, w_b) = 2 \sum_i (|w_{a,i}| + |w_{b,i}| - |w_{a,i} - w_{b,i}|). \quad (1)$$

其中: w_a 和 w_b 表示不同图像的视觉词向量, $w_{a,i}$ 和 $w_{b,i}$ 表示词向量的第 i 维元素.

1.2 语义节点提取及描述

本文所提出的方法将语义区域抽象为语义节点,并计算节点邻域内的语义信息及相对位置信息,作为节点的局部描述子.

1.2.1 节点提取

将图像利用语义分割模型处理后,得到一组带有像素标签的语义区域 $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$.对具有相同语义类别的区域进行边缘检测,获得各个语义区域所对应的最小边界矩阵. 因此,语义区域的信息可以

紧凑地表示为

$$s_i = [(x_1, y_1, x_2, y_2)_i, c_i]. \quad (2)$$

其中: (x_1, y_1) 和 (x_2, y_2) 分别是最小边界矩阵的左上顶点坐标和右下顶点坐标, c_i 为当前区域的语义类别. 为了避免语义分割模型推理而产生的噪声和错误预测,并剔除由于载具移动而变得过远的物体,本文通过阈值对语义区域进行筛选,仅保留图像中足够大的区域. 然后,将保留的语义区域抽象成语义节点,节点在图像中的位置定义为语义区域的中心点,即

$$(o_x, o_y)_i = \left(\frac{1}{2}(x_1 + x_2), \frac{1}{2}(y_1 + y_2) \right). \quad (3)$$

通过上述过程,一张语义图被抽象为语义节点集合 $\mathcal{N} = \{n_1, n_2, \dots, n_m\}$,其中每一个节点 n_i 的信息包含节点位置 $(o_x, o_y)_i$ 和语义类别 c_i .

1.2.2 节点描述

通过语义节点提取步骤,将一张语义图抽象为语义节点集合 \mathcal{N} . 对于其中任意一个节点 n_i ,考虑其他节点相对于该节点的相对位置. 如图2所示,以当前节点 n_i 的横、纵坐标为原点,记录当前节点的上、下、左、右4个方向上其余节点的语义类别,并统计各个语义类别的出现频数,按照上下左右的顺序构造节点的描述子 U . 有

$$U = [Z_u; Z_d; Z_l; Z_r], \quad (4)$$

$$Z_i = [z_1 \ z_2 \ \dots \ z_l]. \quad (5)$$

其中: Z_u 、 Z_d 、 Z_l 、 Z_r 分别代表上下左右4个方向的描述子信息, l 为语义分割模型的语义类别数量, z_i 表示一个方向上某一种语义节点的出现频数. 即矩阵 U 的每一列对应于一种语义类别,每一行对应于一个方向上各类节点的出现频数.

描述子矩阵的计算过程对除当前节点外的每个节点进行两次记录,使得每个节点相对于当前节点

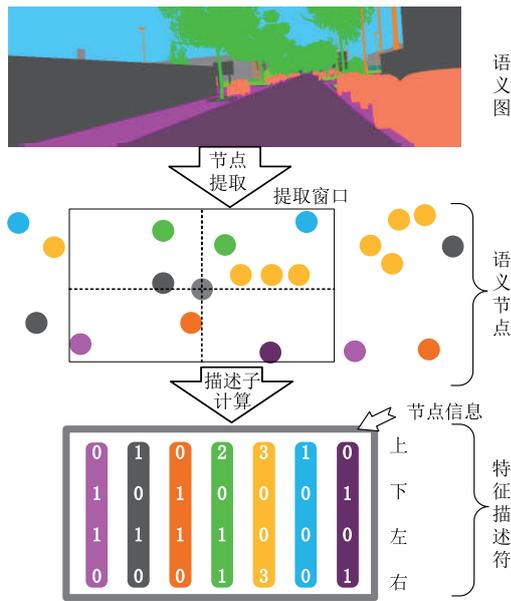


图2 语义节点的特征提取示意图

的上下左右位置关系都得到考虑,完整且紧凑地将语义的位置信息纳入描述子中.若根据当前节点的象限对每个节点进行单次记录,则发生视角变化时,原本节点将被记录在其他象限内,无法鲁棒地描述一个语义节点.对语义图中的每一个语义节点进行描述子提取,可以得到图像的语义节点描述子集合 $\mathcal{D} = \{U_1, U_2, \dots, U_m\}$ 和对应的节点语义类别集合 $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$.

