

doi:10.19306/j.cnki.2095-8110.2021.05.010

一种低数据传输的多机器人实时视觉 SLAM

段胜青¹, 熊智^{1,2}, 赵耀¹, 崔雨晨¹, 周帅琳¹

(1. 南京航空航天大学自动化学院, 南京 211106;

2. 先进飞行器导航、控制与健康管理和工业和信息化部重点实验室, 南京 211106)

摘要:针对多机器人视觉 SLAM 在实际应用中带宽受限的问题, 设计了一种低数据传输的多机器人实时视觉 SLAM 系统。系统中引入了 NetVLAD 神经网络模型, 通过改进 NetVLAD 降低了多机器人回环检测的计算资源占用, 提高了回环检测的实时性。提出了一种针对描述子缺失情况下的特征匹配算法, 提高了回环检测与相对量测的鲁棒性, 并提出了一种增量式多机器人位姿图共享和优化方法。最后, 通过在 KITTI 数据集进行测试, 验证了该 SLAM 系统能有效减少多机器人通信过程中的数据传输, 具有与单机器人 SLAM 相当的定位精度和实时性。

关键词:多机器人 SLAM; 低数据传输; 改进的 NetVLAD; 位姿图优化

中图分类号: TP701

文献标志码: A

文章编号: 2095-8110(2021)05-0071-08

A Real-time Multi-robot Visual SLAM with Lower Data Transmission

DUAN Sheng-qing¹, XIONG Zhi^{1,2}, ZHAO Yao¹, CUI Yu-chen¹, ZHOU Shuai-lin¹

(1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;

2. Key Laboratory of Navigation, Guidance and Health-Management Technologies of Advanced Aircraft, Ministry of Industry and Information Technology, Nanjing 211106, China)

Abstract: A real-time and low-data-transmission multi-robot visual SLAM system is designed for bandwidth limited environment. This system integrates a modified NetVLAD network, which has less computing resource occupation and better real-time performance during the multi-robot loop detection. A feature matching algorithm for the absence of descriptor information is proposed to improve the robustness of loop detection and relative measurement. Moreover, an incremental multi-robot pose graph sharing and optimization method is proposed. Finally, by testing on the KITTI dataset, it is verified that the proposed system has the same positioning accuracy and real-time performance as single-robot SLAM, and can effectively reduce data transmission.

Key words: Multi-robot SLAM; Low data transmission; Modified NetVLAD; Pose graph optimization

收稿日期: 2020-12-11; 修订日期: 2021-01-04

基金项目: 国家自然科学基金(61873125); 江苏省自然科学基金(BK20181291); 中央高校基本科研业务费专项资金(NZ2020004, NZ2019007); 上海航天科技创新基金(SAST2019-085)

作者简介: 段胜青(1996-), 男, 硕士研究生, 主要从事视觉导航及组合导航方面的研究。

通信作者: 熊智(1976-), 男, 研究员, 主要从事惯性导航、视觉导航、组合导航等方面的研究。

0 引言

多机器人系统相比于单个机器人系统在执行任务时具备更高的效率和鲁棒性,可应用于森林/城市救火、自然灾害搜救、水下/空间探测和安全巡检等任务^[1]。多机器人同步定位与构图(Multi-Robot Simultaneous Localization and Mapping, MR SLAM)技术作为多机器人系统的重要组成部分,承担着在全球定位系统(Global Positioning System, GPS)卫星拒止环境下的定位和导航的关键作用。由于相机具有成本低、质量小等特点,以相机为核心的视觉 SLAM 技术近年来得到了广泛的研究和应用。目前,单个机器人 SLAM 技术已经相对成熟,多机器人 SLAM 的主要瓶颈在于通信带宽限制和计算负载,而多机器人 SLAM 的数据交换和计算主要集中在回环检测和后端优化阶段。

回环检测是多机器人 SLAM 的重要基础。在 SLAM 研究的古典时期(classical age)^[2],通常采用激光角点作为路标,并采用最大似然估计的方式对路标进行数据关联^[3-4],此类方法需要持续共享信息矩阵,传输数据量与状态量个数呈平方增长,数据传输量大;此外,由于路标特征简单,容易发生错误关联,因此此类方法适用于人工环境,对于复杂的自然环境鲁棒性较差。A. Howard^[5]和 L. Carlone^[6]等采用机器人携带人工标签的方式实现机器人的检测识别和相对位置量测,特殊的人工标签能有效解决回环检测中错误关联的问题,且仅在检测到回环时交换数据,但要求机器人在执行任务过程中必须相遇,具有较大使用限制。随着图像处理技术的进步,基于特征点的回环检测被应用于单机器人 SLAM 和多机器人 SLAM。P. M. Newman 在 2006 年提出了一种基于尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)特征点的回环检测方法^[7],通过聚类的方法,采用视觉单词对场景进行描述,从而实现实时地回环检测。D. Gálvez-López 提出了一种基于 FAST(Features from Accelerated Segment Test)特征点和 BRIEF(Binary Robust Independent Elementary Features)描述子的词袋模型^[8],实现了更高的实时性,被应用于一些主流的 SLAM 系统^[9-10]。基于词袋模型的回环检测虽然能够减少数据的传输,但准确率有待提高。近年来,神经网络凭借其更好的准确性和鲁棒性开始被应用于 SLAM 回环检测。S. Choudhary 等通过 YOLO(You

