

DOI: 10.11992/tis.201703006

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180131.0858.002.html>

视觉同时定位与地图创建综述

周彦, 李雅芳, 王冬丽, 裴廷睿
(湘潭大学 信息工程学院, 湖南 湘潭 411105)

摘要: 同时定位与地图创建 (simultaneous localization and mapping, SLAM) 自 1986 年提出以来一直是机器人领域的热点问题, 被认为是实现真正全自主移动机器人的关键。其目的是让机器人在未知环境下实现自身定位同时创建出环境地图。视觉 SLAM (visual simultaneous localization and mapping, VSLAM) 是仅用相机作为传感器的定位与制图。随着计算机视觉和机器人技术的发展, VSLAM 已成为无人系统领域的研究焦点。本文对 VSLAM 的最新研究现状进行总结, 阐述了 VSLAM 中的主要问题, 分别介绍了 VSLAM 基于滤波和图优化的实现方法, 并探讨了 VSLAM 的研究与发展方向。

关键词: 计算机视觉; 同时定位与地图创建; VSLAM; 机器人; 滤波; 图优化; 综述; 深度学习

中图分类号: TP24 **文献标志码:** A **文章编号:** 1673-4785(2018)01-0097-10

中文引用格式: 周彦, 李雅芳, 王冬丽, 等. 视觉同时定位与地图创建综述[J]. 智能系统学报, 2018, 13(1): 97-106.

英文引用格式: ZHOU Yan, LI Yafang, WANG Dongli, et al. A survey of VSLAM[J]. CAAI transactions on intelligent systems, 2018, 13(1): 97-106.

A survey of VSLAM

ZHOU Yan, LI Yafang, WANG Dongli, PEI Tingrui

(College of Information Engineering, Xiangtan University, Xiangtan 411105, China)

Abstract: Simultaneous localization and mapping (SLAM), an essential task for an autonomy robot, has been a hot topic in the field of robotics since the concept first proposed in 1986. The purpose is to make a robot locate itself in an unknown environment while simultaneously construct a map of the environment. Visual SLAM (VSLAM) refers to that one using a camera or cameras as the sole sensor. With the development of computer vision and robotics, VSLAM has become the focus in the field of unmanned systems. In this paper, we survey the recent progress of VSLAM. After identifying the main problems in the development of VSLAM, we introduce the VSLAM methods based on both filter and graph optimizations. Finally, the further study and development directions of VSLAM are given.

Keywords: computer vision; simultaneous localization and mapping; VSLAM; robot; filter; graph optimization; survey; deep learning

移动机器人为实现自主导航, 面临着在哪里、到哪里、怎么去 3 个需要解决的关键问题。“在哪里”是机器人对自身的定位, 后两个问题即机器人需要解决的路径规划问题。对自主移动机器人来说, 定位是重中之重, 是路径规划的基石。在定位中, 机器人首当其冲的任务便是感知周围的环境, 并对之加以描述。移动机器人的定位和地图创建已成为

了机器人领域的热点研究问题^[1-2]。目前, 已经有了很多有效方法来解决已知环境中(有环境先验信息)机器人自主定位与已知机器人位置情况下的地图创建问题^[1]。然而在很多环境中, 机器人无法利用全局定位系统进行定位, 而且事先获取环境先验信息很困难, 甚至是不可能的^[3], 此情此景下, 机器人需要在没有环境先验信息的情况下, 在移动过程中一边计算自身位置, 一边构建环境地图, 于是移动机器人的同时定位与地图创建 (SLAM) 问题^[4-5]应运而生。

收稿日期: 2017-03-03. 网络出版日期: 2018-01-31.

基金项目: 国家自然科学基金项目 (61773330, 61372049, 61100140, 61104210); 湖南省自然科学基金项目 (2017JJ2253); 湖南省教育厅优秀青年基金项目 (17B259).

通信作者: 周彦. E-mail: yanzhou@xtu.edu.cn.

SLAM 也称为 CML (concurrent mapping and localization), 最先由 Smith Self 和 Cheeseman 于 1986 年提出^[5-6]。这一理论是实现真正全自主移动机器人的关键已经成为共识^[7-8]。SLAM 以传感器作为划分标准, 主要分为激光、视觉两大类。其中, 激光 SLAM 研究较早, 理论和工程均比较成熟, 视觉 SLAM 尚处于实验室研究阶段^[1]。SLAM 早期研究侧重于使用滤波理论来最小化运动物体的位姿和地图路标点的噪声。自 21 世纪以来, 学者们借鉴运动恢复结构 SfM(structure from motion) 中的方式^[9], 以优化理论为基础求解 SLAM 问题, 该方法通常以位姿图的形式描述机器人各时刻的状态, 又称为基于图优化的 SLAM, 在 VSLAM 领域中取得了主导地位^[10-11]。

1 VSLAM 存在的问题分析

1.1 特征点提取、描述与匹配

图像的特征一般可划分为点特征、直线特征以及边缘、轮廓特征, 其中线、边缘、轮廓等特征在高维空间进行处理, 计算量大; 点特征对遮挡相对鲁棒、提取速度快并且识别性好, 所以应用较多。局部特征点不仅能够保留图像重要特征信息, 而且也使得信息的数据量减少, 使计算速度和匹配速度都加快, 因此基于特征的 VSLAM 普遍采用点特征。图 1 标出了可作为图像特征的部分。

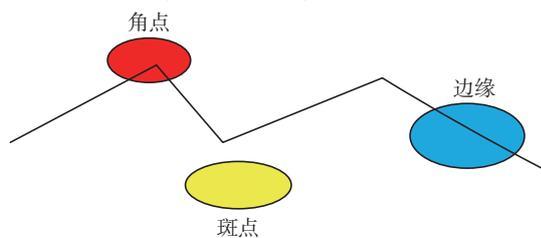


图 1 可作为图像特征的部分: 角点、边缘、斑点

Fig. 1 Parts that can be used as image features: corner, edge, blob

斑点和角点是局部特征点中比较流行的两种。斑点的重要特征是与周围区域有颜色和灰度上的差别。斑点检测方法应用最广泛的是利用高斯拉普拉斯算子检测的方法 (LOG), 以及利用像素点海森矩阵 (二阶微分) 及其行列式值的方法 (DOH)。角点描述的是两条边的交点, 其检测方法常用 Harris 角点检测算法和 FAST 角点检测算法。对特征点的描述有浮点型特征描述子和二进制字符串特征描述子。提取特征点后需对两幅图像进行特征匹配, 特征匹配采用计算描述子间距离的方法, 距离越小相

