

基于图优化的同时定位与地图创建综述

梁明杰¹, 闵华清², 罗荣华¹

(1. 华南理工大学计算机科学与工程学院, 广东 广州 510006; 2. 华南理工大学软件学院, 广东 广州 510006)

摘要: 基于图优化的同时定位与地图创建 (SLAM) 是当前机器人领域的研究热点. 从帧间配准、环形闭合检测以及优化技术 3 个主要方面对基于图优化的同时定位与地图创建进行综述. 对每一个方面, 阐述其关键技术, 介绍最新研究进展, 并探讨相关难点问题及解决思路. 最后, 对基于图优化的同时定位与地图创建的发展作出展望.

关键词: 同时定位与地图创建; 配准; 环形闭合检测; 图优化

中图分类号: TP24

文献标识码: A

文章编号: 1002-0446(2013)-04-0500-13

Graph-based SLAM: A Survey

LIANG Mingjie¹, MIN Huaqing², LUO Ronghua¹

(1. School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China;

2. School of Software Engineering, South China University of Technology, Guangzhou 510006, China)

Abstract: Graph-based simultaneous localization and mapping (SLAM) is currently a hot research topic in the field of robotics. Frame-to-frame alignment, loop closure detection and graph optimization are three main aspects in graph-based SLAM. For each aspect, the key techniques and current progress are discussed, and the difficult problems and the possible solutions are also analyzed. Finally, the potential future issues and research trends are explored.

Keywords: simultaneous localization and mapping; registration; loop closure detection; graph optimization

1 引言 (Introduction)

未知环境中, 机器人的自主地图创建与自定位密切相关, 相互依赖. 同时定位与地图创建 (SLAM) 将二者相结合, 其主要思想是: 一方面, 依靠已创建的地图信息进行自定位; 另一方面, 根据定位结果对地图进行更新. SLAM 已成为近年来机器人领域的热点研究问题, 并被认为是实现真正自主机器人的核心环节^[1]. SLAM 本质上是一个状态估计问题. 当前, 其求解方法可大致分为两大类, 一类是基于滤波器的方法, 另一类则是基于平滑的方法.

基于滤波器的方法主要是利用递归贝叶斯估计原理, 在假定从 0 到 t 时刻的观测信息以及控制信息已知的条件下, 对系统状态 (包括机器人当前位姿以及所有地图特征位置) 的后验概率进行估计. 根据后验概率表示方式的不同, 存在多种基于滤波器的方法. 常用的包括扩展卡尔曼滤波 (EKF) 方法^[2]、扩展信息滤波 (EIF) 方法^[3]、以及粒子滤波 (PF) 方法^[4]等. 为强调其增量式特性, 基于滤波的 SLAM 方法通常也被称为在线 SLAM (on-line SLAM). 对

该类方法, 研究者们已做过系统的分析和总结^[1,5], 在此本文不再累赘. 值得注意的是, 基于滤波器的方法存在线性化及更新效率等问题, 这使得它难以应用于大规模环境的地图创建^[6-7].

与滤波方法中每一步只考虑机器人当前位姿不同, 基于平滑的方法通过所有的观测信息估计机器人完整的运动轨迹及地图, 因而也被称为完全 SLAM 方法 (full SLAM). 由于地图特征可以通过边缘化方法转化为位姿间的约束, 从而简化为对位姿序列的估计^[8]. 这类方法可以用图的方式作直观描述, 所得的图被称作位姿图 (pose graph). 图中的节点对应机器人在不同时刻的位置和姿态, 而边则描述了位姿与位姿间的空间约束关系. 这种约束可以通过里程计或观测信息的配准得到 (后文将作进一步阐述). 在图构造好后, 对图中节点所处的位置 (在姿态空间中) 进行优化, 使其最好地满足边所表示的约束关系, 优化的结果即对应机器人的运动轨迹. 由于平滑方法的这种直观特性, 又常被称为基于图优化的方法. 过去, 基于图优化的方法被认

为太耗费时间, 无法满足 SLAM 的实时性要求, 但随着高效求解方法的出现, 基于图优化的方法重新得到重视. 在经历一个复兴后, 其成为当前 SLAM 研究的一个热点^[9-16].

尽管 SLAM 问题已经获得相当深入的研究, 不少学者也对该研究作过详细的综述, 但这些年来对该问题的研究又有了新进展, 如运动从 3DOF 到 6DOF^[17-18], 地图从 2D 到 3D^[19], 环境有从室内逐步转向室外的趋势, 地图规模不断扩大^[20], 更兴起了对终生地图创建 (lifelong mapping) 的研究^[21]. 另外, 近年来在传感器使用方面也有新变化, 新型 RGB-D 摄像机^[22] 的出现, 为具有丰富 3 维空间信息与颜色纹理信息的地图创建提供了新的可能. 采用 RGB-D 摄像机对环境进行 3 维重构迅速发展成为研究的热点^[23-27]. 在最新研究中, 相当一部分都采用了基于图优化的 SLAM 技术. 而受当时研究方法所限, 已有的关于 SLAM 的综述性文章往往只重点关注了基于滤波器的方法^[1,5], 对基于图优化的方法则缺少分析和总结. 鉴于此, 本文拟从帧间配准、环形闭合检测以及图优化技术 3 方面出发, 对基于图优化的 SLAM 技术进行综述.

2 基于图优化的 SLAM 框架 (Framework for graph-based SLAM)

基于图优化的 SLAM 方法 (graph-based SLAM) 由 Lu 与 Milios^[28] 首先提出. 文中以基于激光传感器的平面地图创建为例, 阐述了基于图优化的 SLAM 技术的基本思想, 即通过保留所有的观测数据帧以及帧间的空间约束关系, 并将约束看作是随机观测, 然后采用最大似然方法估计机器人的位姿. 这种思想可以通过图的方式形象地表现出来. 用图节点表示待求解的机器人位姿, 用节点间的边来描述位姿间的空间约束关系, 则对位姿序列的估计可以转化为图的优化问题, 即通过调整图中节点的位置使其最好地满足边的约束关系. Gutmann 和 Konolige^[29] 在 Lu 与 Milios 的基础上提出了高效的环形闭合检测及图构建方法, 从而形成了基于图优化的增量式 SLAM 算法框架, 其主要包括顺序配准、环形闭合检测以及图优化 3 大部分. 它们之间的关系可以通过图 1 来表示.

顺序配准主要考虑连续数据帧间的匹配及相对姿态估计问题, 而环形闭合检测则主要根据观测数据判断机器人是否处在之前已访问区域. 这两部分都与观测数据密切相关. 从数据处理的观点来看,

两者都是要解决数据关联问题, 前者考虑局部数据关系, 而后者则处理全局数据关系. 从基于图的表示上看, 顺序配准与环形闭合检测都是根据观测信息建立图节点间的约束, 即完成图的构建. 两者一起被称为 SLAM 前端 (SLAM front-end).

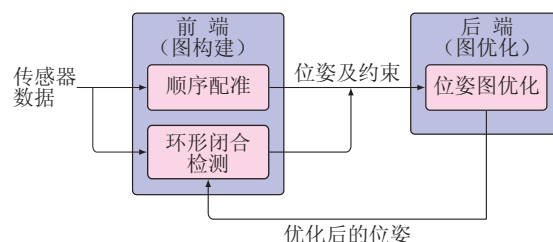


图 1 基于图优化的 SLAM 框架

Fig.1 Framework for graph-based SLAM

由于观测噪声以及配准误差的存在, 通过观测信息配准得到的位姿图往往不具备一致性. 相反, 图中的边与边之间常常存在“冲突”. 举例来说, 若用 T_i 来表示数据帧间的相对变换矩阵而且 T_0, T_1, \dots, T_n 构成一个闭环的话, 从理论上讲, 必然存在 $T_0 T_1 \dots T_n = I$, 其中 I 为单位矩阵. 但通过观测信息配准得到的相对变换矩阵通常不满足该理论约束. 在基于图的形式化表示中, 机器人的位姿被看作是随机变量, 位姿间的约束则是与随机变量相关的观测, 那么图优化结果则对应于位姿的最大似然估计 (详见第 5 节). 与顺序配准及环形闭合检测不同, 图优化部分一般不直接处理观测数据, 而只是对 SLAM 前端构造的图进行优化. 因而, 图优化通常又被称为 SLAM 后端 (SLAM back-end), 与 SLAM 前端相区分.