Look Only Once)神经网络生成视觉单词进行回环检测^[11],但以 YOLO 为代表的用于分类识别的模型主要关注图片的局部特征,信息利用率较低。此外,刘强等指出图片中的环境建筑等背景对于现有网络模型通常属于非感兴趣区域^[12],不利于回环检测。R. Arandjelovic 等根据传统方法中的局部聚合描述子向量(Vector of Local Aggregated Descriptors, VLAD)设计了一种 NetVLAD 神经网络结构^[13],并以 GPS 信息为标签进行弱监督训练,所生成的描述子包含了图片的全局信息,相比于传统方法能以较短的描述子实现更高的准确率,可有效降低数据传输,比其他网络更加适用于多机器人 SLAM 系统。因此,本文将 NetVLAD 应用于多机器人回环检测,但 NetVLAD 对硬件计算能力有较高要求,难以适用于低成本图形处理器(Graphics Processing Unit, GPU),硬件资源占用和实时性有待进一步优化。

多机器人 SLAM 按照后端优化算法可分为基于滤波的多机器人 SLAM 和基于非线性优化的多机器人 SLAM。基于滤波的多机器人 SLAM 常用的滤波方法有卡尔曼滤波^[4,14]、信息滤波^[3]和粒子滤波^[5-6],但对于以相机为主要传感器的视觉 SLAM 系统,大量量测数据造成协方差矩阵和粒子数目增大,数据传输增大,降低了算法的实时性。H. Strasdat 等指出,在视觉 SLAM 领域非线性优化方法相比于滤波方法具有更好的精度和实时性^[15]。对于多机器人系统,理论上可以使每个机器人收集其他所有机器人的量测信息,然后通过非线性优化估计出所有状态量,但实际应用中通信带宽和计算资源都是有限的,因此该方法并不实用。A. Cunningham 等提出了一种具备分布式数据融合(Decentralized Data Fusion, DDF)功能的平滑与构图方法 DDF-SAM^[16],该方法通过高斯消元降低了通信和计算负载。T. Cieslewski 等提出了一种基于 Gauss-Seidel 算法的分布式非线性优化方法^[17],以解决多机器人 SLAM 通信带宽受限的情况。但以上两种方法本质上是对非线性优化过程中求解正规方程的改进,而 SLAM 系统仍需要多次传输数据和迭代才能逼近最优估计值,考虑到网络传输的延迟,将会降低系统的实时性和鲁棒性。目前,位姿图优化方法因其数据少和计算快的优点,在多机器人 SLAM 领域得到了较多的研究和应用^[7,18]。

综上,在多机器人 SLAM 领域的回环检测和后端优化方面,尽管已有很多方法被提出,但无法完全

满足实际场景对系统实时性、鲁棒性以及带宽限制的要求。为此,本文将引入 NetVLAD 神经网络模型,在 NetVLAD 的基础上提出了一种低数据传输的实时多机器人回环检测方法,并根据现有的非线性优化理论对多机器人后端位姿图优化方法进行分析和改进,以减少通信和计算资源占用,并以 ORB-SLAM2 为前端实现了一个实时多机器人 SLAM 系统。

1 系统概述

本文提出系统如图 1 所示,分为前端、后端和

NetVLAD 三部分。前端部分机器人本地图片采用 ORB SLAM2^[9]前端处理,对于外部图片消息采用 4.2 节算法初始匹配后再采用 ORB SLAM2 前端处理。后端本地位姿图优化采用 ORB SLAM2 方法,多机器人位姿图优化和共享采用第 5 节所述方法。NetVLAD 模块用于计算关键帧的局部特征聚合描述符 (Vector of Local Aggregated Descriptors, VLAD) 向量,根据 VLAD 的相似度(4.1 节)进行多机器人回环帧匹配。