似性越高。对于浮点型描述子采用欧氏距离, 对于二进制字符型描述子使用汉明距离 (Hamming distance), 汉明距离指两个描述子 (二进制串) 不同位数的个数。当特征点数量非常大时, 快速近似最近邻 (FLANN) 算法^[12]能够满足 SLAM 的实时性需求。

常用的特征提取和匹配算法有 SIFT 算法、SURF 算法和 ORB 算法。SIFT^[13]算法中使用斑检测方法和浮点型特征描述子, 在建立高斯差分空间金字塔的基础上提取出具有尺度不变性的特征点, 然后对特征点邻域内的点的梯度方向进行直方图统计。特征点的主方向就是直方图中比重最大的方向, 必要时可选一个辅方向。SIFT 特征集旋转不变性、尺度不变性、对图像变形和光照鲁棒等优点于一身, 不足之处是计算量大, 计算速度慢, 需要在 GPU 加速的情况下才可满足 SLAM 的实时性需求。SURF^[14]算法是对 SIFT 算法的改进, 使用基于 DoH 的斑点特征检测方法; 在特征点的描述上, SURF 算法通过积分图, 利用两个方向上的 Harr 小波模板进行梯度计算, 然后对邻域内点的梯度方向以扇形的方式进行统计, 得到特征点的主方向。SURF 算法速度快且稳定性好, 应用也较为广泛。Ethan-Rublee 在 2011 年提出的 ORB^[15]算法使用改进的 FAST 特征点检测算法, ORB 的特征描述子采用改进后的二进制字符串特征描述子 BRIEF^[16]。由于采用速度极快的二进制描述子, ORB 使得整个图像特征提取的环节大大加速。

1.2 特征点深度获取

单目相机无法直接获取深度信息, 深度信息通过反深度法 (inverse depth)、三角测量 (三角化)、粒子滤波法等来获取。Civera 等^[17]提出了反深度法, 该方法旨在减少深度分布非高斯性的影响。反深度法为获得较好的线性效果, 在 EKF 系统里使用深度的倒数进行更新。三角测量最早由高斯提出, 是指通过在两个不同地点观察同一个点的夹角, 确定出该点的距离 (深度)。使用关键帧与稀疏捆集调整 (sparse bundle adjustment, SBA) 框架的 VSLAM 系统, 如文献^[18-21], 均采用了该方法。Davison 等^[22]采用的 Particle Filter 方法会在特征所在的深度方向上生成多个粒子, 通过粒子的匹配、更新来得到特征点深度的概率分布, 不足之处是容易增加系统的不一致性, 致使最后概率估计发散。

双目相机一般由左和右两个水平放置的相机组成, 通过同步采集到的左右相机的图像, 计算图像之间的视差, 来估计每一个像素的深度。

图 2 中, O_L 、 O_R 为左右相机的光圈中心, 黑色框为成像平面, f 为焦距, u_L 、 u_R 为成像的平面坐标, u_R 为负数。根据几何关系, 由相似三角形 $P-P_L-P_R$ 和 $P-O_L-O_R$, 得

$$\frac{z-f}{z} = \frac{b-u_L+u_R}{b}$$

整理得

$$z = \frac{fb}{d}, d = u_L - u_R$$

式中 d 为 P 在左眼相机图像和右眼相机图像中的横坐标之差, 叫做视差。根据视差就可以估计一个像素离相机的距离 z 。

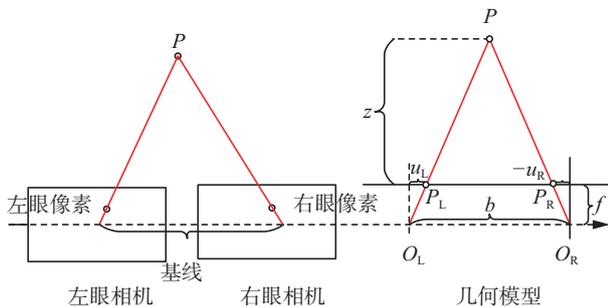


图 2 双目相机模型

Fig. 2 Binocular camera model

深度相机主动测量每个像素的深度直接获取深度信息^[23]。目前的 RGB-D 相机按原理可分为两大类, 即通过红外结构光(structured light)来测量像素距离和通过飞行时间法(time of flight, ToF)测量像素距离。在结构光原理中, 相机向探测目标发射一束光线(通常是红外光), 根据返回的结构光图案, 计算像素离自身的距离。在 ToF 中, 相机向目标发射脉冲光, 然后根据发送到返回之间的光束飞行时间, 确定物体离自身的距离。在测量深度之后, RGB-D 相机完成深度与彩色图像像素之间的配对, 输出一一对应的彩色图和深度图。图 3 是 RGB-D 相机的原理图。

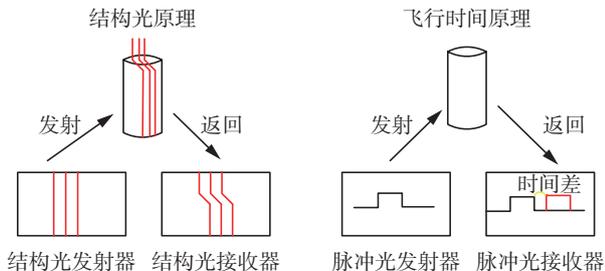


图 3 RGB-D 相机原理图

Fig. 3 Schematic of RGB-D camera

1.3 数据关联的优化问题

SLAM 中数据关联是对两个路标(VSLAM 中路标指图像特征)进行匹配, 确定它们是否对应环

境中的同一物体。在大方向上, 特征匹配解决了 SLAM 中的数据关联问题, 但在这个过程中带有误差, 所以对图像特征匹配的结果优化是必要的, 主要方法有固定区域匹配、Active Matching、1-Point RANAC、几何约束等。