3 帧间配准 (Frame-to-frame alignment)

帧间配准是指根据观测的两帧数据, 建立数据间的对应关系并求解帧间的相对变换, 从而可以将原不同坐标系下的观测数据放置在同一坐标系中. 由于两帧数据分别表示在对应的传感器坐标系下, 所求的相对变换即为传感器的相对位姿. 进行帧间配准时, 用于校准的帧通常被称为模型帧或参考帧, 被校准的帧通常被称为场景帧.

机器人自运动信息的估计是机器人自定位和地图创建的基础. 通常, 自运动信息可以通过编码器 (encoder)、惯量传感器 (IMU) 或者全球定位系统 (GPS) 来获得. 但在缺少这些设备或其不可用的情况下 (如 GPS 不能用于室内、水下及太空环境), 连续帧间的相对运动也可以通过观测信息的配准得到. 与轮子里程计 (wheel odometry)^a 相比, 基于

^a一般又直接称为里程计, 这里加以强调是为了与下文中的视觉里程计 (visual odometry) 相区分.

帧间配准的方法具有不受地形及运动方式影响的优点,而且往往可以具有更高的估计精度^[30].另外,轮子里程计通常只能完成3自由度的运动估计,但帧间配准的方法可以方便地扩展到6自由度(需要观测数据的支持).因而,帧间配准方法不仅用于室内平坦地面环境中工作的机器人,还被广泛用在野外移动机器人^[31-32]、水下机器人^[33]、自主飞行器^[34]甚至火星探测器^[35]当中.值得注意的是,帧间配准除了可以作连续运动估计外,还可以用于计算空间上相邻而时间上不连续的两帧数据间的相对变换,也即构造环形闭合约束.这是能够对不断行进中机器人运动累积误差进行矫正的前提条件,对同时定位与地图创建具有重要意义.下面将分别对空间点配准及图像配准相关技术进行介绍分析.

3.1 空间点配准

迭代最近点(iterative closest point, ICP)^[36-38]是空间点配准的经典方法.这里的空间点包括2维或3维欧氏空间中的点集,例如这些点可以通过2维或3维激光传感器获得,它们表示在以传感器为原点建立起的坐标系中.

假设模型数据集与场景数据集分别用 $\{p_i\}_i^M$ 与 $\{q_j\}_j^N$ 来表示,数据间的相对变换用齐次变换矩阵 T 来表示.标准的迭代最近点方法通过迭代执行以下两个步骤对点集进行配准:

(1) 对应关系建立

即通过最近邻(nearest neighbor, NN)原则建立场景数据集与模型数据集之间的对应关系.对应关系可以用映射 $\eta(j)$ 来表示,其形式化描述如下:

$$\eta_k(j) = \underset{i \in \{1, \dots, M\}}{\operatorname{argmin}} \|T_{k-1} \cdot q_j - p_i\| \quad j = 1, 2, \dots, N \quad (1)$$

其中, $\|\cdot\|$ 表示空间点的距离, k 为迭代次数. 当 $k = 1$, 即第1次迭代时, 假定 T_0 已知. 求解最近邻的一般方法是逐一比较, 由于需要求解 N 个数据的最近邻, 其时间复杂度为 $O(NM)$. K-D 树通过树型结构来描述空间点的分布, 可被用来加速最近邻点的查找^[38].

(2) 相对变换求解

即根据上一步中建立的对应关系, 通过改变变换矩阵以使得对应点间的距离的平方和最小. 可形式化描述为

$$T_k = \underset{T}{\operatorname{argmin}} \sum_{j=1}^N \|T \cdot q_j - p_{\eta_k(j)}\|^2 \quad (2)$$

其中 T_k 为本次迭代中要求解的变换矩阵. 这一步实际上是要解决已知对应关系情况下相对变换的求解问题, 其中较为常用的是基于 SVD 分解的求法^[39].

Eggert 等对这一问题的4种解析求法作了比较^[40]. 另外, 式(2)中使用的点到点间的距离也并非衡量当前配准好坏程度的唯一准则. 如, Yang 等^[37]采用点到面(point-to-plane)的距离减少迭代的次数, 加快了收敛速度; Segal 等^[41]提出面到面(plane-to-plane)距离的解决方案, 提高了配准的精度及鲁棒性, 代表当前发展的高水准.

由于迭代最近点使用的是由最近邻原则建立起的近似对应关系, 其需要依靠迭代的方式进行渐进的修正. 在迭代的过程中, 对应关系 η 以及相对变换 T 都会被不断地更新, 而且两者相互影响. 可以证明, 经过若干次迭代后, 迭代序列 $(\eta_1, T_1), (\eta_2, T_2), \dots, (\eta_n, T_n)$ 将收敛到局部最优. 因而可以选取收敛后的结果作为问题最终解, 即 $T^* = T_n$.

迭代最近点方法一个公认的缺点是对初始值有较大的依赖性. 当场景数据集与模型数据集之间相差较远(存在较大的平移或旋转), 并且不能提供有效初始值(这时候会选择设定 $T_0 = I$) 时, 上述方法往往会陷入到非全局最优的局部区域中, 并最终导致错误的收敛结果. 因而, 迭代最近点通常用于场景与模型相差不大或可以提供较为可靠初始值的环境中, 主要担当配准细化的任务^[42]. 迭代最近点还有很多细致的研究, 包括 Turk 等^[43]通过均匀采样进行控制点选取以减少计算量, Zhang 等^[38]采用设定阈值的方法对错误匹配进行剔除, Godin 等^[44]根据点间距离为对应点赋予不同的权重以平衡对应点的重要性, Fitzgibbon 等^[45]直接采用非线性优化方法对问题进行求解, 并通过实验验证其与基于解析求解的方法具有相当甚至更好的效果等. 文[42,46]对 ICP 进行了较为全面的综述.

3.2 图像配准

图像配准是根据拍摄的图像信息对摄像机的运动状态进行估计, 这一技术在计算机视觉领域又常被称为视觉里程计(visual odometry, VO)^[31,47]. 基于特征的图像配准过程一般包括特征检测、特征匹配以及运动估计3个步骤, 下面将分别进行介绍.

3.2.1 特征检测

特征检测是视觉里程计最为基础但又至关重要的一个步骤, 因为所有的后续处理都建立在所提取特征的基础上. 对图像进行特征提取, 应着重考虑特征的5大特性, 即重现性、准确性、显著性、鲁棒性以及高效性. 一般而言, 以上特性的性能越高, 对图像的配准越有利. 但有些特性在一定程度上可能存在冲突, 如特征的显著性好, 其鲁棒性可能就稍差, 反之亦然. SIFT^[48]具有较好的重现性、准确性

及鲁棒性, 已成功应用于物体识别、目标跟踪、场景分类以及视觉里程计等计算机视觉领域, 被验证具有非常好的实验效果. SURF^[49] 在 SIFT 基础上采用格子滤波来近似高斯滤波, 大大地提高了计算效率, 但在鲁棒性上比 SIFT 稍差. CenSurE^[50] 是专门为图像配准设计的特征, 在重现性、准确性以及鲁棒性上与 SIFT 性能相当, 但计算效率却远远高于后者, 因而更能满足视觉里程计实时性的需要.

3.2.2 特征匹配

特征匹配是要找出图像特征之间的对应关系. 有两种不同的思路: 一是只对一帧图像进行特征提取, 并依靠跟踪技术在另一帧图像中找到其对应点, 这通常只需要在检测特征周围进行局部的检索^[51]; 二是分别对两帧图像进行特征检测, 然后依据特征的相似性建立起对应关系. 我们将其分别称为特征跟踪与特征识别. 前者只作局部搜索, 具有速度快、稳定性好等特点 (较少出现匹配不一致的情况), 其存在的缺陷是只能用于两帧图像变化相对较小的情况; 后者对每个特征作全局的查找, 对帧间变化较大的情况具有较好的适应性, 但效率较低, 而且匹配不一致的情况时有发生, 可以通过距离比测试 (distance ratio test)^[48] 只选择没有歧义的匹配, 或者利用几何约束关系剔除错误的匹配^[52].