随着相机的移动,图像中可见的场景信息将发生变化,语义节点会随之增多或减少.若每个节点的描述子都记录全局的语义信息,则易导致节点描述子受到干扰.为此,本文所提出算法以当前节点为中心,构建一个矩形窗口,在提取窗口内计算描述子.由于语义节点所对应的分割区域面积存在差异,如果面积较大的语义节点使用较小尺寸的窗口,则会丢失过多的语义特征.因此,应该自适应地提取窗口尺寸,窗口尺寸随着当前节点所对应的语义区域最小边界矩阵尺寸而正比例变化.

相机视角变化会导致语义节点在图像中的绝对位置发生变化,本文所提出的节点描述方法记录了当前节点周围的节点相对位置,能够在一定程度上降低视角变化对描述子产生的影响.若图像中存在动态物体,则其对其他节点的影响在描述子中体现为一个节点的变化,而其他节点对动态物体的影响为多个节点的变化.因此,即使动态区域具有相同语义类别,其节点描述子之间的相似度也会相对较低,易于通过阈值剔除动态物体的干扰,而具有不同语义类别的动态节点可以直接通过语义类别剔除,从而保证节点描述子对于动态物体的鲁棒性.

1.3 语义节点匹配

语义节点需要进一步匹配来获得两张图中的节点对应关系,以便于后续的无向图构建和图相似度计算.对于汽车或移动机器人等行驶于地面的移动主体,其视角变化主要会影响场景物体在二维图像中的左右位置关系,对上下位置关系的影响较小.因此,本文设计了相应的加权相似度计算方法和节点匹配方法,以提供准确而稳定的匹配节点.

1.3.1 相似度计算

对于两个语义节点 n_a 和 n_b ,根据其语义类别 c_a , c_b 和节点描述子 U_a , U_b 定义如下相似度计算方法:

$$S_n(n_a, n_b) = \begin{cases} 0, & c_a \neq c_b; \\ \sum_i \zeta(Z_i^{(a)}, Z_i^{(b)}) \cdot q_i, & c_a = c_b. \end{cases} \quad (6)$$

$$\zeta(Z_i^{(a)}, Z_i^{(b)}) = 1 - \frac{\text{sum}(|Z_i^{(a)} - Z_i^{(b)}|)}{\text{sum}(Z_i^{(a)} + Z_i^{(b)})}. \quad (7)$$

其中: $Z_i^{(a)}$ 为节点 n_a 在 i 方向上的描述子; $\text{sum}(\cdot)$ 表示向量中所有元素的和; q_i 表示所对应方向的相似度权重,包含上下位置权重 q_u , q_d 和左右位置权重 q_l , q_r .如前文所述,在地面载具的应用中,视角变化主要影响节点在图像中的左右位置,为了降低左右位置变化在节点相似度计算中的影响,设置 $p_u = p_d = 0.32$, $p_l = p_r = 0.18$.

1.3.2 节点匹配

得到语义节点 n_a 和 n_b 的相似度后,根据阈值判定这对节点是否为匹配节点:超过阈值判定为匹配,反之不匹配.为获得两幅图像中一对一的匹配节点,需根据节点间的相似度进行排序,优先选择相似度高的节点对作为匹配节点.为挑选出具有代表性且足够可靠的匹配节点进行后续处理,让算法适应不同语义丰富程度的场景,本文使用RANSAC算法^[20]在匹配节点对中选择了固定数量 $T = 10$ 对匹配节点.

1.4 基于相对位置网络的验证方法

传统的视觉词袋库模型丢弃了视觉单词的位置信息,因此在具有相似、重复纹理的场景中易遭遇感知混淆问题.利用本文1.2节中提出的语义节点描述子,可以获得足够准确且鲁棒的匹配节点.为进一步利用语义信息的位置信息,本文提出构建相对位置网络,提取出场景中语义间的相对位置信息来剔除由于感知混淆而产生的错误闭环候选帧.