PTAM(parallel tracking and mapping)^[18] 及其改进算法主要使用固定区域匹配的方法。PTAM 假定前后两帧图像中像素距离在一个阈值内, 超出这个阈值就认为是错误匹配, 该法适用于特征点距离相机稍远、深度变化不大的场合, 不适用于相机快速运动的场合。

基于 EKF 滤波的 VSLAM 系统多采用 Davison 提出的 Active Matching^[24]方法。Active Matching 方法中, 在使用 EKF 系统运动模型获得系统状态预测的基础上, 估计环境中的特征点在相机中的投影位置, 再进一步处理即可得到图像中特征点的分布区域。此方法对相机的绝大部分运动情况鲁棒, 但如果出现相机姿态估计协方差较大的情况, 易产生大的特征匹配估计区域, 可能匹配错误。

为去除 Active Matching 中的错误匹配, Civera、Grasa 等提出 1-Point RANSAC^[25-26]方法。该方法用随机选择的一个匹配点的匹配信息来更新相机姿态, 之后计算其他匹配点与估计图像位置的距离, 并判断这个距离是否在一定的阈值范围内, 若不在, 被认为是外点并剔除它, 最后利用得到的内点集来更新整个滤波器状态。该方法主要应用在基于 EKF 滤波的 SLAM 系统中, 由于频繁地更新系统状态, 运算时间代价比较大。

几何约束方法利用 PNP(perspective N points)^[27]、对极几何^[28]等剔除误匹配点。该方法因利用几何求解, 不需要频繁更新系统状态, 故而能获得较好的系统运行速度。但是对于不同的情况该方法需要具体问题具体分析, 使用相对应的几何约束条件, 相应地增加了系统的复杂性。

1.4 累积误差

SLAM 中的误差来源主要为里程计误差、观测误差和错误的关联带来的误差 3 个方面。在 VSLAM 中, 环境的先验信息和机器人的位置都是未知的, 位置误差(视觉里程计误差)不能根据环境先验信息得到有效纠正, 故而随着机器人运动距离的增大位置误差也逐渐累积。位置误差的增大会造成错误的关联, 相应的特征标志的误差也跟着增大; 反过来, 机器人的位置误差因为参考了有误差的特征也会增大。因此, 里程计误差与特征标志之间相互影响使整个 VSLAM 系统产生累积误差, 无法保证地图和轨迹的全局一致性。图 4 中, 累积误

差使得估计轨迹和真实轨迹相差很大。当前 VSLAM 系统多采用回环检测的方式减小这一误差。回环检测是指机器人识别出曾经到达过的场景的能力,当机器人看到两张相似图片时,计算图像数据的相似性,如果回环检测成功,可以显著地减小累积误差。回环检测在 VSLAM 中意义重大,既关系到估计的地图和轨迹在长时间下的正确性,也可在跟丢时进行重定位,大大增强了系统的鲁棒性。

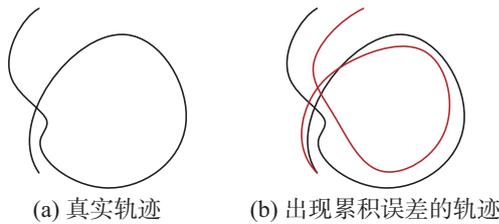


图 4 真实轨迹与出现累积误差的轨迹

Fig. 4 Real track and track with accumulated error

2 VSLAM 实现方法

VSLAM 的实现方法分为基于滤波器的方法和基于图优化的方法。其中,基于滤波器的方法只估计当前时刻的位姿,是一种增量式算法;基于图优化的方法根据所有观测到的信息,对整个机器人运动轨迹进行估计。前者又称为在线 SLAM,后者又称为全 SLAM(FULL SLAM)。表 1 给出了常用的开源 VSLAM 方案,其中有使用滤波方法的,也有使用优化方法的,本文 2.1 和 2.2 节将对典型方案详述。

表 1 常用开源 VSLAM 方案

Table 1 Commonly used open source VSLAM solutions

方案名称	传感器形式	地址
MonoSLAM	单目	https://github.com/hanmekim/SceneLib2
PTAM	单目	http://www.robots.ox.ac.uk/~gk/PTAM/
ORB-SLAM	单目为主	http://webdiis.unizar.es/~raulmur/orbslam/
LSD-SLAM	单目为主	http://vision.in.tum.de/research/vslam/lsdslam
SVO	单目	https://github.com/uzh-rpg/rpg_svo
DTAM	RGBD	https://github.com/anuranbaka/OpenDTAM
DVO	RGBD	https://github.com/tum-vision/dvo_slam
RTAB-MAP	双目/RGBD	https://github.com/introlab/rtabmap
RGBD-SLAM-V2	RGBD	https://github.com/felixendres/rgbdslam_v2
Elastic Fusion	单目	https://github.com/mp3guy/ElasticFusion

2.1 基于滤波器的实现方法

2.1.1 基于扩展卡尔曼滤波器 (extended kalman filter, EKF) 的 EKF-VSLAM

21 世纪之前,SLAM 中的状态估计主要使用滤波的方法。在 SLAM 中,系统的状态由机器人的位姿和地图信息(路标)组成。用卡尔曼滤波器(KF)实现 SLAM 必须遵循运动方程和观测方程都符合线性高斯模型、系统的状态服从高斯分布这两个假设。基于 KF 的 SLAM 由系统状态预测和更新两步组成,与此同时,对地图进行加入新路标、删除旧路标等操作。KF 中,假设系统都是线性的,但是现实中,机器人的运动模型与观测模型往往都是非线性的。对此,通常采用一阶泰勒展开来近似表示非线性模型的扩展卡尔曼滤波器(extended Kalman filter, EKF)方法来实现 SLAM。