3.2.3 运动估计

特征的位置信息可以直接使用图像坐标来表示, 也可以通过其它方式获取特征在 3 维空间中的坐标^b, 将其描述在 3 维物理空间中. 受特征描述方式的影响, 匹配关系有 3 种不同的表示方式, 分别为 2D 到 2D、3D 到 3D 以及 3D 到 2D.

(1) 2D 到 2D 在 2D 到 2D 的匹配关系描述中, 特征的位置用图像坐标来表示. 经校准的两帧图像 I_{k-1} 与 I_k 间的几何关系可以通过本质矩阵 (essential matrix) 来描述. 借助于极线约束, 本质矩阵可以从 2D 到 2D 的对应关系中求得. 理论证明, 求解本质矩阵的对应关系的数目至少为 5 组. Nister 等^[53] 提出的 5 点法是求解该问题的经典方法.

(2) 3D 到 3D 当特征及其对应关系直接描述在 3 维空间中时, 变换关系的求解与空间点配准方法相似. 不同点在于通过特征匹配建立的对应关系是确定的, 因而不需要进行迭代修正. 评价的标准依然是匹配点间距离的平方和, 具体可以参考式 (2) 以及 3.1 节相关内容.

(3) 3D 到 2D 3D 到 2D 的匹配表示描述 3 维特征及其在图像上投影之间的对应关系. 与 3D 到

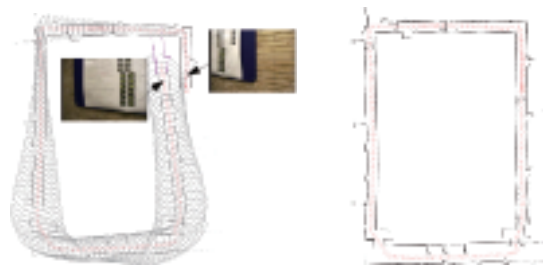
3D 中直接优化空间点间的距离不同, 3D 到 2D 通常采用最小化重投影误差 (re-projection error) 的方法来计算帧间的相对变换. 为了表示方便, 不妨假定已建立好对应关系 $\{(q_i, u_i)\}$, 其中 q_i 表示特征点在 3 维空间中的位置, u_i 表示其在图像上投影的坐标, 则变换矩阵求解如下:

$$T^* = \operatorname{argmin}_T \sum_i \|f(q_i; T) - u_i\|^2 \quad (3)$$

其中, f 是投影函数. 这就是 n 点透视问题 (perspective-n-point, PnP)^[54]. 这一问题已经被相当广泛地研究, 并存在很多不同的解法. 3 组对应关系是求解该问题的最小规模 (称为 P3P^[55]), 其在求解问题中具有重要价值, 如可以结合 RANSAC^[52] 剔除错误匹配. Nister^[47] 通过实验证明, 采用 3D 到 2D 匹配关系并通过最小化重投影误差的方法所取得的效果要比采用 3D 到 3D 匹配关系并直接最小化空间距离的方法得到的效果更好.

4 环形闭合检测 (Loop closure detection)

在 SLAM 中, 环形闭合检测是指根据传感器信息判断机器人当前是否处在之前已经访问过的某个区域, 或者说, 机器人当前所处的位置在创建的地图中是否已有相应的描述. 环形闭合检测的重要性体现在, 正确的闭环信息可以用于修正里程计误差, 从而得到信息一致的地图 (见图 2); 而错误的闭环信息不仅会对后续图优化处理造成干扰, 甚至可能完全毁坏已有的地图创建结果. 本节先分析造成环形闭合问题难解的主要原因, 在此基础上有针对性地对环形闭合检测相关技术进行阐述与分析.



(a) 累积误差导致地图欠缺一致性 (b) 利用闭环约束优化后的结果图 2 环形闭合检测及地图创建结果^[56]

Fig. 2 Loop closure detection and the map built^[56]

4.1 存在的问题

环形闭合检测是未知环境地图创建中的一个难点问题. 其之所以困难, 主要体现在 3 个方面:

(1) 感知歧义 即相似的观测不一定来自相同的场景, 从而引起错误的环形闭合判断. 一方面传

^b单目及双目摄像机可以通过三角化的方法获得, 而 RGB-D 传感器则可以根据深度图像通过简单的等比原理获得.

感器只能获得环境的部分信息,观测数据可分辨性有限;另一方面,无论是在室内还是在室外环境中,都有大量相似的事物存在,如相似的桌子、椅子、墙壁、树木、建筑等,加剧了判断的难度^[57]。

(2) 数据规模大 在进行环形闭合检测时,需要将当前的观测数据与之前已有的观测数据作比较,以计算它们之间的相似性或对它们出自同一地点的概率进行估计。要处理的数据随运行时间或访问过的地点的增加而不断增长,要求能处理的规模达成千上万帧^[58-59]。

(3) 评价要求高 环形闭合检测的作用是可以增加新的约束,从而减少甚至是消除增量式估计中引起的累积误差。但错误的环形闭合不仅会对结果的精度产生影响甚至可能导致优化时无法收敛或收敛到错误的结果^[15]。可见,判断的准确性对环形闭合检测极其关键。常用于评价环形闭合检测性能的一个指标是准确率-召回率曲线。准确率描述了正确检测出来的闭环占总检测闭环的百分比,召回率则描述了正确检测出来的闭环占有所有真实闭环的百分比。准确率要非常高,要求接近 100%,同时召回率也不能太低,否则无法通过闭环信息对运动轨迹进行有效的纠正。

4.2 环形闭合检测技术

环形闭合检测以其重要性及挑战性吸引了广大学者参与其研究。在众多的环形闭合检测技术中,以基于视觉信息的环形闭合检测最为成功,其已经广泛应用在各式 SLAM 系统中。基于以上原因,本节介绍以基于视觉的环形闭合检测为主。下面分场景表示、假设产生以及假设验证 3 个方面对环形闭合检测的相关技术进行阐述。

4.2.1 场景表示

(1) 基于特征表示

场景表示最为直接且简单的方法是利用从传感器数据上提取的特征进行描述,每一帧数据上的特征构成一个场景。特征可分为全局特征与局部特征。如, S nderhauf 等^[60]提出计算图像的 BRIEF-Gist 描述子作为场景的描述, Liu 等^[61]采用图像的 Gabor-Gist 描述子,并通过 PCA(主元分析)降维以提高计算效率及节省存储空间。Newman 等^[56]先通过信息熵检测图像中的显著区域并提取位于区域内的稳定特征(如 MSER^[62]),然后利用提取的特征来构建场景数据库。Zhang^[58]则利用不同尺度下特征匹配重现性的差异对特征进行选择,提高了场景的可分辨性。相似地, Steder^[63]也采用局部特征表示方法,并将所有的特征视为高维空间中的点,通

过 K-D 树组织在一起,从而提高了查询效率。

与全局特征相比,局部特征对部分可见、遮挡情况下的辨识具有更好的适应性和鲁棒性,因而也更有利于减少感知歧义,增强判断的可靠性,但缺点是其通常需要存储并处理更多的数据。

(2) 基于词袋表示

词袋表示的主要思想是将从图像中提取的局部特征进行聚类(如通过 K 均值聚类方法^[64]),将连续变化的特征转变为离散化的“词”,然后采用词的统计直方图对场景进行描述^[65]。这实际上是借鉴了文本信息检索中的词袋(bag of word, BoW)^[66]的思想,在计算机视觉领域又常被称为视觉词袋(bag of visual word)方法。离散化的好处是不仅能归类相似的特征及压缩存储空间,而且可以通过倒排技术快速检索到包含指定特征词的场景,从而避免逐帧比较。Nister 等^[67]在词袋思想基础上提出基于树型结构的存储和管理方式,从而大大地提升了检索效率。Schindler 等^[68]分析了词典构造中特征的选取问题,并利用信息增益对特征进行评估,只挑选区分性好的特征进行词典构造,从而使方法的性能和扩展性显著提高。Cummins 等^[57,59]考虑词与词的相互关系,并采用 Chow-Liu 树来近似描述它们间的相关性,通过利用上下文信息来减少感知歧义。Angeli 等^[69]则探讨了增量式词典构造的相关问题。

Zhang^[58]通过实验证实,离散化会降低特征的可分辨性,从而可能引起感知歧义。但基于词袋的表示已成功应用于不同场合下的环形闭合检测,取得了非常好的实验效果。尤其是 Cummins 等提出的 FAB-Map^[57,59],代表了视觉环形闭合检测当前发展的高水准。