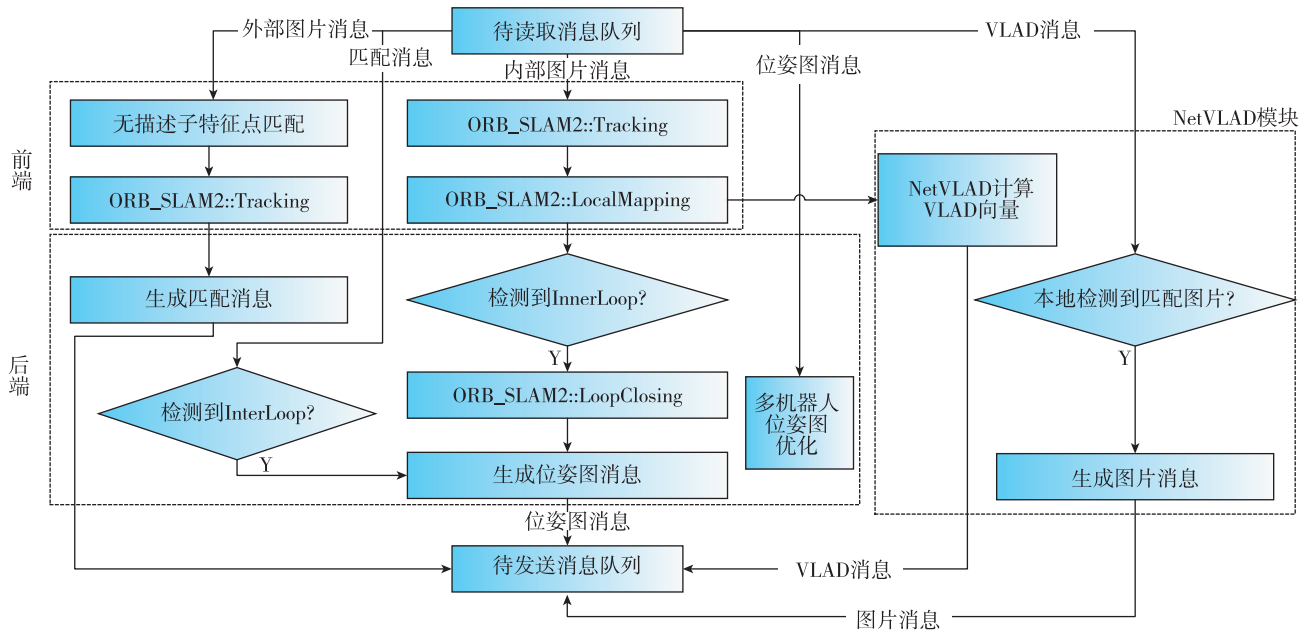


图 1 本文系统流程框图

Fig. 1 Block diagram of our system

2 背景知识

通常 SLAM 问题通过概率模型来求解。设状态量 \mathbf{X} , 量测量 \mathbf{Z} , 则有

$$P(\mathbf{X} | \mathbf{Z}) = P(\mathbf{Z} | \mathbf{X})P(\mathbf{X})/P(\mathbf{Z}) = \eta \prod_{x_i, x_j \in X, z_{ij} \in Z} P(z_{ij} | x_i, x_j) \prod_{x_i \in X} P(x_i) \quad (1)$$

先验分布 $P(\mathbf{X})$ 未知时,可看作实数域上的均匀分布,故可忽略。 \mathbf{X} 的最优估计即为

$$\mathbf{X}_{opt} = \arg \min_{\mathbf{X}} - \ln P(\mathbf{X} | \mathbf{Z}) = \arg \min_{\mathbf{X}} - \sum_{\substack{x_i, x_j \in X \\ z_{ij} \in Z}} \ln P(z_{ij} | x_i, x_j) \quad (2)$$

为了方便描述和分析问题,通常将式(1)和式(2)用因子图的形式描述,节点(Vertex)表示待估计

状态量,边(Edge)表示量测约束 $P(z_{ij} | x_i, x_j)$ 。当式(2)中的概率分布 $P(z_{ij} | x_i, x_j)$ 服从正态分布时,可通过非线性最小二乘方法进行求解。SLAM 系统中的 BA 优化和位姿图优化均是基于上述原理。

3 多机器人 SLAM 模型描述

记 α, β, γ 为机器人的序号,并且 α, β, γ 可指代任意同一个机器人或不同机器人,状态量 $\mathbf{x}_{\gamma\alpha i}$ 定义为关键帧 $F_{\alpha i}$ 在机器人 γ 初始关键帧 $F_{\gamma 0}$ 对应相机坐标系下的位姿,记作

$$\mathbf{x}_{\gamma\alpha i} \triangleq \begin{bmatrix} R_{\gamma\alpha i} & t_{\gamma\alpha i} \\ 0^T & 1 \end{bmatrix} \in SE(3) \quad (3)$$

量测量 $\mathbf{z}_{\alpha i \beta j}$ 定义为当前机器人 β 的关键帧 $F_{\beta j}$ 相对于机器人 α 历史关键帧 $F_{\alpha i}$ 的位姿,记作

$$\mathbf{z}_{\alpha_i\beta_j} \triangleq \begin{bmatrix} \hat{R}_{\alpha_i\beta_j} & \hat{t}_{\alpha_i\beta_j} \\ 0^T & 1 \end{bmatrix} \in SE(3) \quad (4)$$

当 $\alpha = \beta$ 时, $\mathbf{z}_{\alpha_i\beta_j}$ 采用 ORB SLAM2 前端进行求解; 当 $\alpha \neq \beta$ 时, 求解方法见本文 4.3 节。

用 $\rho(\cdot)$ 表示李群到李代数的运算, 量测约束概率模型定义如下

$$\mathbf{e}_{\alpha_i\beta_j} = \rho(\mathbf{z}_{\alpha_i\beta_j} \cdot \mathbf{x}_{\gamma_0\alpha_i}^{-1} \cdot \mathbf{x}_{\gamma_0\beta_j}) \sim N(0, \Sigma) \quad (5)$$