卡尔曼滤波器是实现 SLAM 的基本方法之一^[29]。其协方差矩阵描述了机器人的位置和地图的不确定信息。当机器人连续观测到环境中的特征标志时,所有协方差矩阵子阵的行列式呈单调递减。每一时刻机器人能观测到路标不会很多,只有少数几个。基于卡尔曼滤波器的 SLAM 的时间复杂度为 $O(n^2)$, n 表示地图中的特征标志数^[30]。为了达到降低 SLAM 的时间复杂度的目的,Leonard 等^[31]提出了 DSM (decoupled stochastic mapping) 方法。DSM 中机器人位置估计被各子地图分别保存,当机器人从 1 个子地图运动到另 1 个子地图时,将前 1 个子地图的信息以 EKF 的方式传送给后 1 个子地图。Williams 等^[32]提出的基于 CLSF (constrained local submap filter) 的 SLAM 方法涉及全局坐标已知的子地图,首先构建出这些子地图,然后机器人运动过程中只利用观测信息更新自身位置和局部子地图中的特征标志,并且在时效范围内向全局地图传递局部子地图信息。Guivant 等^[33]提出了 1 种没有任何信息丢失的 SLAM 优化算法 CEKF (compressed extended Kalman filter)。在 CEKF 中,已观测到的地图路标一分为二分成 A 与 B 两部分,比较特别的是,用 A 来记录活动子地图(机器人当前位置的邻域)。当机器人在 A 中运动时,机器人的位置与地图 A 通过观测信息得到实时更新,与此同时,地图 B 受到子观测信息的影响被递归地记录;当机器人运动到 A 的区域之外时,观测信息被传送给子地图 B,地图 B 进行一次性更新,新的活动子地图同时被创建。

为了降低 SLAM 的时间复杂度,Thrun 等^[34]也提出去相关的方法,即基于稀疏信息滤波器(sparse extended information filter, SEIF)的 SLAM 方法,该

方法中, 只对约束关系进行局部更新, 这种局部更新使得信息矩阵近似于系数矩阵, 有效降低 SLAM 的时间复杂度。

Davison^[35]于 2007 年提出的 MonoSLAM, 是第一个基于 EKF 方法实时的单目 VSLAM 系统, 虽然初步解决了实时的问题, 能够在线创建稀疏地图, 漂移多少仍然不能确定, 目前已经停止对其的开发。

图 5 是基于 EKF 的单目 VSLAM 流程图。

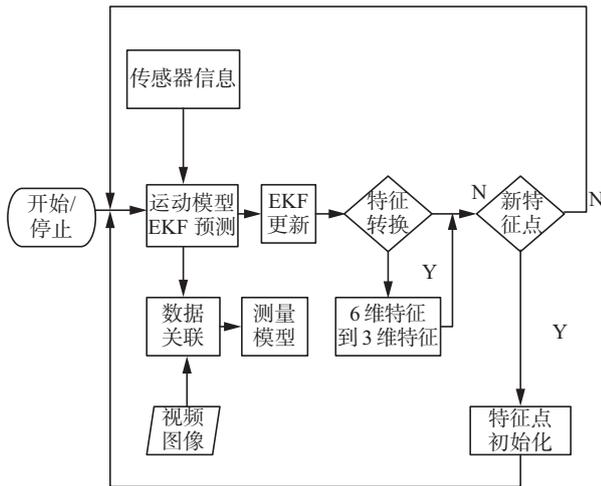


图 5 基于 EKF 的单目 VSLAM 流程图

Fig. 5 Flowchart of EKF-based monocular VSLAM

2.1.2 基于粒子滤波器的 FastSLAM

M. Montemerlo 等^[36-37]提出了 1 种基于粒子滤波器 (particle filter, PF) 的 FastSLAM 方法。FastSLAM 包含了机器人定位和特征标志位置估计两个过程。粒子滤波器法中机器人可能的运动路径用粒子表示, 1 个粒子对应着 1 种可能, 每条路径的好坏由利用观测信息计算得到的粒子权重来评价。对于每个粒子来说机器人的运动路径是确定的, 故特征标志之间相互独立且其观测信息只与机器人的位姿有关。FastSLAM 的时间复杂度为 $O(kn)$, 其中 k 为粒子个数^[30]。用树形数据结构优化后的时间复杂度可以达到 $O(k \log n)$ ^[30]。FastSLAM 能够比较好地表示机器人的非线性、非高斯运动模型。

EKF 存在非线性误差, 且需要存储、维护和更新状态量的均值和方差。如果把路标也加入状态的话, 由于 VSLAM 中路标数量很大, 这个存储量是相当大的, 且与状态量呈平方增长 (因为要存储协方差矩阵)。因此, EKF 普遍被认为不适用于大型场景。PF 采样所需的粒子数量, 随维度的增加呈指数增长, 所以仅限于低维的问题, 对高维问题不适用。除此之外, 滤波器方法在一定程度上假设了马尔可夫性, 如果当前帧与很久之前的帧有关 (例如回环), 那么滤波器就会难以处理这种情况。因为滤波这些明显的缺点, 在同等计算量的情况下, 非

线性优化的方法 (现代 SLAM 系统) 可以取得更好的效果^[10]。

另外, 在 2011 年滤波方面出现了基于 RFS (random finite set) 的方法^[38]。RFS 是滤波中新兴的潮流^[38], RFS 是以集合为元素的集合, 此集合中的元素及元素个数都是随机变量。文献^[38]对环境地图和传感器观测信息用 RFS 建模, 构造联合目标状态变量的 RFS。依据贝叶斯滤波框架, 利用概率假设密度滤波 (probability hypothesis density, PHD)^[39]实现对机器人位姿和环境地图同时估计。该算法避免了数据关联的问题, 相对于 EKF 和 PF 能更有效地表达 SLAM 问题。