(3) 基于场景地图表示

对创建度量地图(metric map)的 SLAM 方法而言,通过以上方法进行环形闭合检测均需要额外的存储空间,如保存特征、词袋以及词典等。另外一种思路是直接利用创建的度量地图,这样可以减少不必要的存储开销。Williams 等^[70]提出通过建立当前帧特征与地图特征之间的对应关系来求解摄像机可能的位姿,也即利用重定位(relocalization)来进行环形闭合检测。另外,为了充分利用上下文信息,还可以通过构建子地图,然后采用地图与地图相匹配的方法作检测^[71],这种方法尤其适用于通过子地图来构建全局地图的 SLAM 技术。

4.2.2 假设产生

环形闭合假设产生是要从规模庞大的场景中快速地找出若干环形闭合的候选。根据该过程中是否

依赖位姿及其不确定性的估计, 可分为基于空间位置与基于外观相似的两种计算方法。

(1) 基于空间位置的方法

环形闭合是指机器人再次回到之前已经访问过的区域, 因而判断给定的两个位姿是否处在同一场景的最为直观的方式就是计算位姿间的距离。倘若距离值足够小, 则可能构成环形闭合。但由于累积误差的存在, 在经过长时间运动后, 即使机器人真实处在相同的位置, 估计的位姿间也会存在较大的偏差。若能考虑运动过程中的不确定性, 则可以得到一个合理的解决方案。Bosse 等^[72]提出通过 Dijkstra 最短路径算法来计算位姿间的最小不确定性路径。在此基础上, Olson 考虑了传感器测量范围, 并通过位姿间的马氏距离判断观测信息存在重合的可能性^[73]。

这类方法的局限性在于, 当实际误差大于估计误差时, 正确的环形闭合信息将会被忽略掉^[74]。如, 判定环形闭合假设的一个常用准则是落入估计值的 3 倍标准差之内, 若超出此范围则直接被忽略。

(2) 基于外观相似的方法

Newman 等^[56]认为, 要保证环形闭合检测的鲁棒性, 就不能过分依赖地图创建过程中的同源信息(如机器人的姿态信息), 并提出利用视觉显著特征来判断环形闭合假设的方法。通过观测信息来判定环形闭合即是要估计观测信息来自同一场景的可能性。针对不同的表示, 所使用评价标准会有所不同。如对采用二进制的特征描述, Sünderhauf^[60]等直接计算特征间的海明距离来衡量观测的相似性; 对基于词袋的表示, 常采用的是先进行 TF-IDF(词频-逆向文件频率)权重计算再进行余弦相似度求解的方法^[65]; 在考虑词间相互关系的情况下, Cummins 等^[57]为每一个地点建立概率观测模型, 并利用贝叶斯框架计算不同的观测来自同一地点的概率。

在获得相似性度量后, 可通过设定阈值来选取符合条件的环形闭合假设, 但更为常见的方法是对相似性进行排序, 然后选取前 N 项作为候选。对候选数目的选定, Konolige 等^[75]在实验中选取前 15 项作为假设, 获得以下结果: 对 97% 的帧至少可以找到一个好的环形闭合假设, 对 90% 的帧可以找到两个; 对所有的帧, 约 60% 的正确环形闭合都出现在前 15 个假设中。

4.2.3 假设验证

由于环形闭合信息的至关重要性, 对环形闭合假设进行验证是必要的。在此, 我们将验证方法分

为单环形闭合假设验证与多环形闭合假设验证。

(1) 单环形闭合假设验证

单环形闭合假设验证的主要依据是特征的一致性。在给定一组可能的特征匹配的情况下, 对特征作空间结构的一致性检验最为常用的方法是通过 RANSAC 以及极线约束来求解本质矩阵或相对变换^[59,76-77]。若能得到合适的本质矩阵或相对变换则表示检验通过, 否则为不通过。另外, Cadena 等^[78]提出利用条件随机场匹配(CRF-matching)对特征的外观、分布等作一致性检验, 并通过实验验证了方法的有效性。

(2) 多环形闭合假设验证

与单环形闭合假设验证方法不同, 多环形闭合假设验证的主要依据是位姿的一致性。我们知道, 对构成环形闭合的一组相对变换, 理论上, 它们之间的乘积等于单位矩阵。对给定的一组环形闭合假设, 可以利用上述性质对它们两两之间的一致性进行检验, 得到“一致性矩阵”(consistency matrix), 并依此选出最大自洽子集(maximally self-consistent set)作为多环形闭合假设验证的输出结果^[79]。

4.3 复杂度对比

本节的目的并不是对以上各类方法的计算复杂性进行系统的、严格的分析, 而在于通过简单的对比以发现存在的问题, 以及在实际应用中需要对方法作选择时提供有用的参考。由于问题规模对算法复杂度的影响主要体现在环形闭合假设产生阶段, 在此只对该阶段的若干方法进行比较。

在时间上, 基于特征表示的方法通常需要逐帧进行比较, 其时间复杂度与数据帧呈线性关系。考虑到基于局部特征的方法需要建立特征间的关联, 其时间复杂度为 $O(M^2n)$, 其中 n 为帧数, M 为特征数(假定每帧图像特征数相等)。将特征离散化, 并采用基于词袋的表示后, 计算性能主要由聚类过程(词化)决定, 时间复杂度为 $O(MN)$, 其中 N 为类别数, M 为特征数。可见, 在基于词袋的表示中, 在词典规模固定后, 计算的复杂性与帧数 n 无关。当环境的规模较小(帧数在数千左右)时, 基于特征的表示具有较高的计算效率, 但对于更大的环境规模, 基于词袋的表示在效率上更优^[58]。

在空间上, 基于度量地图的场景表示不需要额外的存储空间, 其它方法需要保存相应的特征或词袋以及字典。因而, 基于特征表示的空间复杂度为 $O(n)$, 基于词袋的表示为 $O(n+N)$ 。

可见, 为了达到最好的时间、空间效率, 帧的选取以及词数的选定是关键。

5 图优化 (Graph optimization)

通过帧间配准 (详见第 3 节) 以及环形闭合检测 (详见第 4 节) 完成图的构建后, SLAM 问题可以转化为图优化问题, 即根据已建立的约束关系求解最优的位姿序列. 下面, 分问题建模、优化方法以及面临的问题 3 方面作阐述.

5.1 问题建模

在基于图优化的 SLAM 中, 机器人的位姿被表示为图中的节点, 观测信息在经过处理后转变为机器人位姿间的约束关系, 并通过连接节点间的边来表示. 机器人的绝对或相对位姿可以通过 $SE(2)$ 或 $SE(3)$ 中的 2D 或 3D 变换来表示. 以 6 自由度移动机器人在 3 维环境中的地图构建为例, 图的每个节点是一个 6 元组或 7 元组 (取决于位姿的描述, 其唯一确定机器人姿态 $P_i = \langle x_i, y_i, z_i, \phi_i, \theta_i, \psi_i \rangle$), 边则描述了相连的两个节点 P_i 及 P_j 间相对变换的概率分布, 通常由一个均值向量 \bar{T}_{ij} 及对应的信息矩阵 Σ_{ij}^{-1} (协方差矩阵的逆) 来表示 (见图 3 左图). 基于图优化的形式化方法可以从两个视角来看, 一个对应于不确定性观测的最大似然估计, 另外一个则是物理系统的最小能量状态. 两者最终都可以转化为非线性最小二乘问题.