基于Delaunay三角剖分算法,以前文得到的 T 对匹配节点为顶点,构建出每幅图像中节点间的连接关系(边).这种连接关系(边)所形成的三角形需满足两条基本准则:1)所形成三角形的外接圆内不包含其

他点;2)所形成三角形的最小角最大化. 这两条准则保证了Delaunay三角剖分法构建网络的唯一性.

对经过RANSAC筛选后的匹配语义节点集合 $\mathcal{N}' = \{n_1, n_2, \dots, n_T\}$ 进行三角剖分处理,可以得到一组三角形边 $e_i = (n_a, n_b)$, 其中 n_a 和 n_b 为该边所连接的两个节点. 本文将构建好的三角无向图称为“相对位置网络”, 并利用当前帧与闭环候选帧的相对位置网络相似度来判断候选帧是否正确. 为便于度量相似度, 需要从整体上描述一张图像中物体间的相对位置关系, 即将相对位置网络用一个矩阵或向量表示, 因此, 将相对位置网络转化为邻接矩阵 A . 取任意一个节点作为起始点, 依次将网络每一条边的信息写入邻接矩阵的对应位置, 有

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1T} \\ a_{21} & a_{22} & \dots & a_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ a_{T1} & a_{T2} & \dots & a_{TT} \end{bmatrix}, \quad (8)$$

其中每一个矩阵元素 a_{ij} 表示两个节点 n_i 与 n_j 间的连接关系, 如果存在连接边 $e = (n_i, n_j)$, 则 a_{ij} 为1; 反之则为0. 在当前帧与候选帧上, 分别通过上述方法得到两个邻接矩阵 A_1 和 A_2 .

为方便计算邻接矩阵间的相似度, 两幅图像中相互匹配的语义节点应位于邻接矩阵中所对应的行和列. 令邻接矩阵的对角线元素(即每个节点与自身的相邻情况)为0, 因此, 计算两个邻接矩阵间的差异程度只需比较邻接矩阵左下或右上三角区域内元素的差异, 这里通过下式计算两个临界矩阵间的相似度:

$$S_g(A_1, A_2) = 1 - \frac{\sum_{i=2}^k \sum_{j=1}^{i-1} |a_{ij}^{(1)} - a_{ij}^{(2)}|}{(k^2 - k)/2}. \quad (9)$$

其中: $a_{ij}^{(1)}$ 表示邻接矩阵 A_1 中 a_{ij} 的值, 分母 $(k^2 - k)/2$ 表示 k 阶矩阵左下三角区域的元素数量. 本文对相似度进行了归一化处理, 数值越大, 代表邻接矩阵 A_1 和 A_2 的误差越小, 也代表其所属的两张图像为闭环的可能性越大. 在实验中, 对当前帧与闭环候选帧构建相对位置网络并计算相似度, 位置网络的相似度低于阈值的候选帧将被剔除.

2 实验与结果分析

2.1 实验环境及数据介绍

实验所使用的计算机硬件配置: CPU为i7-7700HQ处理器, 主频2.8GHz, 内存16GB. 由于本文所提出的闭环检测算法需要较高精度的语义分割结果, 为了充分验证所提出模型的有效性, 实验使用语义分割真值进行. 然而, 现有的闭环检测数据集没有提供语义分割的真值, 因此, 本文以公开的街景数据集为基础, 构建一个闭环检测数据集.

实验部分使用Virtual KITTI2数据集^[21], 此数据集包含多条虚拟道路上的不同视觉条件下的图像数据, 并且提供了真值的语义信息. 本文使用Scene 01和Scene 02两条道路场景下的部分数据构建两个实验场景, 如图3所示. 对于每一场景, 本文将3种视觉条件(视角或天气)下的数据集根据位置信息截取等长序列并直接拼接, 构成一条场景序列. 如此构建的数据集包含了天气变化和多视角变化等复杂信息, 常见的动态物体, 以及真值的语义分割结果, 能够全面地评估所提出算法的性能. 得到数据集后, 本文根据图像的位置信息和研究经验, 人工标定了闭环检测算法图像级别的真值闭环, 如图3(b)所示. 所构建数据集的详细信息如表1所示.