2.2 现代 SLAM 系统: 基于非线性优化的方法

现代 SLAM 系统分为两个部分: 前端和后端^[1]。前端提取传感器数据构建模型用于状态估计, 后端根据前端提供的数据进行优化。这个架构如图 6 所示。

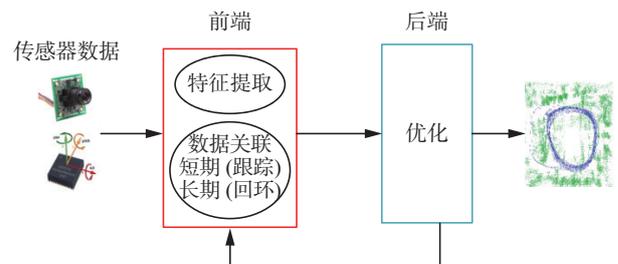


图 6 典型 SLAM 系统

Fig. 6 Typical SLAM system

当前 SLAM 事实标准形成来源于 Lu 和 Milios^[40], 它是 Gutmann 和 Konolige^[41]研究的后续。典型的 SLAM 系统如图 6 所示, 前端进行特征提取、数据关联和初值优化。前端的数据关联模块包括 1 个短期 (局部) 数据关联模块和 1 个长期 (回环) 数据关联模块。通常意义下的数据关联问题在 SLAM 中是指递增定位与建图过程中如何确定当前连续的传感器观测之间或者当前时刻的观测与最近所创建的局部地图中特征间的关联关系, 这也称为短期 (局部) 数据关联; 回环检测中的数据关联研究机器人沿不同的路径回到某一循环的起点时, 如何确定当前创建的局部地图中的特征与以前所创建的循环起点处地图中的特征间的关联关系, 这称为长期 (回环) 数据关联。短期的数据关联模块负责关联传感器中连续的观测值对应的特征: 得到 1 帧图像数据后, 对其进行预处理, 筛选出关键帧, 对图像进行特征提取、匹配以及运动求解并得到局部地图, 也就是视觉里程计 (visual odometry, VO); 长期的数据关联负责将新的观测值关联到旧的路标上, 也就是回环 (loop closure)。

前端对视觉传感器得到的图像的处理考虑两帧的信息,出现错误匹配后便无法进行,视觉里程计也存在累积误差;后端用来解决这些问题,对全局进行优化。后端优化不仅仅考虑相邻两帧之间的信息,而且把所有帧(包括很久之前的状态)的信息都考虑进来,其本质上是1个线性最小二乘问题。与此同时,后端也将信息反馈给前端用于回环检测、验证等。上述优化问题在SLAM中表示成图(图论意义上的图)的形式。1个图由若干个顶点,以及连接着这些顶点的边组成;用顶点表示优化变量,用边表示误差项。于是对任意1个上述形式的非线性最小二乘问题可以构建与之对应的1个图。后端优化工具常用g2o、Google Ceres、GTSAM等。

2.2.1 基于特征提取的VSLAM

2007年,Klein等^[18]提出的PTAM(parallel tracking and mapping)是基于关键帧的单目VSLAM系统。PTAM提出并实现了跟踪与建图过程的并行化。这是视觉SLAM中首次出现前后端的概念,对后来许多视觉SLAM系统的设计提供了方向。PTAM将姿态跟踪和建立地图设计为两个线程:姿态跟踪线程(前端)不修改地图,只是用已知地图来快速跟踪;建立地图线程(后端)引入关键帧机制,把若干个关键帧串起来继而优化其轨迹和地图,此过程专注于地图的建立、维护和更新。PTAM在选取关键帧、匹配特征、估计相机位置方面十分出色,PTAM在重定位上(跟踪失败后需要进行重定位以继续跟踪)也表现突出,然而该算法没有设计回环检测部分,且缺少帧数过多时的处理,视图不变特性较差,鲁棒性不够,需要人工干预地图的构建过程。

RGBD-SLAM-V2^[42]是F. Endres在2014年提出的一种使用深度相机计算SLAM的系统。RGBD-SLAM-V2前端提取3-D点的图像特征,进行匹配,渲染点云,后端建立位姿图并用g2o进行优化,最后输出地图。RGBD-SLAM-V2仅使用深度相机——RGBD相机,同时采用SLAM领域里的图像特征提取、回环检测、点云、图优化等目前非常流行的技术,效果良好,但特征点提取、点云渲染都是费时的环节,算法实时性有待改进。

2015年,ORB-SLAM^[43]算法被提出。ORB-SLAM整合当前各种主流的方式计算SLAM,是现代SLAM系统中非常完善且易用的算法之一。ORB-SLAM整个系统都是围绕着ORB特征实现的,对输入的图像信息处理时使用ORB特征,在视觉里程计与回环检测环节使用ORB字典。ORB-SLAM创新地使用实时跟踪特征点线程、局部捆集优化线

程以及全局位姿图的回环检测与优化线程三线程完成SLAM,能够确保地图与轨迹的全局一致性。其中,实时跟踪特征点线程对每帧新来的图像提取ORB特征,并与最近的关键帧比较,计算特征点的位置并粗略估计相机位姿;局部捆集优化线程求解1个包括局部空间内的特征点与相机位姿的BA(bundle adjustment)问题;全局位姿图的回环检测与优化线程对全局的地图关键帧进行回环检测,消除累积误差。继PTAM的双线程结构之后,ORB-SLAM的三线程结构又给研究者带来新的思路。相比于PTAM,ORB-SLAM有着更为宽松的关键帧选取策略。ORB-SLAM在优化位姿时循环优化4次,确保得到更多的正确匹配。所有的这些改进使得ORB-SLAM具有更好的鲁棒性,即使在较差的场景也能顺利工作。ORB-SLAM2^[44]提供单目、双目和RGBD接口,这使得它具有更好的泛用性。然而,ORB的建图部分只含有稀疏的地图点,只能满足定位需求,用于导航就不足了。另外,整个ORB-SLAM系统都采用特征点进行计算,对于每张图都要计算ORB特征,这使得ORB-SLAM系统非常耗时。

2.2.2 基于直接法的VSLAM

特征点法在VSLM中占据着主流地位,但关键点的提取与描述子的计算非常耗时,而且使用特征点时,只使用到特征点的信息,其他的像素点都被忽略了。针对这些缺点,出现了直接法VSLAM,该法既不必提取关键点也不必计算描述子,而是基于灰度不变假设,根据像素来直接计算相机运动。直接法是从光流^[45]演变而来的,光流描述了像素在图像中的运动,而直接法则附带1个相机运动模型。直接法中,假设在各个视角下,1个空间点成像的灰度值是恒定不变的。该方法直接根据像素亮度信息,估计相机的运动。直接法完全舍去了关键点提取和描述子计算,不仅避开了耗时的特征提取,而且保留了所有特征信息。