其中, Σ 表示量测模型的方差; $\mathbf{e}_{\alpha_i\beta_j}$ 对应因子图中的一条边。此时

$$\mathbf{X}_{\text{opt}} = \arg \min \sum \mathbf{e}_{ij}^T \Sigma^{-1} \mathbf{e}_{ij} \quad (6)$$

4 多机器人回环检测

4.1 NetVLAD 神经网络模型

NetVLAD 是一种专用于位置识别的神经网络模型, 相比其他神经网络模型具有更高的准确率, 但对 GPU 资源要求较高。为了减少该神经网络模型对 GPU 资源占用、提高回环检测速度, 本文直接对文献[13]训练完成的 NetVLAD 模型做出参数调整而无需重新训练。

4.1.1 conv1_1 卷积层合并

NetVLAD 网络 conv1_1 层卷积核的维度为 $3 \times 3 \times 3 \times 64$, conv1_1 层输出可表示如下

$$O_i = \sum_{c=1}^3 (I_c - a_c) \otimes K_{ic} + b_i \quad (7)$$

其中, O_i 表示卷积层第 i 个输出通道; I_c 对应输入图像的 3 个通道; $a_c \in R$ 为预处理时减去的一个常数; K_{ic} 为卷积核; b_i 为偏置项; \otimes 表示卷积运算。本文以灰度图作为输入, $I_1 = I_2 = I_3$, 因此

$$O_i = I_1 \otimes \sum_{c=1}^3 K_{ic} + b_i - \sum_{c=1}^3 (a_c \sum_{k_{mn} \in K_{ic}} k_{mn}) \quad (8)$$

可以对 conv1_1 层进行合并

$$\begin{aligned} K_{i\text{new}} &= \sum_{c=1}^3 K_{ic} \\ b_{i\text{new}} &= b_i - \sum_{c=1}^3 (a_c \sum_{k_{mn} \in K_{ic}} k_{mn}) \end{aligned} \quad (9)$$

最终 conv1_1 卷积核维度变为 $3 \times 3 \times 1 \times 64$ 。

4.1.2 WPCA 层裁剪

NetVLAD 模型将 vgg16 网络的输出经过 NetVLAD core 处理后, 通过加权主成分分析 (Weighted Principal Component Analysis, WPCA) 将输出降维至 4096 维^[13]。考虑到 PCA 矩阵中越靠后的列包含信息越少, 本文将 WPCA 层 32768×4096 的矩阵裁剪为 32768×128 维, 只保留矩阵的前 128 列, 并

将偏置项裁剪为 128×1 维, 从而极大地减少参数数目, 降低计算量。

4.1.3 相似度计算

图片 U 和 V 根据上面所述改进 NetVLAD 进行计算, 得到归一化 128 维向量 \mathbf{u} 和 \mathbf{v} , 相似度定义如下

$$S(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{128} \mathbf{u}_i \mathbf{v}_i \quad (10)$$

$S(\mathbf{u}, \mathbf{v})$ 越大, 图片越可能构成回环。

4.2 特征匹配

为减少数据传输, 本文所述方法仅发送左目相机中图片特征点像素坐标和对应单词 Id, 以量测相对于其他机器人历史关键帧的位姿。

机器人在接收到上述图片信息后, 由于缺失特征点的描述子信息, 采用 ORB SLAM2 前端会产生大量误匹配 (见第 6 节图 4), 导致相对位姿解算失败。为此, 本文提出了以下方法用于解决描述子缺失情况下的特征匹配。

设图片 i 宽和高分别为 W 和 H , 词袋为 B_i , 特征点集合为 P_i , 对于特征点 $p \in P_i$, 对应单词为 w_p , 像素坐标为 $\mathbf{c}_p = [x_p, y_p]^T$, 记 $\tau_d = (W^2 + H^2)/4$, 匹配算法描述如下:

算法 1 描述子缺失情况下的特征匹配

```

输入  $B_1, P_1, B_2, P_2$ 
 $B = B_1 \cap B_2, M = \emptyset$ 
For 所有  $w \in B$ 
  For 所有  $p_1 \in \{p \mid p \in P_1, w_p = w\}$ 
    For 所有  $p_2 \in \{p \mid p \in P_2, w_p = w\}$ 
      If  $p_2$  未被匹配 &&  $\|c_{p_1} - c_{p_2}\|^2 < \tau_d$ 
         $M = M \cup \{(p_1, p_2)\}$ 
    End
  End
End
End
创建空直方图  $H \triangleq \{h_i \mid 0 \leq i < 30, h_i = \emptyset\}$ 
For 所有  $(p_1, p_2) \in M$ 
   $i = \text{floor}(\arctan(y_{p_1} - y_{p_2}, x_{p_1} - x_{p_2}) / 12^\circ)$ 
   $h_i = h_i \cup \{(p_1, p_2)\}$ 
End
将  $H$  的  $h_i$  按点对个数降序排序  $H_{\text{ordered}} = \text{Sort}(H)$ 
取  $H_{\text{ordered}}$  中分布在  $1\sigma$  范围内 (前 68.3%) 的点对,  $R = H_{\text{ordered}}(0.683)$ 
输出匹配结果  $R$ 