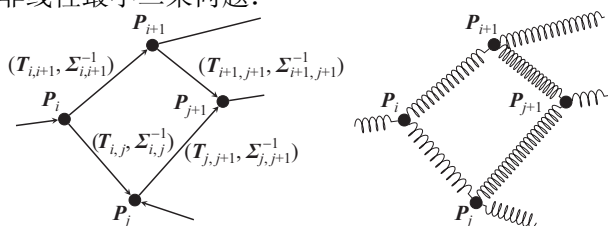


图 3 基于图优化的 SLAM 模型

Fig.3 Graph-based SLAM model

5.1.1 最大似然估计模型

位姿间的约束可以看作是“虚拟的观测”, 我们可以根据这些观测对系统内部状态 (即位姿序列) 作最大似然估计. 用 $P = \{P_i\}$ 表示机器人位姿集合, $T = \{T_{ij}\}$ 表示位姿间的相对变换集, 则在给定观测信息 T 的条件下, 对 P 作最大似然估计可描述为

$$L(P|T) = Pr(T|P) \quad (4)$$

$$= \prod_m Pr(T_m|P) \quad // \text{独立性假设} \quad (5)$$

$$= \prod_m Pr(T_m|P_{R(m)}) // \text{马尔可夫性} \quad (6)$$

$$= \prod_{(i,j) \in E} Pr(T_{ij}|P_i, P_j) // \text{符号替换} \quad (7)$$

其中 E 表示图中所有边的集合. 式 (5) 使用了条件独立假设, 即假定给定机器人姿态信息的条件下各观测之间相互独立. 式 (6) 应用了马尔可夫性^c, $P_{R(m)}$ 构成了 T_m 的马尔可夫邻域. 由于每个约束 T_{ij} (即每条边) 只与两个位姿节点 P_i 和 P_j 相关, 作简单的符号替换后, 可以得到式 (7).

若假定给定位姿 P_i 及 P_j 的条件下, T_{ij} 服从均值为 $\bar{T}_{ij} = P_j \ominus P_i$ ^d、方差为 Σ_{ij} 的高斯分布, 即 $T_{ij} \sim \mathcal{N}(P_j \ominus P_i, \Sigma_{ij})$, 则不难推导, 对 P 作最大似然估计与以下非线性最小二乘问题等价:

$$P_{ML} = \underset{P}{\operatorname{argmin}} \sum_{(i,j) \in E} [(P_j \ominus P_i) \ominus T_{ij}]^t \Sigma_{ij}^{-1} [(P_j \ominus P_i) \ominus T_{ij}] \quad (8)$$

$$= \underset{P}{\operatorname{argmin}} \sum_{(i,j) \in E} \|(P_j \ominus P_i) \ominus T_{ij}\|_{\Sigma_{ij}^{-1}}^2 \quad (9)$$

5.1.2 弹簧-质点物理模型

另一种较为直观的表示是, 将机器人的位姿看作是带质量的节点 (质点), 而约束则看作是连接这些节点的弹簧^[80], 参见图 3 右图. 由于每个约束都是根据与之相关的观测独立求解的, 它们之间不存在不一致性, 此时弹簧处在受力形变状态. 这样的物理系统通常并不稳定. 当弹簧对质点的作用力使系统重新达到平衡时, 系统处在能量最小状态, 此时质点的分布即代表最优的位姿序列. 弹簧的系数通过观测的不确定性来表示. 观测的不确定性越小 (协方差越小), 弹簧的强度越大, 使其形变需要的外部作用力也就越大 (不满足该约束会产生较大的代价). 相反, 观测的不确定性越大 (协方差越大), 弹簧的强度就越小, 使其形变需要的外部作用力也就越小 (不满足该约束的代价相对较小). 系统的能量最小状态对应式 (9) 所表示的非线性最小二乘问题的最优解.

5.2 优化方法

第 1 节曾探讨过基于图优化与基于滤波的 SLAM 方法的不同, 基于滤波的方法只考虑机器人当前的位姿, 但基于图优化的方法会对整个位姿序列进行估计. 因而基于图优化的 SLAM 方法需要处理更高维的数据. 假定环境中有 m 个特征 (每个特征用 x, y, z 三个变量表示), 当前机器人轨迹中包含 n 个姿态信息 (每个位姿用 $x, y, z, \phi, \theta, \psi$ 六个变量表示), 则滤波方法要处理问题的规模是 $6 + 3m$, 但平滑方法要处理的规模是 $6n + 3m$. 即使使用第 3 节中的帧间配准方法将观测信息全部转为位姿间约

^c即在给定马尔可夫邻域 (Markov Blanket) 的条件下, 随机变量的取值与其它非邻域随机变量的取值无关.

^d注, 符号 \ominus 表示位姿复合的逆操作, 其含义与文 [28] 中定义相同.

束, 问题的规模仍然为 $6n$. 随着需要考虑位姿数目 n 的增长, 问题的规模会变得越来越^e. 我们需要高效的方法来应对如此大规模的优化问题.

5.2.1 基于松弛的优化方法

Duckett 等提出采用 Gauss-Seidel 松弛方法来求解 SLAM 中的优化问题. 其基本思想是“依次选取每一个节点, 并将其移动到‘它的邻节点认为它该处在的位置’”^[81], 换句话说, 即根据其相邻节点的位置及它们之间的约束关系重新计算并更新节点的位置信息, 并且每次迭代都遍历所有节点. 在假定方位角已知(如通过电子罗盘测量)的情况下, Duckett 等证明了其必收敛于最优解. 该方法可用于增量式的 SLAM 中, 它可以在每次有新的观测到来时直接在上一轮的结果上进行更新. 但存在的缺陷是, 当某条边的误差较大时, 需要多次迭代才能将误差分配到其它边中, 而这正是出现环形闭合时所需要应对的情况. Frese 等^[9]提出多层次松弛的优化策略, 并利用多重网格方法来求解偏微分方程, 从而大大地提高了出现环形闭合时节点的优化效率.

5.2.2 基于随机梯度下降的方法

Olson 等人^[10]将随机梯度下降方法(stochastic gradient descent, SGD)应用到 SLAM 的优化中, 每次迭代随机选取图中的一条边作为当前约束并计算相应的梯度下降方向, 然后在该方向上对目标函数寻优. 随机梯度下降方法具有不容易陷入局部极值的优点, 对初始值具有较高的鲁棒性. 实验证明, 即使初始值与最优值相差较远, 甚至在不刻意初始化的情况下(如全零初始值或随机初始值), 随机梯度下降方法也能取得较好的收敛结果^[10-11,82]. Grisetti 等^[11]对 Olson 等所提的方法进行了改进和拓展, 采用树型结构来描述位姿间的关系并通过增量方式表示待求解的状态, 从而能更有效地对位姿进行更新. 此外, Grisetti 等还将树型表示以及随机梯度方法应用到 6DOF 位姿优化中, 并验证了方法的有效性^[17].

5.2.3 非线性最小二乘 + 稀疏结构

SLAM 中的图优化问题可以看作是一个非线性最小二乘问题(见 5.1 节), 而求解该类问题的一般方法是: 对目标函数在当前解处进行线性化(如通过 1 阶泰勒展开对非线性函数作线性近似, 从而得到线性最小二乘问题), 求解线性系统(线性最小二乘问题具有凸特性, 对函数求导并令导数等

于零即可构造需求解的线性系统, 所得的解是近似问题的最优解), 然后依此进行迭代直到问题收敛或达到最大迭代步数. 这便是 Gauss-Newton 以及 LM (Levenberg-Marquardt) 方法的主要思想. LM 与 Gauss-Newton 方法的不同点在于, LM 通过引入阻尼因子及回溯性, 从而能有效控制算法的收敛^f.

但是, 如果全然不考虑 SLAM 问题的结构特性并假定图节点数为 n , 则 LM 算法的时间复杂度为 $O(n^3)$ ^[9], 在实际问题求解中将远不能满足实时性要求^[10]. SLAM 问题具有非常好的稀疏性质, 可以被充分利用进而提高问题的求解效率.

在基于滤波的方法中, 协方差矩阵及信息矩阵都会随着位姿节点变量的边缘化而逐渐变得密集. 平滑方法保留位姿序列, 虽然需要考虑更多的变量, 但却保持了信息矩阵及观测函数雅可比矩阵的稀疏性^[83]. Dellaert 与 Kaess^[84]充分利用平滑 SLAM 问题中的稀疏结构特性, 通过稀疏矩阵分解(如稀疏乔列斯基分解等)对线性系统进行求解. Kaess 等^[12]对平滑信息矩阵作 QR 分解, 并选择性对其进行增量式更新, 从而避免每次重新计算平滑信息矩阵, 这大大地提高了更新的效率. Konolige 等^[82]提出一种根据给定图约束快速构造稀疏矩阵的方法. Kummerle 等^[14]则提供了基于稀疏矩阵分解的开源解决方案.