文献^[46]里的DTAM(dense tracking and mapping),继续使用关键帧的架构,但对关键帧的处理与传统的特征点提取大相径庭。DTAM中的直接法(direct method)在默认环境亮度不变的前提下,对每一个像素的深度数据进行反深度(inverse depth)计算和不断优化,从而建立稠密地图并且稳定地跟踪。DTAM准、稳,但对每一个像素都计算使得实时计算困难,需通过GPU加速。

文献^[47]中J. Engel等在2014年提出了LSD-SLAM(large-scale direct monocular SLAM),LSD-SLAM是直接法在单目SLAM中的实现。LSD-

SLAM 放弃了 BA (bundle adjustment) 的方法,利用直接法构建了半稠密地图,可在实时应用中运行,非常适合机器人使用。LSD-SLAM 对相机内参和相机曝光非常敏感,且相机快速运动时不容易跟踪。LSD-SLAM 中相机定位的精度明显比 PTAM 和 ORB 低^[43]。在回环检测部分,目前尚未有在直接法上实现的回环检测方式,LSD-SLAM 仍需依赖基于特征的回环检测方法。

文献[48]中 Forster 等提出了介于直接方式和基于特征的方法之间的半直接法 SVO。SVO 中特征点与直接法混合使用,跟踪一些关键点(无描述子的角点),然后用直接法,根据这些关键点周围的信息,估计相机的运动和位置。SVO 不必计算描述子,处理的信息少,速度极快,在 4 轴飞行器上效果很好。然而,SVO 为了速度和轻量化,没有回环检测、后端优化部分,这导致了 SVO 必然存在累计误差,且跟丢后难以重定位。

直接法可以构建半稠密乃至稠密的地图,而特征点法因为自身局限是无法做到的。对比稀疏的点云地图,比如 ORB-SLAM 或 PTAM,构建稠密的地图对相机定位和导航意义更加深刻,作用也更加突出。直接法只要求有像素梯度即可估计相机运动,而无须特征点。上文中的 DTAM 和 LSD-SLAM 可以对环境进行稠密或半稠密的地图构建,相机位姿通过图像像素的灰度直接优化。直接法避开了提取特征这一繁琐和耗时的过程,大大提高了系统效率,且在图像模糊、低纹理环境和高频纹理的情况下有着更好的鲁棒性。

然而,直接法有其局限性。这些方法采用一个表面反射模型,灰度值不变这一严苛的假设需要严格遵守。灰度间的差异计算是直接法赖以实现的基础,如果环境亮度或者图像整体灰度变化,例如环境突然发生光线变换或者相机突然调整曝光参数,整体的图像就会变亮或变暗,于是灰度不变假设不再成立,算法失败。梯度不明显的像素对环境重构的精度影响很大,对于非深度相机需要更多的观测信息以弥补深度的不确定性。直接法可能会受到快门、自动增益和自动曝光的影响,而且对相机的图像采集速率也有较高的要求。相反,基于特征的 VSLAM 方法得益于其较好的视图不变特性,可以在更宽的条件下匹配特征、捆集调整和优化相机位姿。文献[49]也指出了特征点法相比直接法的优势。

3 结束语

综上所述,随着机器视觉技术的兴起,VSLAM 成为机器人领域越来越热的问题,该领域的研究者

们致力于降低算法复杂度、减小误差、提高效率 and 精度、提高鲁棒性等。但是在以下几个方面还需要更进一步研究。

1) 静态环境到动态环境

目前,大部分 VSLAM 系统仍然局限于研究理想情况下的静态环境,然而事与愿违,现实中的环境通常是复杂的动态环境,在动态环境下研究 VSLAM 具有重要的意义。动态环境中在语义地图下做 VSLAM 是一个可选择的方法。利用语义地图先判断物体是否为可动物体,然后继续下一步的动作。

2) 与人工智能相结合的 VSLAM 实现方法

可以将人工智能领域的方法引入到 VSLAM 中^[50],寻找更有效的 VSLAM 算法,如文献[51]。上文提及的语义地图也可尝试用深度学习的方式实现。

3) 多传感器融合

本文所提到的算法都是在仅用相机作为传感器的前提下,而在算法的实用方面,仅用单一传感器的 VSLAM 有固有局限,如相机快速运动下不容易跟踪、难以处理动态的障碍物等,因此,将不同的传感器数据融合起来进行互补,可使系统更加鲁棒、精度更高。例如,惯导组合的 VSLAM 更能适用于复杂的场景。惯性传感器(IMU)能够测量传感器本体的加速度和角速度,与相机传感器互补,两者融合之后能得到更完善的 VSLAM 系统,然而如何有效结合这两者是值得深入探讨的问题。

4) 多机器人协作

对于单个机器人的 VSLAM 系统已经有不少的解决方案,然而多机器人 VSLAM 领域还有通信拓扑、任务规划和地图融合等方面有待研究^[52]。

上述 VSLAM 发展方向中,结合深度学习的 VSLAM 和多传感融合 VSLAM 将在 SALM 领域占据重要地位;动态环境下的 VSLAM 和多机器人协作 VSLAM 仍需砥砺前行。

参考文献:

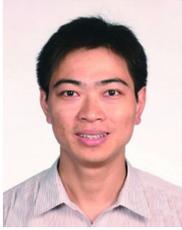
- [1] CADENA C, CARLONE L, CARRILLO H, et al. Simultaneous localization and mapping: present, future, and the robust-perception age[Z]. Computer science, 2016.
- [2] FUENTES-PACHECO J, RUIZ-ASCENCIO J, RENDÓN-MANCHA J M. Visual simultaneous localization and mapping: a survey[J]. Artificial intelligence review, 2015, 43(1): 55-81.
- [3] 于金霞,王璐,蔡自兴.未知环境中移动机器人自定位技术[M].北京:电子工业出版社,2011.
YU Jinxia, WANG Lu, CAI Zixing. Self-localization technologies of mobile robot in unknown environment[M]. Beijing: Publishing House of Electronics Industry, 2011.