```

通过上述方法可以得到误匹配较少的匹配结果, 此后通过 RANSAC PnP (Random Sample Consensus Perspective-n-Point) 方法求解初始位姿, 再由初始位姿通过反投影方式增加匹配点对。

4.3 回环检测

为减少数据传输,回环检测分为五步:

1) 计算机器人 α 新增关键帧图片 U 的 NetVLAD 特征向量 \mathbf{u} , 并广播给其他机器人;

2) 任意其他机器人 β 收到 \mathbf{u} 后, 与本地 NetVLAD 特征库进行对比, 如果存在特征向量为 \mathbf{v} 的关键帧图片 V 满足 $S(\mathbf{u}, \mathbf{v}) > \tau_{\text{match}}$ (τ_{match} 为相似度阈值), 则将图片 V 的特征点信息(坐标、单词) 发送给机器人 α , 否则结束回环检测;

3) 采用算法 1 进行特征点匹配;

4) 机器人 α 用 RANSAC PnP 方法求解 U 与 V 的相对位姿 $\mathbf{z}_{\alpha U \beta V}$, 如果求解成功, 则认为 U 和 V 对应同一地点, 并广播机器人 α 和 β 的关键帧匹配消息;

5) 每个机器人通过接收关键帧匹配消息判断自身是否与其他机器人轨迹构成回环或间接回环(例如 $A \leftrightarrow B \leftrightarrow C \leftrightarrow A$)。

5 增量式多机器人位姿图优化

先验分布 $P(\mathbf{X})$ 未知时, 因子图只存在双边约束, 可将因子图按照下面的规则分割为多个子图:

1) 构成一个环路的节点和边属于同一子图;

2) 有至少一条公共边的 2 个子图应合并为一个子图;

3) 不经过任何环路的边和其连接 2 个节点构成一个子图。

约定依照上述规则所得的子图称为回环子图。记完整因子图为 G , 回环子图 $S_i \in G$, 经过上述分割, 子图间无法再形成环路, 本文参考 iSAM2^[19] 用贝叶斯树表示各个子图的邻接关系(如图 2 所示)。相比于 iSAM2 的贝叶斯树, 本文中树的节点为非完全子图, 无需三角化, 每次检测到新的回环时只需通过并查集操作合并子图, 从而简化计算。

观察式(5)发现

$$\begin{aligned} \mathbf{e}_{\alpha_i \beta_j} &= \rho(\mathbf{z}_{\beta_j \alpha_i} \cdot \mathbf{x}_{\gamma_0 \alpha_i}^{-1} \cdot \mathbf{x}_{\gamma_0 \beta_j}) \\ &= \rho(\mathbf{z}_{\beta_j \alpha_i} \cdot (T\mathbf{x}_{\gamma_0 \alpha_i})^{-1} \cdot (T\mathbf{x}_{\gamma_0 \beta_j})) \end{aligned} \quad (11)$$

可知式(6)的最优解位姿变换之后仍为它的一组最优解。因此可通过 3 个步骤求解 \mathbf{X}_{opt} :

1) 根据上述规则将因子图划分为回环子图;

2) 根据式(6)求解各个子图的最优估计;

3) 根据式(11)特性, 按贝叶斯树从根节点向下的顺序依此将各个子图的解做整体位姿变换, 以确保相连子图的共享变量数值一致性, 即可求得整个

因子图 G 的最优解。

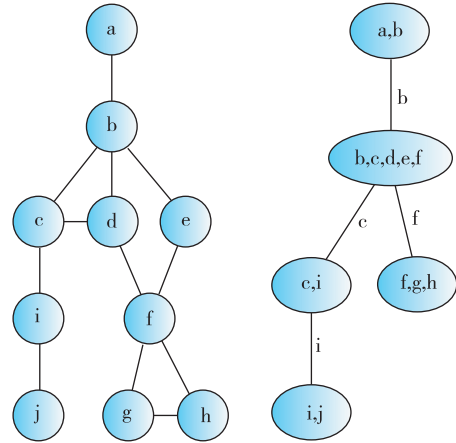


图 2 位姿图(左)和贝叶斯树(右)

Fig. 2 Pose graph (left) and Bayes tree (right)

以图 2 为例, 左图为原始因子图, 右图为所得贝叶斯树, 右图每个节点表示一个回环子图, 每条边上的字母表示 2 个相邻子图的共有节点。分别求出子图 ab 、 $bcdef$ 、 ci 、 fgh 、 ij 的最优解, 再依次对 $bcdef$ 、 ci 、 ij 、 fgh 进行位姿变换, 以保证状态变量 b 、 c 、 f 、 i 的一致性, 即可求得因子图最优解。