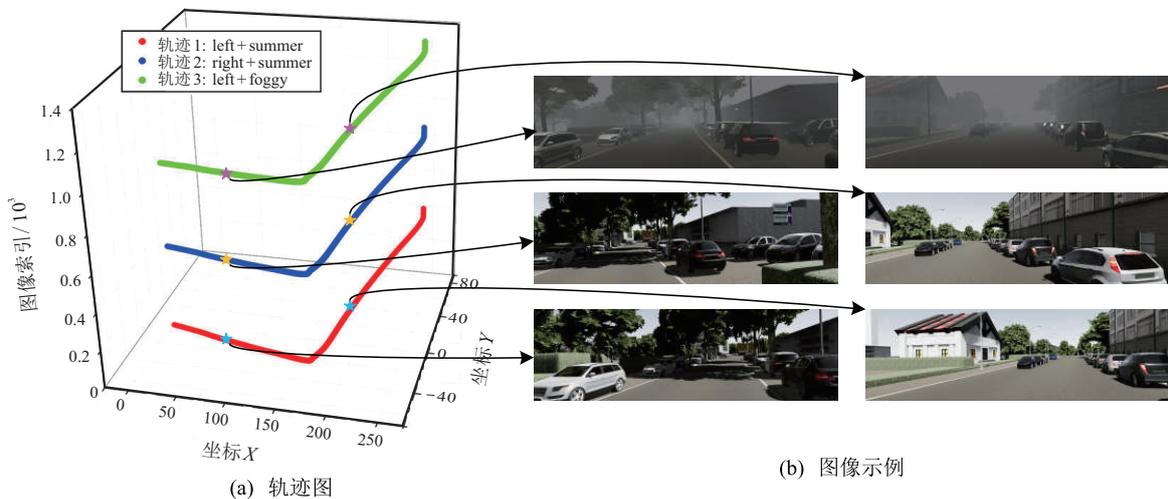


图3 基于Virtual KITTI2构建的Scene 01场景数据轨迹及图像示例

表1 闭环检测数据集介绍

数据集场景	描述	图像数量
scene 01	轨迹1 左视相机,晴天	447
	轨迹2 右视相机,晴天	447
	轨迹3 左视相机,雾天	447
scene 02	轨迹1 左视相机,晴天	233
	轨迹2 右视相机,晴天	233
	轨迹3 左视相机,雾天	233

2.2 评价指标说明

评估闭环检测结果可能会出现表2中的4种情况^[7]. 其中: TP表示算法预测正确的闭环个数, FP表示算法预测错误的闭环个数, FN表示未检测出的真值闭环数量. 闭环检测算法常用的评价指标如下:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}. \quad (10)$$

其中: P 为准确率, R 为召回率. 准确率反映了算法检测正确的能力, 召回率反映了算法检测全面的能力. 改变当前帧与算法预测的闭环结果之间相似度的判定阈值, 可以得到不同的 P 和 R 值, 由此可得到准确率-召回率曲线 (P - R 曲线).

表2 闭环检测结果分类

数据集真值	闭环	非闭环
算法	闭环	true positive (TP)
	非闭环	false positive (FP)
		false negative (FN)
		true negative (TN)

2.3 语义节点匹配实验

本文算法利用语义节点作为局部特征构建拓扑图模型, 完成闭环验证, 因此, 语义节点应对场景视觉干扰的鲁棒性将直接影响算法的性能. 为了直观且量化地评估语义节点描述及节点匹配的准确率, 本文设计了相应的节点匹配实验.

在街景场景中, 动态物体及视角变化对特征描述和匹配会产生剧烈的干扰. 为了验证本文所提出的语义节点匹配方法应对场景中此类干扰的鲁棒性, 根据视觉干扰的类型从 Virtual KITTI2 数据集中挑选出部分测试图像, 包含 40 对含视角变化的图像、40 对有动态干扰的图像和 20 对包含两种干扰的图像, 如图4所示. 人工标记了这些图像对中真实存在的匹配节点数量, 然后通过本文方法对图像对中的语义节点进行匹配, 记录正确的匹配节点数量, 实验结果如图5所示.