- [4] 刘浩敏, 章国锋, 鲍虎军. 基于单目视觉的同时定位与地图构建方法综述[J]. 计算机辅助设计与图形学学报, 2016, 28(6): 855–868.
LIU Haomin, ZHANG Guofeng, BAO Hujun. A survey of monocular simultaneous localization and mapping[J]. Journal of computer-aided design and computer graphics, 2016, 28(6): 855–868.
- [5] SMITH R, SELF M, CHEESEMAN P. Estimating uncertain spatial relationships in robotics[J]. Machine intelligence and pattern recognition, 1988, 1(5): 435–461.
- [6] SMITH R C, CHEESEMAN P. On the representation and estimation of spatial uncertainty[J]. The international journal of robotics research, 1986, 5(4): 56–68.
- [7] CSORBA M. Simultaneous localisation and map building[D]. Oxford: University of Oxford, 1997: 699–704.
- [8] DISSANAYAKE M W M G, NEWMAN P, CLARK S, et al. A solution to the simultaneous localization and map building (SLAM) problem[J]. IEEE transactions on robotics and automation, 2001, 17(3): 229–241.
- [9] AGARWAL S, SNAVELY N, SIMON I, et al. Building Rome in a day[C]//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan, 2009: 72–79.
- [10] STRASDAT H, MONTIEL J M M, DAVISON A J. Real-time monocular SLAM: why filter?[C]//Proceedings of 2010 IEEE International Conference on Robotics and Automation. Anchorage, AK, USA, 2010: 2657–2664.
- [11] 梁明杰, 闵华清, 罗荣华. 基于图优化的同时定位与地图创建综述[J]. 机器人, 2013, 35(4): 500–512.
LIANG Mingjie, MIN Huaqing, LUO Ronghua. Graph-based SLAM: a survey[J]. Robot, 2013, 35(4): 500–512.
- [12] MUJA M, LOWE D G. Fast approximate nearest neighbors with automatic algorithm configuration[C]//Proceedings of the 4th International Conference on Computer Vision Theory and Applications. Lisboa, Portugal, 2009: 331–340.
- [13] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91–110.
- [14] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer vision and image understanding, 2008, 110(3): 346–359.
- [15] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//Proceedings of 2011 IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 2564–2571.
- [16] CALONDER M, LEPETIT V, STRECHA C, et al. BRIEF: binary robust independent elementary features[C]//Proceedings of the 11th European Conference on Computer Vision. Heraklion, Crete, Greece, 2010: 778–792.
- [17] CIVERA J, DAVISON A J, MONTIEL J M M. Inverse depth parametrization for monocular SLAM[J]. IEEE transactions on robotics, 2008, 24(5): 932–945.
- [18] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces[C]//Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara, Japan, 2007: 225–234.
- [19] KLEIN G, MURRAY D. Improving the agility of keyframe-based SLAM[C]//Proceedings of the 10th European Conference on Computer Vision. Marseille, France, 2008: 802–815.
- [20] WEISS S, SIEGWART R. Real-time metric state estimation for modular vision-inertial systems[C]//Proceedings of 2011 IEEE International Conference on Robotics and Automation. Shanghai, China, 2011: 4531–4537.
- [21] JAMA M, SCHINSTOCK D. Parallel tracking and mapping for controlling VTOL airframe[J]. Journal of control science and engineering, 2011, 2011: 413074.
- [22] DAVISON A J. SLAM with a single camera[C]//Proceedings of Workshop on Concurrent Mapping and Localization for Autonomous Mobile Robots in Conjunction with ICRA. Washington, DC, USA, 2002.
- [23] 祝风翔. 基于图像的深度获取方法研究[D]. 杭州, 中国: 浙江大学, 2016.
ZHU Fengxiang. Depth map acquisition method study based on image[D]. Hangzhou, China: Zhejiang University, 2016.
- [24] DAVISON A J. Active search for real-time vision[J]. Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China, 2005: 66–73.
- [25] CIVERA J, GRASA O G, DAVISON A J, et al. 1-point RANSAC for EKF-based structure from motion[C]//Proceedings of 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, MO, USA, 2009: 3498–3504.
- [26] CIVERA J, GRASA O G, DAVISON A J, et al. 1-Point RANSAC for extended Kalman filtering: application to real-time structure from motion and visual odometry[J]. Journal of field robotics, 2010, 27(5): 609–631.
- [27] GUERRA E, MUNGUIA R, BOLEA Y, et al. Validation of data association for monocular SLAM[J]. Mathematical problems in engineering, 2013, 2013: 671376.
- [28] EADE E, DRUMMOND T. Scalable monocular SLAM [C]//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA, 2006: 469–476.