根据上述特性, 在多机器人位姿图优化时, 只需要发送与其他机器人构成回环子图的部分边和节点即可, 从而实现了节省带宽的目的。

多机器人后端优化采用如下策略:

1) 当检测到自身与其他机器人构成回环时, 广播位姿图给其他机器人, 且仅广播与其他机器人构成回环子图的边和节点;

2) 广播时, 如果在上次广播到本次广播期间机器人没有发生内部闭环和 BA 优化, 仅广播新增位姿图节点和边, 已广播的节点不需要再传输;

3) 当机器人发生内部闭环和 BA 优化时, 对内部闭环优化影响到回环子图上的边和节点进行重新广播;

4) 每个机器人接收其他机器人广播的信息, 并周期性检测与机器人内部位姿图是否构成回环, 若构成回环, 仅对新增的且包含本地位姿图的回环子图进行优化, 本地其他位姿图按照贝叶斯树做位姿变换即可。

6 实验与分析

本文系统前端和后端部分由 C++ 编程开发, NetVLAD 采用 Python3 和 Tensorflow 实现, 每个机器人都运行有 NetVLAD 回环检测和 SLAM 前

后端 2 个进程,进程间通过机器人操作系统(Robot Operating System, ROS)收发消息,机器人间也通过 ROS 通信,并通过广播的方式进行信息发送,从而实现一对多信息传递。

6.1 NetVLAD 回环检测

采用混淆矩阵的方式检验回环检测效果,并将混淆矩阵用灰度图的形式显示,矩阵的行列对应 KITTI00 数据集中图片编号(0~4540),图片相似度越高,对应像素的颜色越黑。图 3(a)中矩阵计算规则如下

$$a_{ij} = \|t_i - t_j\|_2 > 4? 0 : 1 \quad (12)$$

t_i 和 t_j 表示图片对应的真实位置,即二者距离大于 4m 时 a_{ij} 为 0(白色),否则为 1(黑色)。图 3(b)表示原始 NetVLAD 网络计算所得特征向量根据式(10)计算所得的混淆矩阵,图 3(c)表示根据第 5 节修改后 NetVLAD 网络计算所得特征向量根据式(10)计算所得的混淆矩阵,图 3(d)为对图 3(c)的结果设置 0.85 阈值的二值化处理结果,此时图 3(d)的准确率为 99.44%,召回率为 39.16%,准确度能够满足 SLAM 需求,召回率虽然偏低,但 SLAM 系统在一段连续图片序列中只需要找到一个回环帧即可,因此仍能满足回环检测的需求。

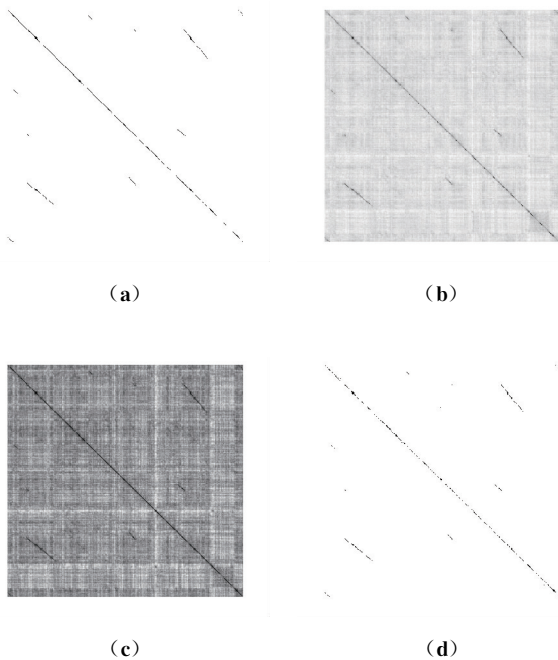


图 3 NetVLAD 混淆矩阵

Fig. 3 Confusion matrices of NetVLAD

分别用不同显卡设备搭载 NetVLAD 网络处理 KITTI00 数据集(4541 张图片),对比结果如表 1 和

表 2 所示。

表 1 Nvidia GTX1650Ti 4G 测试结果

Tab. 1 Test results on Nvidia GTX1650Ti 4G

网络模型	图片处理时间均值/ms	显存占用/MB
原始 NetVLAD	172.49	3617
改进 NetVLAD	32.08	2747

表 2 GTX965M 2G 测试结果

Tab. 2 Test results on Nvidia GTX965M 2G

网络模型	图片处理时间均值/ms	显存占用/MB
原始 NetVLAD	显存不足,无法运行	—
改进 NetVLAD	64.51	1650

通过上述结果可知,改进后的 NetVLAD 拥有更快的计算速度和更少的资源占用,可应用于低成本 GPU,同时能够满足回环检测准确率和实时性需求。

6.2 描述子缺失情况下的特征匹配

图 4 和图 5 所示分别为 ORB SLAM2 和本文算法 1 对 KITTI00 数据集图片 000391 和 002442 的匹配结果,可知 ORB SLAM2 前端匹配算法在无描述子情况下,当 2 幅待匹配图片存在视角偏移时会产生较多误匹配,影响机器人的定位结果(参见图 6),而算法 1 仅存在少量误匹配,不影响系统后续运行。