5.2.4 流形优化

以上的方法均假定优化是在欧氏空间中进行的, 但机器人的位姿(包括平移与旋转两部分)却处在非欧氏空间中. 为了避免奇异值问题, 可以采用冗余的表示方式, 如将位姿中的旋转分量用旋转矩阵或四元数来表示. 但如此, 在优化过程中会打破变量间的约束(如旋转矩阵或单位四元数的相关性), 产生额外的自由度, 从而引入不必要的误差. Grisetti 等^[85]提出在流形空间中进行优化的思想. 这样不仅克服了奇异值问题, 也被实验验证可以取得更高的精度. 最近, 研究者们提供了能用于流形优化的开源工具^[14,86], 可以大为提高开发效率.

5.3 面临的问题

5.3.1 效率问题

在基于图优化的 SLAM 方法刚提出时, Lu 等采用的是基于非线性最小二乘的求解方法. 但由于其没有考虑 SLAM 问题中的稀疏结构, 甚至直接采用矩阵求逆方法对问题进行求解, 因此求解效率非常低下. 基于松弛与随机梯度下降方法的提出, 在

^e对大小固定的环境, m 通常是固定的, 但 n 往往随轨迹的长度呈线性增长.

^f对高度非线性情况, Gauss-Newton 方法可能无法收敛.

一定程度上提高了求解效率,但其并未充分利用非线性最小二乘问题的优点^[87],如随机梯度下降方法只利用函数的 1 阶特性,在接近最优解时收敛非常缓慢.在迭代次数受限的情况下,其最终结果的精度受到影响.最近发表的基于非线性最小二乘的求解方法^[12,14,82],不但充分利用了 SLAM 中的稀疏结构,也吸取了稀疏线性代数方面所取得的重要研究成果.这使得问题的求解效率大大提高,能处理问题的规模也大为提升^[82],代表当前该领域的高水准.

基于非线性最小二乘方法的效率主要受所求解线性系统复杂度的影响,而求解线性系统的效率与其稀疏程度有关.对采用稀疏矩阵分解的方法,其最佳复杂度为 $O(n)$,最坏情况下为 $O(n^3)$.

5.3.2 鲁棒性问题

求解方法的鲁棒性主要从两方面来考虑,一是对初始值的依赖,二是对错误闭环信息的适应能力.

(1) 对初始值的鲁棒性

由于里程计(包括轮子里程计及其它依靠观测信息作自运动估计的方法)信息可能会存在较大的累积误差,由此得到的位姿序列的初始值与真实值之间存在较大的偏差.因而减少方法对初始值的依赖,增强方法的收敛域至关重要.解决思路有两种:一是提高优化方法本身的全局搜索能力,二是通过其它方法快速获取一个可以接受的初始值.上一小节提到的方法中,随机梯度下降方法对初始值的鲁棒性较好,而非线性最小二乘方法则容易陷入局部最优^[82].Carlone 等^[88]提出对基于图优化的 SLAM 作线性近似并给出解析求解的方法,该方法对初始值不存在依赖,可以利用其结果作为非线性最小二乘方法的初始值.但目前该方法只适用于 2D SLAM.

(2) 对错误闭环信息的鲁棒性

传统的图优化方法通常以图具有正确的拓扑结构为前提^[10-11].若图中引入不恰当的环形闭合信息(false positive loop closure),将可能导致其产生错误的收敛结果.这是因为最小二乘优化方法本身对局外点(outlier)的鲁棒性不好.为了尽可能保证收敛时所得到的地图是正确的,可以对环形闭合施加严格的约束,使检测的错误率足够低,并通过核函数方法^[14]降低其影响.但这样做的效果仍然有限.与之前认为错误的环形闭合信息仅仅是前端的问题不同,Sünderhauf 等允许在图优化的过程中改变图的拓扑结构以剔除错误的环形闭合^[15,89],从而

大大地提高了方法的鲁棒性.

5.3.3 扩展性问题

基于图优化的 SLAM 方法以机器人的位姿为节点.通常情况下,机器人走过的轨迹越长,需处理的位姿节点也就越多,这不利于方法的扩展.显然,当机器人在大小固定的环境中行走时,图的节点数目应该跟环境的大小相关,而不应与运动轨迹的长度相关.可见,要使 SLAM 方法具备良好的扩展性,关键是对图节点进行有效的控制.

减少图中节点最为直观的方法是对节点间的距离进行限制,即只有节点间的距离超过一定的阈值时才添加到图中^[13].Kretschmar 等^[21]从观测所含的信息出发,评估观测帧的信息增益,并依此对图进行剪枝,以控制节点数目.这种方法在保持节点规模的同时具有最小的信息损失,因而也最大程度上保证了地图信息的完整性.对节点进行剪枝实际对应节点的边缘化过程,这可能导致图的结构变得密集.文[90]提出采用 Chow-Liu 树对节点间的关系作近似描述,以保证节点连接的稀疏性.

6 发展趋势(Development trend)

基于图优化的 SLAM 方法的研究已取得了很大的进展,尤其体现在基于视觉的环形闭合检测以及图优化方法上.纵观近些年该领域的文献资料,基于图优化的 SLAM 技术有以下发展趋势:

(1) 深化 SLAM 问题非线性结构的研究:非线性与稀疏结构是 SLAM 问题的两大特性,对问题的高效求解具有重要的指导意义.到目前为止,对 SLAM 的稀疏结构性质已经有了较为深入的了解^[84,91].相比之下,对非线性结构的研究则显得欠缺.Huang 等人^[92-93]对 SLAM 的非线性结构作了初步的理论分析,其研究显示通过选择合适的状态表示,SLAM 可描述为具有众多二次项的非线性优化问题.

(2) 构建鲁棒性好的后端:传统的图优化研究以正确的数据关联为前提,即作为输入的图必须具有正确的拓扑结构,否则优化可能全然失效.这使得这些方法具有很大的局限性.Sünderhauf^[15,94]等提出允许在优化过程中对图的拓扑结构作修改,以发现并剔除错误的环形闭合信息.Olson 等^[95]也采用了类似的思想,并通过混合模型来描述和推断错误的环形闭合信息.实验证明,鲁棒性好的后端在传统方法失效的情况下仍能很好地对地图进行恢复.

(3) 终生地图创建:传统的研究多趋向于将地图学习与应用分开,如先通过 SLAM 学习未知环境地图,然后利用学习好的地图进行自定位或作运动规

划等. 终生地图创建考虑直接将机器人长期运行于未知环境当中, 其需要应对环境的变化, 并持续地对地图进行更新. 对短期的地图创建, 基于图优化的 SLAM 方法已被证实非常有效. 一些学者对终生地图创建环境下, 基于图优化方法的扩展性问题进行了研究^[21,96-97].

(4) 语义地图创建 (semantic mapping): 传统的地图描述缺乏语义信息, 这对其应用造成很大的限制和制约. 为了使机器人能够更好地为人类提供各类服务以及与人类进行自然的交互, 机器人需要具备理解场景以及辨识物体的能力, 而构造含有语义信息的地图则成为解决该问题的一种重要途径. Koppula 等^[98]利用 RGB-D 传感器对室内环境进行 3 维重构, 并通过点云标注为创建的地图添加语义信息. Bao 等^[99]则提出了新型的从运动中恢复结构 (structure from motion, SFM) 的方法, 该方法同时考虑了几何与语义约束, 因而能恢复具有语义信息的地图. 总体而言, 当前对语义地图创建的研究还相对较少, 鉴于图优化 SLAM 方法在传统地图创建中已取得的效果, 通过其来构造具有语义信息的地图将是 SLAM 发展的一个可能趋势.