- [29] LEONARD J J, DURRANT-WHYTE H F. Simultaneous map building and localization for an autonomous mobile robot[C]//Proceedings of Intelligence for Mechanical Systems, Proceedings IROS'91. IEEE/RSJ International Workshop on Intelligent Robots and Systems'91. Osaka, Japan, 1991: 1442–1447.
- [30] 罗荣华, 洪炳镕. 移动机器人同时定位与地图创建研究进展[J]. 机器人, 2004, 26(2): 182–186.
LUO Ronghua, HONG Bingrong. The progress of simultaneous localization and mapping for mobile robot[J]. Robot, 2004, 26(2): 182–186.
- [31] LEONARD J J, FEDER H J S. Decoupled stochastic mapping[J]. IEEE journal of oceanic engineering, 2001, 26(4): 561–571.
- [32] WILLIAMS S B. Efficient solutions to autonomous mapping and navigation problems[D]. Sydney, Australia: University of Sydney, 2001.
- [33] GUIVANT J E, NEBOT E M. Optimization of the simultaneous localization and map-building algorithm for real-time implementation[J]. IEEE transactions on robotics and automation, 2001, 17(3): 242–257.
- [34] THRUN S, LIU Yufeng, KOLLER D, et al. Simultaneous localization and mapping with sparse extended information filters[J]. The international journal of robotics research, 2004, 23(7/8): 693–716.
- [35] DAVISON A J, REID I D, MOLTON N D, et al. MonoSLAM: real-time single camera SLAM[J]. IEEE transactions on pattern analysis and machine intelligence, 2007, 29(6): 1052–1067.
- [36] MONTEMERLO M, THRUN S, KOLLER D, et al. FastSLAM: a factored solution to the simultaneous localization and mapping problem[C]//Proceedings of the 18th International Conference on Artificial Intelligence. Pittsburgh, PA, USA, 2002: 2004.
- [37] MONTEMERLO M, THRUN S. Simultaneous localization and mapping with unknown data association using FastSLAM[C]//Proceedings of 2003 IEEE International Conference on Robotics and Automation. Taipei, Taiwan, China, 2003: 1985–1991.
- [38] MULLANE J, VO B N, ADAMS M D, et al. A random-finite-set approach to Bayesian SLAM[J]. IEEE transactions on robotics, 2011, 27(2): 268–282.
- [39] ADAMS M, VO B N, MAHLER R, et al. SLAM gets a PHD: new concepts in map estimation[J]. IEEE robotics and automation magazine, 2014, 21(2): 26–37.
- [40] LU F, MILIOS E. Globally consistent range scan alignment for environment mapping[J]. Autonomous robots, 1997, 4(4): 333–349.
- [41] GUTMANN J S, KONOLIGE K. Incremental mapping of large cyclic environments[C]//Proceedings of 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Monterey, CA, USA, 1999: 318–325.
- [42] ENDRES F, HESS J, STURM J, et al. 3-D mapping with an RGB-D camera[J]. IEEE transactions on robotics, 2014, 30(1): 177–187.
- [43] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE transactions on robotics, 2015, 31(5): 1147–1163.
- [44] 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, 23(5): 1255–1262.
- [45] BLESER G, HENDEBY G. Using optical flow as lightweight SLAM alternative[C]//Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality. Orlando, FL, USA, 2009: 175–176.
- [46] NEWCOMBE R A, LOVEGROVE S J, DAVISON A J. DTAM: dense tracking and mapping in real-time[C]//Proceedings of 2011 IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 2320–2327.
- [47] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: large-scale direct monocular SLAM[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland, 2014: 834–849.
- [48] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: fast semi-direct monocular visual odometry[C]//Proceedings of 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China, 2014: 15–22.
- [49] TORR P H S, ZISSERMAN A. Feature based methods for structure and motion estimation[M]. TRIGGS B, ZISSERMAN A, SZELISKI R. Vision Algorithms: Theory and Practice. Berlin, Heidelberg, Germany: Springer, 2000: 278–294.
- [50] 林辉灿, 吕强, 张洋, 等. 稀疏和稠密的 VSLAM 的研究进展[J]. 机器人, 2016, 38(5): 621–631.
LIN Huican, LYU Qiang, ZHANG Yang, et al. The sparse and dense VSLAM: a survey[J]. Robot, 2016, 38(5): 621–631.
- [51] GAO Xiang, ZHANG Tao. Unsupervised learning to detect loops using deep neural networks for visual SLAM system[J]. Autonomous robots, 2017, 41(1): 1–18.
- [52] 张国良, 汤文俊, 曾静, 等. 考虑通信状况的多机器人 CSLAM 问题综述[J]. 自动化学报, 2014, 40(10): 2073–2088.

ZHANG Guoliang, TANG Wenjun, ZENG Jing, et al. An overview on the cooperative SLAM problem of multi-robot systems considering communication conditions[J]. Acta automatica sinica, 2014, 40(10): 2073–2088.

作者简介:



周彦,男,1978年生,副教授,博士,主要研究方向为多传感器信息融合、图像处理与机器视觉。发表学术论文 40 余篇,其中被 SCI 收录 10 余篇。目前主持国家自然科学基金项目 1 项,参与国家自然科学基金项目 4 项;曾参与及主持“973”、国家自然科

学基金等项目 15 项。



李雅芳,女,1993年生,硕士研究生。主要研究方向为图像处理与机器视觉。



王冬丽,女,1980年生,副教授,博士,主要研究方向为模式识别与机器视觉。发表学术论文 30 余篇。目前主持国家自然科学基金项目 1 项。曾参与国家自然科学基金、上海市自然科学基金等项目。

2018 第 10 届 IEEE 通信软件和网络国际会议 (ICCSN2018) 2018 10th International Conference on Communication Software and Networks (ICCSN 2018)

2018 年第十届通信软件和网络国际会议将于 2018 年 7 月 6—9 日在中国成都召开。会议旨在促进通信软件和网络等领域的学术交流与合作,热忱欢迎从事相关技术研究的专家、学者和专业技术人员踊跃投稿并参加大会。IEEE,电子科技大学联合举办,中电 54 所,通信网信息传输与分发技术重点实验室,是本次会议的协办方,广东工业大学为技术支持单位,《电子技术应用》是会议的赞助方。

出版和检索说明:

文章将会根据主题推荐 JCM 或者 IEEE 录用,请以通知书为准完成注册。

1. Journal of Communications (JCM) 期刊为 Scopus 检索期刊,目前该 JCM 期刊 Scopus 检索正常。

2. ICCSN 2018 由 IEEE 出版至 ICCSN 会议论文集,并按照 IEEE 要求会后按时提交到 IEEE 审查,审查通过的论文集将进入 IEEE Xplore 并被 Ei Compendex 检索。

ICCSN 2017 论文集已整体进入 IEEE Xplore 数据库,检索结果即将公布。

征稿:

主题:关于通信软件和网络。

详细信息请见:<http://www.iccsn.org/cfp.html>

投稿:

1. 直接把文章发到会议邮箱:iccsn_conference@163.com

2. 上传文章到电子投稿系统---<http://www.easychair.org/conferences/?conf=iccsn2018>

详细信息请见---<http://www.iccsn.org/guide.html>

会议联系方式:

会议邮箱: iccsn_conference@163.com

会议网址: <http://www.iccsn.org/>