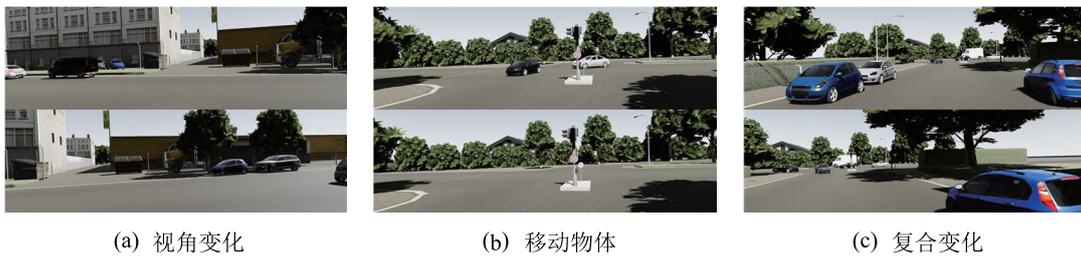


图4 节点匹配实验所用图像示例

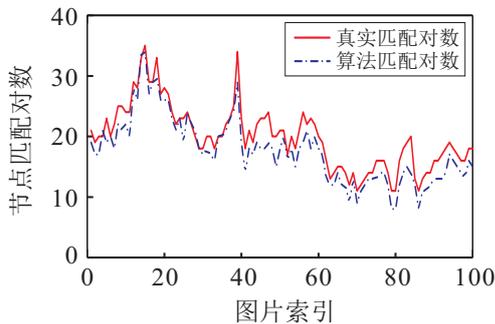


图5 语义节点匹配实验结果

在图5中, 实线表示图像对中的真值匹配节点对数, 点划线表示算法匹配正确的节点对数. 可以看到, 本文所提出的语义节点描述和匹配方法在大部分图像上表现良好, 面对复杂场景中的视觉干扰有足够的鲁棒性. 为了量化这一性能, 这里将算法正确匹配数与真值的比值作为准确率, 计算出算法在3种视觉干扰下的平均匹配准确率, 如表3所示.

表3 语义节点匹配平均准确率

变化类别	视角变化	物体变化	复合变化
匹配平均准确率/%	90.5	88.5	85.7

由实验结果可以看出, 本文所提出的节点描述和匹配算法能够有效应对场景中存在的两种主要干扰. 对于视角变化, 由于算法采用语义节点之间的相对位置和语义分布情况来描述节点, 并且添加窗口模式以避免相机运动对语义节点数量造成的影响, 因此, 语义节点的描述子对视角变化有足够的鲁棒性. 对于动态干扰, 本文所提出的节点描述可以较好地地区分动态语义节点和静态语义节点, 在节点匹配过程中, 通过语义类别和节点描述子相似度两方面约束, 有效剔除了动态物体的干扰. 综上, 本文的语义节点描述和匹配方法对于复杂场景中的主要干扰具有较好的不变性, 能够为后续的位置网络构建过程提供准

确的匹配节点. 在语义节点匹配之后, 对匹配的语义节点对进行基于RANSAC算法的筛选, 以剔除可能存在的误匹配点对, 使准确率进一步提高, 并选择具有代表性的最优匹配节点对进行后续处理.

在图6中, 本文对比了ORB特征匹配与语义节点匹配的效果. 可以看到, 经过RANSAC算法的筛选, ORB特征也可以达到较高的匹配精度, 但是, 匹配的特征集中在局部区域, 很难全面反映整张图像的视觉信息, 用这样的特征构建无向图难以有效地利用位置信息. 与之相比, 本文算法用语义节点进行匹配, 并且为之设计了专门的节点描述和匹配方法, 可以看到, 匹配的语义节点分布更加合理, 准确率也满足要求.



(a) ORB特征匹配结果 (b) 语义节点匹配结果

图6 匹配效果对比

2.4 闭环验证实验

本文利用语义节点构建了位置网络, 对词袋模型检索得到的闭环候选帧进行验证. 为了验证所提出的算法对词袋模型的性能提升效果, 本文使用目前应用最广泛的词袋模型DBoW3^[22] (DBoW2的改进版本) 作为检索框架. 对于系统中需要提取图像特征的部分, 不失一般性, 使用ORB特征算法进行提取. 首先在实验中预训练一个视觉词典, 通过DBoW3框架加载该词典. 在检测闭环候选帧过程中, 提取图像特征后, 利用DBoW3框架将局部特征转换为一个视觉词向量, 用于计算图像间相似度并检索候选帧. 实验中, 用于对比的基本算法只检索出相似度最大的1张候选帧, 记为“DBoW3”. 而本文方法在上述过程中检索出相似度最大的5张候选帧, 对每个候选帧与当前帧进行语义节点提取、匹配和位置网络构建, 计算位置网络的相似度, 将位置网络相似度最大的候选帧作为算法检测到的闭环结果. 因为这一过程的功能为验证DBoW3框架得到的闭环结果, 所以记为“DBoW3+语义位置验证”. 此外, 为验证语义位置信息的优势, 本文设置基于特征的位置验证实验. 实验通过在当前帧与候选帧间分别构建匹配特征点间的相对位置网络(同1.4节中所述方法), 进而计算出候选帧与当前帧间的相似度, 从5张候选帧中选择位置验证相似度最大的一张作为最终的闭环结果. 此实验记作“DBoW3+特征点位置验证”.