参考文献 (References)

- [1] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: Part I. The essential algorithms[J]. IEEE Robotics and Automation Magazine, 2006, 13(2): 99-108.
- [2] Smith R C, Cheeseman P. On the representation and estimation of spatial uncertainty[J]. International Journal of Robotics Research, 1986, 5(4): 56-68.
- [3] Thrun S, Liu Y F, Koller D, et al. Simultaneous localization and mapping with sparse extended information filters[J]. International Journal of Robotics Research, 2004, 23(7/8): 693-716.
- [4] Montemerlo M, Thrun S, Koller D, et al. FastSLAM: A factored solution to the simultaneous localization and mapping problem[C]//Proceedings of the National Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2002: 593-598.
- [5] Thrun S. Robotic mapping: A survey[M]// Exploring Artificial Intelligence in the New Millennium. San Francisco, USA: Morgan Kaufmann, 2002: 1-35.
- [6] Huang S D, Dissanayake G. Convergence and consistency analysis for extended Kalman filter based SLAM[J]. IEEE Transactions on Robotics, 2007, 23(5): 1036-1049.
- [7] Thrun S, Burgard W, Fox D. Probabilistic robotics[M]. Cambridge, USA: MIT Press, 2005.
- [8] Thrun S, Montemerlo M. The graph SLAM algorithm with applications to large-scale mapping of urban structures[J]. International Journal of Robotics Research, 2006, 25(5/6): 403-429.
- [9] Frese U, Larsson P, Duckett T. A multilevel relaxation algorithm for simultaneous localization and mapping[J]. IEEE Transactions on Robotics, 2005, 21(2): 196-207.
- [10] Olson E, Leonard J, Teller S. Fast iterative alignment of pose graphs with poor initial estimates[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2006: 2262-2269.
- [11] Grisetti G, Stachniss C, Grzonka S, et al. A tree parameterization for efficiently computing maximum likelihood maps using gradient descent[M]//Robotics: Science and Systems III. Cambridge, USA: MIT Press, 2008: 65-72.
- [12] Kaess M, Ranganathan A, Dellaert F. iSAM: Incremental smoothing and mapping[J]. IEEE Transactions on Robotics, 2008, 24(6): 1365-1378.
- [13] Konolige K, Agrawal M. FrameSLAM: From bundle adjustment to real-time visual mapping[J]. IEEE Transactions on Robotics, 2008, 24(5): 1066-1077.
- [14] Kummerle R, Grisetti G, Strasdat H, et al. g²o: A general framework for graph optimization[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2011: 3607-3613.
- [15] Sünderhauf N, Protzel P. Towards a robust back-end for pose graph SLAM[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2012: 1254-1261.
- [16] Johannsson H, Kaess M, Fallon M, et al. Temporally scalable visual SLAM using a reduced pose graph[R]. Cambridge, USA: Computer Science and Artificial Intelligent Laboratory, MIT, 2012.
- [17] Grisetti G, Grzonka S, Stachniss C, et al. Efficient estimation of accurate maximum likelihood maps in 3D[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2007: 3472-3478.
- [18] Borrmann D, Elseberg J, Lingemann K, et al. Globally consistent 3D mapping with scan matching[J]. Robotics and Autonomous Systems, 2008, 56(2): 130-142.
- [19] Wurm K M, Hornung A, Bennewitz M, et al. OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems[C]//ICRA 2010 Workshop: Best Practice in 3D Perception and Modeling for Mobile Manipulation. 2010.
- [20] McDonald J, Kaess M, Cadena C, et al. Real-time 6-DOF multi-session visual SLAM over large-scale environments[J]. Robotics and Autonomous Systems, 2012.
- [21] Kretzschmar H, Grisetti G, Stachniss C. Lifelong map learning for graph-based SLAM in static environments[J]. Künstliche Intelligenz, 2010, 24(3): 199-206.
- [22] Cruz L, Lucio D, Velho L. Kinect and RGBD images: Challenges and applications[C]//SIBGRAPI - Conference on Graphics, Patterns and Images Tutorials. Washington, USA: IEEE Computer Society, 2012: 36-49.
- [23] Henry P, Krainin M, Herbst E, et al. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments[J]. International Journal of Robotics Research, 2012, 31(5): 647-663.
- [24] Huang A S, Bachrach A, Henry P, et al. Visual odometry and mapping for autonomous flight using an RGB-D camera[C]//International Symposium on Robotics Research. 2011.
- [25] Endres F, Hess J, Engelhard N, et al. An evaluation of the RGB-D SLAM system[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2012: 1691-1696.
- [26] Izadi S, Kim D, Hilliges O, et al. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera[C]//24th Annual ACM Symposium on User Interface Software and Technology. New York, USA: ACM, 2011: 559-568.

- [27] Whelan T, Kaess M, Fallon M, et al. *Kintinuous: Spatially extended KinectFusion*[M]//Robotics: Science and Systems VII. Cambridge, USA: MIT Press, 2012.
- [28] Lu F, Milios E. Globally consistent range scan alignment for environment mapping[J]. *Autonomous robots*, 1997, 4(4): 333-349.
- [29] Gutmann J S, Konolige K. Incremental mapping of large cyclic environments[C]//IEEE International Symposium on Computational Intelligence in Robotics and Automation. Piscataway, USA: IEEE, 1999: 318-325.
- [30] Scaramuzza D, Fraundorfer F. Visual odometry. Part I: The first 30 years and fundamentals[J]. *IEEE Robotics and Automation Magazine*, 2011, 18(4): 80-92.
- [31] Nister D, Naroditsky O, Bergen J. Visual odometry for ground vehicle applications[J]. *Journal of Field Robotics*, 2006, 23(1): 3-20.
- [32] Konolige K, Agrawal M, Sola J. Large-scale visual odometry for rough terrain[C]//13th International Symposium on Robotics Research. Berlin, Germany: Springer, 2011: 201-212.
- [33] Saez J M, Hogue A, Escolano F, et al. Underwater 3D SLAM through entropy minimization[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2006: 3562-3567.
- [34] Steder B, Grisetti G, Stachniss C, et al. Visual SLAM for flying vehicles[J]. *IEEE Transactions on Robotics*, 2008, 24(5): 1088-1093.
- [35] Maimone M, Cheng Y, Matthies L. Two years of visual odometry on the mars exploration rovers[J]. *Journal of Field Robotics*, 2007, 24(3): 169-186.
- [36] Besl P J, McKay N D. A method for registration of 3-D shapes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(2): 239-256.
- [37] Yang C, Medioni G. Object modeling by registration of multiple range images[J]. *Image and Vision Computing*, 1992, 10(3): 145-155.
- [38] Zhang Z Y. Iterative point matching for registration of free-form curves and surfaces[J]. *International Journal of Computer Vision*, 1994, 13(2): 119-152.
- [39] Arun K S, Huang T S, Blostein S D. Least-squares fitting of two 3-D point sets[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, 9(5): 698-700.
- [40] Eggert D W, Lorusso A, Fischer R B. Estimating 3-D rigid body transformations: A comparison of four major algorithms[J]. *Machine Vision and Applications*, 1997, 9(5/6): 272-290.
- [41] Segal A, Haehnel D, Thrun S. *Generalized-ICP*[M]//Robotics: Science and Systems V. Cambridge, USA: MIT Press, 2009.
- [42] Salvi J, Matabosch C, Fofi D, et al. A review of recent range image registration methods with accuracy evaluation[J]. *Image and Vision Computing*, 2007, 25(5): 578-596.
- [43] Turk G, Levoy M. Zippered polygon meshes from range images[C]//21st Annual Conference on Computer Graphics and Interactive Techniques. New York, USA: ACM, 1994: 311-318.
- [44] Godin G, Rioux M, Baribeau R. Three-dimensional registration using range and intensity information[C]//Proceedings of the SPIE, vol.2350. Bellingham, USA: SPIE, 1994: 279-290.
- [45] Fitzgibbon A W. Robust registration of 2D and 3D point sets[J]. *Image and Vision Computing*, 2003, 21(13/14): 1145-1153.
- [46] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm[C]//Third International Conference on 3-D Digital Imaging and Modeling. Los Alamitos, USA: IEEE Computer Society, 2001: 145-152.
- [47] Nister D, Naroditsky O, Bergen J. Visual odometry[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2004: 652-659.
- [48] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [49] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [50] Agrawal M, Konolige K, Blas M R. CenSurE: Center surround extremas for realtime feature detection and matching[C]//European Conference on Computer Vision. Berlin, Germany: Springer, 2008: 102-115.
- [51] Shi J B, Tomasi C. Good features to track[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 1994: 593-600.
- [52] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [53] Nister D. An efficient solution to the five-point relative pose problem[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2003: 195-202.
- [54] Moreno-Noguer F, Lepetit V, Fua P. Accurate non-iterative O(n) solution to the PnP problem[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2007: 2168-2175.
- [55] Kneip L, Scaramuzza D, Siegwart R. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2011: 2969-2976.
- [56] Newman P, Ho K. SLAM – Loop closing with visually salient features[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2005: 635-642.
- [57] Cummins M, Newman P. Probabilistic appearance based navigation and loop closing[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2007: 2042-2048.
- [58] Zhang H. BoRF: Loop-closure detection with scale invariant visual features[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2011: 3125-3130.
- [59] Cummins M, Newman P. Appearance-only SLAM at large scale with FAB-MAP 2.0[J]. *International Journal of Robotics Research*, 2011, 30(9): 1100-1123.
- [60] Sünderhauf N, Protzel P. Brief-gist-closing the loop by simple means[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2011: 1234-1241.
- [61] Liu Y, Zhang H. Visual loop closure detection with a compact image descriptor[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2012: 1051-1056.