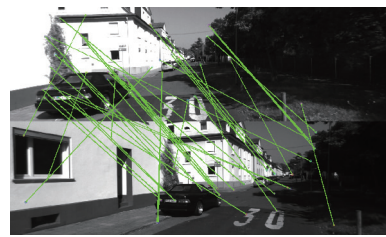


图 4 ORB SLAM2 特征匹配结果

Fig. 4 Feature matching result of ORB SLAM2

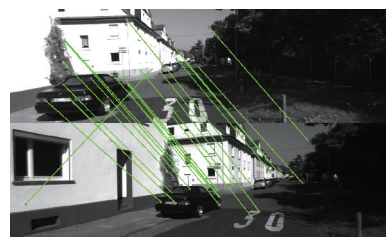


图 5 本文方法特征匹配结果

Fig. 5 Feature matching result of our method

6.3 多机器人协同定位导航轨迹

将 KITTI00 数据集按照图片序号平均划分为 4 段,分别配置在 4 台 i7-6700HQ/Nvidia GTX965M 2G 笔记本电脑上,配置 4 个多机器人 SLAM 系统节点,并且机器人初始位置未知,最终运行结果如图 6、图 7 和表 3、表 4 所示。

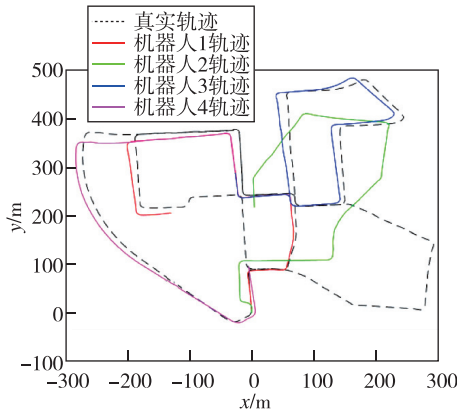


图 6 匹配失败情况下的 MR SLAM 轨迹

Fig. 6 MR SLAM trajectories in case of matching failure

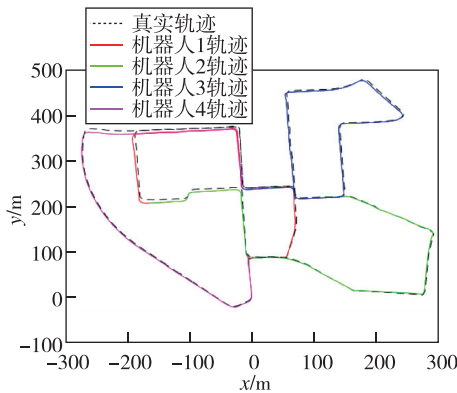


图 7 本文方案轨迹

Fig. 7 MR SLAM trajectories of our system

表 3 本文方案与 ORB SLAM2 精度对比

Tab. 3 Comparison of accuracy: our system vs ORB SLAM2

绝对位置误差(APE)	本文方案	ORB SLAM2
最大误差/m	10.19	12.24
误差均值/m	5.57	6.92
误差中值/m	5.57	6.87
均方根误差/m	6.09	7.64

表 4 本文方案与 ORB SLAM2 实时性对比

Tab. 4 Comparison of real-time performance: our system vs ORB SLAM2

计算速度指标	本文方案	ORB SLAM2
图片处理时间中值/ms	62.29	62.21
图片处理时间均值/ms	62.49	62.25

通过以上数据可以看出,本文提出的多机器人 SLAM 系统能够实现与 ORB SLAM2 系统同样数量级的定位精度和实时性,多机器人 SLAM 的定位精度要略高于 ORB SLAM2,其原因来自于多机器人 SLAM 仅进行了位姿图优化,而未进行 BA 优化,由于 ORB SLAM2 前端特征点匹配存在误差,BA 优化带来了过拟合效应。表 4 中图片处理时间均值小于表 2 中的均值,原因是本文方案中 NetVLAD 仅对关键帧进行处理。

图 8 所示为部署的 4 个机器人节点在运行期间的数据传输统计图,可见在未检测到回环帧时,带宽保持在 10kb/s 以下,在发现回环时发送图片数据会占用 25kb/s 的带宽,位姿图占用带宽与机器人所经过路径长度有关,145s 时刻由于多个机器同时构成回环,造成了位姿图流量高峰,但不超过 40kb/s。整个过程传输数据总量为 1.18MB,位姿图消息 105.16kB,图片消息 265.22kB(11 张图片),其他消息(主要为 NetVLAD)814.13kB,证明了本文提出方法可有效降低带宽负载。