如图7和图8所示, 本文分别在scene01和scene02两个场景上进行了闭环检测实验. 可以看出: 本文方法相对于DBoW3算法, 准确率有大幅度提高; 相对于“DBoW3+特征点位置验证”的闭环检测方法, 也有准确率上的优势. 在本文所构建的两个场景中, 包含了街景中常见的动态物体干扰、天气变化和视角变化等干扰, 对基于视觉信息的闭环检测算法产生了较大的影响. 在两个场景中, 经典的DBoW3算法准确率表现较差, 这是由于传统的词袋模型在特征提取和量化为视觉词向量这一过程中丢弃了位置信息, 容易在包含相似、重复纹理的场景中产生感知混淆. 在包含动态物体的场景中, 当图像的特征点位于动态物体上时, 当前帧与闭环帧之间的相似度会降低, 从而影响闭环检测的性能. “DBoW3+特征点位置验证”的方法虽然通过位置验证可以降低相似场景中感知混淆的影响, 但是, 由于特征点提取过程受天气和动态物体的影响较大, 不能稳定地应对多变的复杂真实环境. 而本文提出的算法在DBoW3算法的基础上加入了基于语义信息的验证过程, 构建了鲁棒性较好的语义节点, 并利用匹配准确的节点构建了相对位置网络, 高效地将语义区域的类别信息和位置信息加入算法中, 从而增强了算法在复杂场景中应对各种干扰的鲁棒性, 从准确率-召回率表现上可以看出, 算法的准确率有了显著提升.

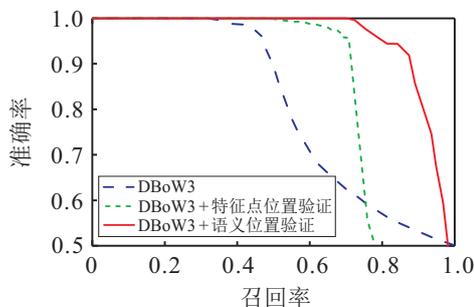


图7 scene01场景的P-R曲线图

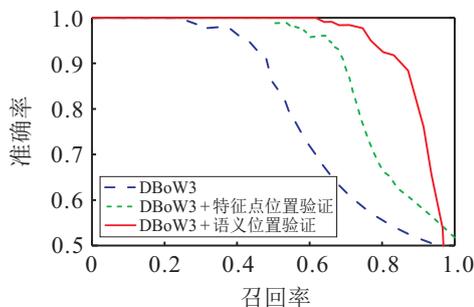


图8 scene02场景的P-R曲线图

2.5 闭环检测算法整体性能

为了更全面地评估本文方法, 实验按2.4节所述将所提出的验证方法与DBoW3算法相结合, 并与基

本的DBoW3算法、FAB-MAP 2.0算法^[23]、SeqSLAM2算法^[24]和文献[15]中的方法进行比较。其中,各算法皆采用其作者提供的代码及默认参数进行实验。为公平比较,本文将文献[15]算法中的深度学习特征替换为ORB特征。由于在V-SLAM系统中错误的闭环将导致地图构建的失败,本文更关注当准确率达到100%时算法可以达到的最大召回率。各算法的最大召回率如表4所示。

表4 100%精度时不同算法的最大召回率 %

实验场景	scene 01	scene 02
DBoW3	33	26
FAB-MAP 2.0	51	47
SeqSLAM2	60	52
文献[15]	50	49
本文算法	71	63

