

doi:10.19306/j.cnki.2095-8110.2017.02.001

视觉导航技术的发展

张 海¹, 吴克强¹, 张晓鸥²

- (1. 北京航空航天大学自动化科学与电气工程学院, 北京 100191;
2. 北京航空航天大学无人驾驶飞行器设计研究所, 北京 100191)

摘要:近20年来视觉导航技术发展迅速,在原理方法与技术实现上取得了重要突破,规模化应用已经成为可能。以视觉导航中光流、视觉里程计、视觉同步定位与地图构建等主流研究方向为对象,对其发展概况、基本原理、技术优势与难点等进行了简要介绍与分析,并从近期应用推广及长远发展角度对视觉及视觉组合导航技术进行了展望。

关键词:SFM; 视觉里程计; VSLAM; 光流; 组合导航

中图分类号:TP391.4 **文献标志码:**A

文章编号:2095-8110(2017)02-0001-08

The Development of Visual Navigation

ZHANG Hai¹, WU Ke-qiang¹, ZHANG Xiao-ou²

- (1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;
2. Unmanned Aerial Vehicle Research Institute, Beihang University, Beijing 100191, China)

Abstract: Along with the fast development of visual navigation in the last two decades, some significant progresses have been achieved and applications in wide fields are ready to come. The developments, theories, advantages and difficulties are introduced according to the main visual navigation algorithms of optical flow, Visual Odometry and VSLAM. Finally, the future researches of coming and long term stages are analyzed in brief.

Key words: SFM; VO; VSLAM; Optical flow; Integrated navigation

0 引言

导航伴随着人类文明的发展已有数千年的历史,20世纪以来更出现了跨越式的进步,以惯性、无线电、天文、地磁、光学为基础的导航系统发展日新月异,在军事领域取得举足轻重地位的同时,也正在加速进入民用领域。美国国防部高级研究计划局(DARPA)2004年开始组织自动驾驶车辆大赛,极大地推动了相关技术的发展,也客观促成了今天Google、Uber、Tesla、百度等无人驾驶技术的示范应用。亚马逊公司为提高仓库工作效率,收购了

Kiva system公司,并将Kiva机器人应用于自动化仓库,实现了大规模机器人全自动商品搬运,成为自动导引车辆(AGV)应用范例。我国洋山港自动化码头,在可靠的组合导航系统支持下,已采用130台AGV实现集装箱装卸、转运自动化,能24h工作,车辆定位精度达到2cm,港区的工作人员减少了75%,但工作效率却提升了30%。

视觉是自然界绝大多数动物采用的导航手段,因图像中含有丰富的空间信息,特别是视觉作为一种自主的测量方式,在卫星导航(GNSS)受干扰、遮挡而无法工作的情况下仍能有效发挥作用,相对惯

收稿日期:2017-02-07;修订日期:2017-03-08

基金项目:国家重点研发计划(2016YFB0502004);国家自然科学基金国际合作与交流项目(61320106010)

作者简介:张海(1970-),男,博士,副教授,主要从事惯性、视觉及组合导航技术等方面的研究。

E-mail: zhanghai@buaa.edu.cn

性导航具有良好的误差累积抑制能力,在导弹武器、火星探测机器人、无人驾驶汽车、小型无人机等领域已经得到有效的应用,并具有进一步小型化、低成本化能力,发展空间十分广阔。

20世纪50年代开始的计算机视觉理论研究奠定了视觉导航的基础^[1],70年代后 Moravec 等学者就开始了面向导航的视觉应用探究^[2],90年代后因计算机处理能力的快速进步,计算机视觉及视觉导航应用研究工作得到了快速发展,Hartley 在其研究成果基础上对视觉解算、几何解算进行了全面总结^[30],特别是2000年后 Nistér、Davison、Klein 等学者的创造性研究工作极大地推动了该领域研究的快速发展,视觉及视觉组合导航已经进入了一个全新的阶段。下面对视觉导航主要技术分支及其发展情况进行简要介绍、分析,并给出视觉及视觉相关导航技术的未来发展趋势。

1 视觉导航技术研究进展

视觉或图像技术在导航中的应用众多,如目标跟踪、景象匹配、距离测量等,但不能提供完整导航功能,如不具备仅利用该方法自身实现摄像机速度、位置、姿态等完整导航信息输出的能力,故未作为视觉导航方法进行讨论。

光流的概念在20世纪50年代由 Gibson 提出,已被证明是昆虫、鸟类有效的导航方式。光流利用连续采集图像获得速度、转动估计,是仿生导航重要的研究对象^[3]。与视觉里程计、视觉同时定位与地图构建不同,它不能仅依据图像内容的空间几何关系进行导航计算,而对相机相对环境的瞬时运动具有较好的测量能力, Lucas-Kanade^[4]、Horn-Schunck^[5]是两种最为经典的计算方法, Mukauwa^[6]针对光流计算中不连续性问题进行改进。光流已被广泛应用于障碍物检测与着陆^[7], Ohnishi 等^[8]利用光流法实现场景中支配性平面检测,进而为机器人提供导航服务。近年来,马里兰大学研究团队从仿生角度出发提出了 Wide Field Integration 处理模型^[9-10],证明了光流法可以满足6自由度自主导航应用。