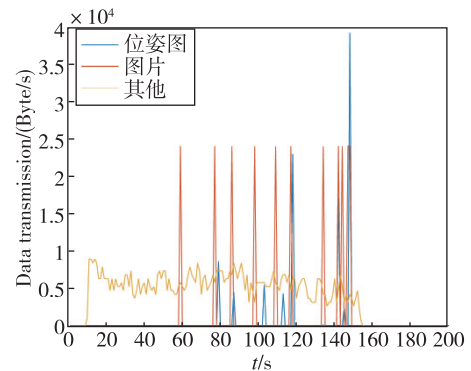


图 8 运动过程中每秒数据传输量

Fig. 8 Data transmission per second during the movement

7 结论

本文提出了一种低带宽占用的多机器人 SLAM 系统。试验结果表明:

1)本文对 NetVLAD 神经网络的改进可有效提高 NetVLAD 计算速度和回环检测实时性,并降低显存占用和通信带宽占用。

2)本文提出的图片消息压缩方法和对应的特征点匹配算法能极大地减少数据发送,并保证匹配和量测的鲁棒性。

3)本文提出的多机器人视觉 SLAM 系统能够达到和单机器人 ORB SLAM2 同样的精度和实时性,同时有效降低了机器人之间的数据传输。

4)本文的图片消息采用发送所有特征点的方式,对于通信带宽限制较大的场景,应该对特征点进行进一步筛选。位姿图共享也需要进行合理规划,分批发送,避免造成流量高峰时的拥堵情况。

参考文献

- [1] Saeedi S, Trentini M, Seto M, et al. Multiple-robot simultaneous localization and mapping: a review[J]. *Journal of Field Robotics*, 2016, 33(1): 3-46.
- [2] Cadena C, Carlone L, Carrillo H, et al. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age[J]. *IEEE Transactions on Robotics*, 2016, 32(6): 1309-1332.
- [3] Thrun S, Liu Y. Multi-robot SLAM with sparse extended information filters[C]// *Proceedings of 11th International Symposium Robotics Research (ISRR)*. Siena, Italy. DBLP, 2003.
- [4] Zhou X S, Roumeliotis S I. Multi-robot SLAM with unknown initial correspondence: the robot rendezvous case [C]//*Proceedings of IEEE/RSJ International Conference on Intelligent Robots & Systems*. IEEE, 2006.
- [5] Howard A. Multi-robot simultaneous localization and mapping using particle filters[J]. *The International Journal of Robotics Research*, 2006, 25(12): 1243-1256.
- [6] Carlone L, Ng M K, Du J, et al. Simultaneous localization and mapping using rao-blackwellized particle filters in multi robot systems[J]. *Journal of Intelligent & Robotic Systems*, 2011, 63(2): 283-307.
- [7] Newman P M, Cole D M, Ho K. Outdoor SLAM using visual appearance and laser ranging[C]// *Proceedings of IEEE International Conference on Robotics & Automation*. IEEE, 2006.
- [8] Gálvez-López D, Tardós J D. Real-time loop detection with bags of binary words[C]// *Proceedings of IEEE/RSJ International Conference on Intelligent Robots & Systems*. IEEE, 2011.
- [9] Mur-Artal R, Tardos 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.
- [10] Tong Q, Li P, Shen S. VINS-Mono: a robust and versatile monocular visual-inertial state estimator[J]. *IEEE Transactions on Robotics*, 2017(99): 1-17.
- [11] Choudhary S, Carlone L, Nieto C, et al. Distributed mapping with privacy and communication constraints: lightweight algorithms and object-based models[J]. *International Journal of Robotics Research*, 2017, 36(12): 1286-1311.
- [12] 刘强, 段富海, 桑勇, 等. 复杂环境下视觉 SLAM 闭环检测方法综述[J]. *机器人*, 2019, 41(1): 112-123.
- Liu Qiang, Duan Fuhai, Sang Yong, et al. A survey of loop-closure detection method of visual SLAM in complex environments[J]. *Robot*, 2019, 41(1): 112-123(in Chinese)
- [13] Arandjelovic R, Gronat P, Torii A, et al. NetVLAD: CNN architecture for weakly supervised place recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(6): 1437-1451.
- [14] Fenwick J W, Newman P M, Leonard J J. Cooperative concurrent mapping and localization[C]//*Proceedings of IEEE International Conference on Robotics & Automation*. IEEE, 2002.
- [15] Strasdat H, Montiel J M M, Davison A J. Visual SLAM: why filter? [J]. *Image and Vision Computing*, 2012, 30(2): 65-77.
- [16] Cunningham A, Indelman V, Dellaert F. DDF-SAM 2.0: consistent distributed smoothing and mapping [C]// *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2013.
- [17] Cieslewski T, Choudhary S, Scaramuzza D. Data-efficient decentralized visual SLAM[C]// *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018.
- [18] Chen Y, Zhao L, Lee K, et al. Broadcast your weaknesses: cooperative active pose-graph SLAM for multiple robots[J]. *IEEE Robotics and Automation Letters*, 2020, 5(2): 2200-2207.
- [19] Kaess M, Johannsson H, Roberts R, et al. iSAM2: incremental smoothing and mapping using the Bayes tree[J]. *The International Journal of Robotics Research*, 2011, 31(2): 216-235.

(编辑:孟彬)