由实验结果可以看出,所提出的算法在两个实验场景中都达到了最优的表现。与其他算法相比,本文在DBoW3算法的ORB局部特征之外巧妙地利用了语义信息,因而获得了更鲁棒的表征性能。实验结果验证了有效利用语义信息能够显著提升闭环检测算法应对动态物体、感知混淆和天气变化等视觉干扰的鲁棒性。

为了分析本文算法的执行效率,将算法分为5个阶段:候选帧检索、语义节点提取、节点描述、节点匹配和位置网络构建及验证。表5中给出了上述5个阶段运行的平均时间,并且对比了算法的运行时间。

表5 算法运行时间 单位: ms

实验场景	scene 01	scene 02
候选帧检索	22.73	13.04
语义节点提取	25.63	22.45
语义节点描述	1.30	0.93
语义节点匹配	4.24	3.65
位置网络验证	1.92	1.87
本文算法时间总和	55.82	41.94
FAB-MAP 2.0	287.2	184.1
SeqSLAM2	164.3	103.7
文献[15]	36.2	30.1

由表5可知,本文算法在语义节点提取步骤上花费时间较多,因为算法需要遍历语义图中所有的语义种类,并提取最小边界。就总体表现而言,本文所提出的算法可以在约50ms内完成一次闭环检测,与经典的FAB-MAP 2.0算法和SeqSLAM2算法相比耗时更少。文献[15]算法由于不需要使用额外的语义信息构建后验过程,在耗时上相对较少;但本文算法在实

时性表现相近的情况下,可以显著提升检测准确率,因此更具优势。

综上,本文所提出的闭环检测算法在复杂场景中获得了较好的检测性能,能够以较高的准确率检测场景中存在的大部分闭环,并且计算耗时较少,对于V-SLAM系统具有较高的实用价值。

3 结论

本文提出了一种基于语义位置验证的鲁棒闭环检测算法,旨在提升闭环检测应对场景中各种视觉干扰的鲁棒性。该方法使用词袋模型作为候选帧检索的方法,得到当前帧所对应的闭环候选帧;然后,分别对当前帧和候选帧进行语义分割,基于语义分割结果进行节点提取、节点描述和匹配,并通过三角剖分法建立了语义节点间的相对位置关系,即相对位置网络;最后,通过位置网络的相似度验证闭环。在复杂场景中的实验结果表明,本文所提出的算法能够更好地应对场景中动态物体干扰、视角变化,并结合位置信息有效解决了感知混淆问题,显著提升了基于词袋模型的闭环检测算法的准确率,并且在运行时间上也具有一定的优势。在未来工作中,将进一步优化从历史数据中检索候选帧的能力,并增强算法在语义丰富度较低的场景中的适应能力,以及应对不良语义分割结果的能力。

参考文献(References)

- [1] 李云天, 穆荣军, 单永志. 无人系统视觉SLAM技术发展现状简析[J]. 控制与决策, 2021, 36(3): 513-522. (Li Y T, Mu R J, Shan Y Z. A survey of visual SLAM in unmanned systems[J]. Control and Decision, 2021, 36(3): 513-522.)
- [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.
- [3] Ho K L, Newman P. Loop closure detection in SLAM by combining visual and spatial appearance[J]. Robotics and Autonomous Systems, 2006, 54(9): 740-749.
- [4] 佟国峰, 张佳伟, 刘满堂, 等. 基于高效回环检测和重定位的SLAM算法[J]. 控制与决策, 2020, 35(3): 587-592. (Tong G F, Zhang J W, Liu M T, et al. SLAM algorithm based on efficient loop detection and relocalization[J]. Control and Decision, 2020, 35(3): 587-592.)
- [5] 张云洲, 胡航, 秦操, 等. 基于栈式卷积自编码的视觉SLAM闭环检测[J]. 控制与决策, 2019, 34(5): 981-988. (Zhang Y Z, Hu H, Qin C, et al. Loop closure detection for visual SLAM based on stacked convolutional autoencoder[J]. Control and Decision, 2019, 34(5): 981-988.)