- [62] Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. *Image and Vision Computing*, 2004, 22(10): 761-767.
- [63] Steder B, Grisetti G, Burgard W. Robust place recognition for 3D range data based on point features[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2010: 1400-1405.
- [64] Hartigan J A, Wong M A. A k-means clustering algorithm[J]. *Applied Statistics*, 1979, 28(1): 100-108.
- [65] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos[C]//IEEE International Conference on Computer Vision. Los Alamitos, USA: IEEE Computer Society, 2003: 1470-1477.
- [66] Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval [M]. New York, USA: ACM Press, 1999.
- [67] Nister D, Stewenius H. Scalable recognition with a vocabulary tree[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2006: 2161-2168.
- [68] Schindler G, Brown M, Szeliski R. City-scale location recognition[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2007: 1-7.
- [69] Angeli A, Filliat D, Doncieux S, et al. Fast and incremental method for loop-closure detection using bags of visual words[J]. *IEEE Transactions on Robotics*, 2008, 24(5): 1027-1037.
- [70] Williams B, Cummins M, Neira J, et al. An image-to-map loop closing method for monocular SLAM[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2008: 2053-2059.
- [71] Clemente L A, Davison A J, Reid I, et al. Mapping large loops with a single hand-held camera[C]//Robotics: Science and Systems. Cambridge, USA: MIT Press, 2007: 297-304.
- [72] Bosse M, Newman P, Leonard J, et al. An Atlas framework for scalable mapping[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2003: 1899-1906.
- [73] Olson E B. Robust and efficient robotic mapping[D]. Cambridge, USA: Department of Electrical Engineering and Computer Science, MIT, 2008.
- [74] Ho K L, Newman P. Detecting loop closure with scene sequences[J]. *International Journal of Computer Vision*, 2007, 74(3): 261-286.
- [75] Konolige K, Bowman J, Chen J D, et al. View-based maps[J]. *International Journal of Robotics Research*, 2010, 29(8): 941-957.
- [76] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge, UK: Cambridge University Press, 2000.
- [77] Galvez-Lopez D, Tardos J D. Real-time loop detection with bags of binary words[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2011: 51-58.
- [78] Cadena C, Gálvez-López D, Ramos F, et al. Robust place recognition with stereo cameras[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2010: 5182-5189.
- [79] Olson E. Recognizing places using spectrally clustered local matches[J]. *Robotics and Autonomous Systems*, 2009, 57(12): 1157-1172.
- [80] Golfarelli M, Maio D, Rizzi S. Elastic correction of dead-reckoning errors in map building[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 1998: 905-911.
- [81] Duckett T, Marsland S, Shapiro J. Learning globally consistent maps by relaxation[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2000: 3841-3846.
- [82] Konolige K, Grisetti G, Kummerle R, et al. Efficient sparse pose adjustment for 2D mapping[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2010: 22-29.
- [83] Strasdat H, Montiel J M M, Davison A J. Real-time monocular SLAM: Why filter?[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2010: 2657-2664.
- [84] Dellaert F, Kaess M. Square root SAM: Simultaneous localization and mapping via square root information smoothing[J]. *International Journal of Robotics Research*, 2006, 25(12): 1181-1203.
- [85] Grisetti G, Kummerle R, Stachniss C, et al. Hierarchical optimization on manifolds for online 2D and 3D mapping[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2010: 273-278.
- [86] Wagner R, Birbach O, Frese U. Rapid development of manifold-based graph optimization systems for multi-sensor calibration and SLAM[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2011: 3305-3312.
- [87] Nocedal J, Wright S J. Numerical optimization[M]. Berlin, Germany: Springer Verlag, 1999.
- [88] Carlone L, Aragues R, Castellanos J A, et al. A linear approximation for graph-based simultaneous localization and mapping[J]. *Robotics: Science and Systems VII*. Cambridge, USA: MIT Press, 2012: 41-48.
- [89] Sünderhauf N. Robust optimization for simultaneous localization and mapping[D]. Chemnitz, Germany: Department of Electrical Engineering and Information, Chemnitz University of Technology, 2012.
- [90] Kretschmar H, Stachniss C, Grisetti G. Efficient information-theoretic graph pruning for graph-based SLAM with laser range finders[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2011: 865-871.
- [91] Konolige K, Garage W. Sparse sparse bundle adjustment[C]//Proceedings of the British Machine Vision Conference. 2010.
- [92] Huang S D, Lai Y W, Frese U, et al. How far is SLAM from a linear least squares problem?[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2010: 3011-3016.
- [93] Huang S D, Wang H, Frese U, et al. On the number of local minima to the point feature based SLAM problem[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2012: 2074-2079.
- [94] Sünderhauf N, Protzel P. Switchable constraints for robust pose graph SLAM[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: 2012: 1879-1884.
- [95] Olson E, Agarwal P. Inference on networks of mixtures for robust robot mapping[M]//Robotics: Science and Systems 12. Cambridge, USA: MIT Press, 2012.

- [96] Konolige K, Bowman J. Towards lifelong visual maps[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2009: 1156-1163.
- [97] Walcott-Bryant A, Kaess M, Johannsson H, et al. Dynamic pose graph SLAM: Long-term mapping in low dynamic environments[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2012: 1871-1878.
- [98] Koppula H S, Anand A, Joachims T, et al. Semantic labeling of 3D point clouds for indoor scenes[C]//25th Annual Conference on Neural Information Processing Systems. New York, USA: Curran Associates Inc., 2011.
- [99] Bao S Y, Savarese S. Semantic structure from motion[C]//IEEE

Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2011: 2025-2032.

作者简介:

梁明杰 (1986-), 男, 博士生. 研究领域: 环境建模, 自主导航, 概率推理.

闵华清 (1956-), 男, 教授, 博士生导师. 研究领域: 智能机器人, 智能软件, 自动系统.

罗荣华 (1975-), 男, 副教授, 硕士生导师. 研究领域: 机器视觉, 场景理解, 概率模型.

(上接第 499 页)

- [13] 朱博, 戴先中, 李新德. 一种基于物联网技术的场所感知系统初探[J]. 计算机科学, 2012, 39(2): 216-221.
- Zhu B, Dai X Z, Li X D. Preliminary study on a kind of place perception system based on IOT technology[J]. Computer Science, 2012, 39(2): 216-221.
- [14] Morel J-M, Yu G S. ASIFT: A new framework for fully affine invariant image comparison[J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 438-469.
- [15] MIT Media Lab. Lifelong Kindergarten[OL]. [2012-08-25]. <http://llk.media.mit.edu/projects.php?id=135>.
- [16] Microsoft Robotics. Microsoft Robotics Developer Studio[OL]. [2012-08-25]. <http://www.microsoft.com/robotics/#About>.
- [17] Bodurov V. Using Microsoft Robotics Developer Studio with Lego Mindstorms NXT Robot[OL]. (2010-08-10) [2012-08-25]. <http://www.simple-talk.com/dotnet/.net-framework/using-microsoft-robotics-developer-studio-with-lego-mindstorms-nxt-robot/>.
- [18] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis

and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.

- [19] Zhu B, Dai X Z, Li X D, et al. Place concept teaching through sketch map for robot place perception based on prototype mechanism[C]//6th International Conference on Intelligent Systems and Knowledge Engineering: vol.124. Berlin, Germany: Springer-Verlag, 2011: 79-89.

作者简介:

朱 博 (1981-), 男, 博士生. 研究领域: 空间定性推理, 机器人交互与导航, 语义地图创建.

戴先中 (1954-), 男, 博士, 教授. 研究领域: 复杂控制理论, 机器人控制, 电力系统控制, 测量与信号处理.

李新德 (1975-), 男, 博士, 副教授. 研究领域: 智能机器人, 机器感知, 信息融合, 不确定推理和机器视觉.