- [6] 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.
- [7] 张翠军, 张玉河. 基于HHO算法的SLAM闭环检测方法研究[J]. *激光与光电子学进展*, 2021, 58(12): 452-460.
(Zhang C J, Zhang Y H. Research on SLAM loop closure detection method based on HHO algorithm[J]. *Laser & Optoelectronics Progress*, 2021, 58(12): 452-460.)
- [8] 郭纪志, 刘凤连, 杨馨竹, 等. 基于深度学习的视觉SLAM闭环检测方法[J]. *光电子·激光*, 2021, 32(6): 628-636.
(Guo J Z, Liu F L, Yang X Z, et al. The closed loop detection method of vision SLAM based on deep learning[J]. *Journal of Optoelectronics · Laser*, 2021, 32(6): 628-636.)
- [9] Tian Y L, Chang Y, Herrera Arias F, et al. Kimera-multi: Robust, distributed, dense metric-semantic SLAM for multi-robot systems[J]. *IEEE Transactions on Robotics*, 2022, 38(4): 2022-2038.
- [10] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [11] Alonso I, Riazuelo L, Murillo A C. MiniNet: An efficient semantic segmentation ConvNet for real-time robotic applications[J]. *IEEE Transactions on Robotics*, 2020, 36(4): 1340-1347.
- [12] Yu C Q, Gao C X, Wang J B, et al. BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation[J]. *International Journal of Computer Vision*, 2021, 129(11): 3051-3068.
- [13] Yu C, Liu Z X, Liu X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, 2018: 1168-1174.
- [14] Chen H S, Zhang G, Ye Y D. Semantic loop closure detection with instance-level inconsistency removal in dynamic industrial scenes[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(3): 2030-2040.
- [15] Yue H S, Miao J Y, Yu Y, et al. Robust loop closure detection based on bag of SuperPoints and graph verification[C]. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macao, 2019: 3787-3793.
- [16] Bernuy F, Ruiz del Solar J. Semantic mapping of large-scale outdoor scenes for autonomous off-road driving[C]. 2015 IEEE International Conference on Computer Vision Workshop. Santiago, 2015: 124-130.
- [17] Gawel A, Don C D, Siegwart R, et al. X-view: Graph-based semantic multi-view localization[J]. *IEEE Robotics and Automation Letters*, 2018, 3(3): 1687-1694.
- [18] 张括嘉, 张云洲, 吕光浩, 等. 基于局部语义拓扑图的视觉SLAM闭环检测[J]. *机器人*, 2019, 41(5): 649-659.
(Zhang K J, Zhang Y Z, Lü G H, et al. Loop closure detection based on local semantic topology for visual SLAM system[J]. *Robot*, 2019, 41(5): 649-659.)
- [19] Qin C, Zhang Y Z, Liu Y D, et al. Semantic loop closure detection based on graph matching in multi-objects scenes[J]. *Journal of Visual Communication and Image Representation*, 2021, 76: 103072.
- [20] Korman S, Litman R. Latent ransac[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 6693-6702.
- [21] Gaidon A, Wang Q, Cabon Y, et al. VirtualWorlds as proxy for multi-object tracking analysis[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 4340-4349.
- [22] Rmsalinas. Dbow3[EB/OL]. (2017-02-17). <https://github.com/rmsalinas/DBoW3>.
- [23] Cummins M, Newman P. Appearance-only SLAM at large scale with FAB-MAP 2.0[J]. *The International Journal of Robotics Research*, 2011, 30(9): 1100-1123.
- [24] Siam S M, Zhang H. Fast-SeqSLAM: A fast appearance based place recognition algorithm[C]. 2017 IEEE International Conference on Robotics and Automation. Singapore, 2017: 5702-5708.

作者简介

赵一夫(1995—), 男, 硕士生, 从事视觉SLAM、机器视觉的研究, E-mail: 1456368460@qq.com;

苗津毓(1997—), 男, 硕士生, 从事视觉SLAM、深度学习的研究, E-mail: mjj0519@buaa.edu.cn;

郭方洪(1988—), 男, 副教授, 博士, 从事移动机器人等研究, E-mail: fhguo@zjut.edu.cn;

吴祥(1990—), 男, 助理研究员, 博士, 从事网络化运动控制的研究, E-mail: xiangwu@zjut.edu.cn;

董辉(1979—), 男, 教授, 博士生导师, 从事智能控制、工业物联网等研究, E-mail: hdong@zjut.edu.cn;

俞立(1961—), 男, 教授, 博士生导师, 从事鲁棒控制、网络化控制等研究, E-mail: llyu@zjut.edu.cn.