SFM (Structure From Motion,运动重建结构)是利用二维图像序列恢复空间三维结构及相机位姿信息方法的总称。虽然 SFM 具有相机位姿估计的功能,但其初衷是以环境的三维重构为首要目标,相机运动甚至可以是已知或可直接测量的。

SFM 概念形成较早,20世纪80年代 Longuet-Higgins^[11]、Harris^[12]等的研究工作实质性地推动了 SFM 的发展,SFM 算法的研究热点也从借助惯性等辅助测量^[13]向纯视觉处理方向转变^[14],且研究重点逐渐向单目视觉、相机精确定位、快速处理、稠密重构等方向集中。

VO (Visual Odometry,视觉里程计)是一个十分活跃的视觉导航研究方向,其功能是对相机的运动速度、轨迹及相机位姿进行准确估计,因此 VO 具有完整的导航应用功能,并被认为是一种特殊的 SFM^[15]。Nistér 在2004年提出了系统的本质矩阵五点估计法^[16],进而提出了 VO 的概念^[17],并给出了可实时处理的单目/双目 VO 实现方案,被认为是 VO 的开创性工作。迄今绝大多数 VO 采用特征点匹配的稀疏方案,精度较高。近年来出现了采用图像亮度而非特征的直接计算半稠密、稠密 VO^[18-19],在获得相机位姿信息的同时能够对环境进行大范围深度感知。目前 VO 的研究已经能够实现单目视觉下实时图像特征提取及相机定位,但在鲁棒性、准确性方面尚有很大的改善空间,VO 与惯性测量结合能够更有效应对复杂非结构化环境,取得良好的导航效果^[20]。

VSLAM (Visual Simultaneous Localization and Mapping,视觉同步定位与地图构建)是在经典 SLAM 算法基础上发展起来的视觉导航方案,以视觉取代传统 SLAM 中激光雷达等方式实现地标点测量,因其基于不同时刻二维图像进行导航定位,故也属于 SFM,研究难点主要集中在如何实现单目纯视觉条件下快速相机定位与地图构建^[21-22]。VSLAM 从其估计精度优化方式出发可以分为滤波、非滤波方法两大类^[15,21],滤波方法通过 EKF (Extended Kalman Filter,扩展 Kalman 滤波)、UKF (Uncented Kalman Filter,无迹 Kalman 滤波)以及 PF (Particle Filter,粒子滤波)等滤波解算提高导航精度,非滤波方法则是利用多幅图像构建对相机及特征点的联合几何约束进行导航参数优化。因受到单目视觉原理上的距离尺度模糊、地标数量造成的计算量过大两个共性难点的限制,直至2003年,两类研究仍然进展缓慢。

在滤波 VSLAM 方面,Davison 在2003年提出的一种单目实时 VSLAM 方案^[22]具有开创性,通过滤波模型设计及尺度概率估计方式得到小空间内有效的定位结果。针对单目视觉尺度模糊问题,

Nützi^[14]在 VSLAM 姿态位置估计结果基础上,建立了载体位置、速度、加速度及尺度因子的 EKF 估计模型,以 VSLAM 位置及加速度计测量为输入,实现了运动参数尺度修正估计;Montemerlo 等设计了基于 PF 的快速滤波方案,并证明了其滤波稳定性^[40]。非滤波 VSLAM 主要采用的优化方法有 BA (Bundle Adjustment, 光束平差法)、Pose Graph,其中 BA 应用最为普遍^[21],因其实现需要利用一段时间内众多关键帧建立几何约束关系,进而对相机运动参数进行最优搜索,计算量巨大,长期以来限制了实际应用。Klein 与 Murray 在 2007 年公布的 PTAM 算法通过高效的处理机制打破了 BA 研究的僵局^[23],使实时应用成为可能。因 PATM 是一种框架性的解决方案,目前已经成为 BA 类方法普遍采用模式,DTAM^[38]、ORB-SLAM^[32]、LSD-SLAM^[19] 等实时性算法都采用 PTAM 方案。

VO、VSLAM 同属于 SFM 方法,从功能上都能完整实现导航,且在特征提取、精度优化等实现步骤上可采用完全相同的技术,其区别主要体现在 VSLAM 针对往复运动区域,为保证地图及导航精度需要进行地图闭环处理 (Loop closure 或 Loop Closing),而 VO 不考虑闭环应用,在使用过程中仅保存最近的地图信息。

VO 及 VSLAM 两种最为主要的视觉导航方法已具有较好的实时处理能力,克服了传统 SFM 算法仅能事后离线处理的不足。使用极线几何或其他亮度信息计算方法^[18-19],能够实现针对图像中每个像素或局部区域每个像素的稠密、半稠密三维重构,进而可以获得丰富的场景信息,极大地推动了 AR (Augmented Reality, 增强现实) 领域的发展,基于 PTAM 框架的 VO、VSLAM 已经成为 AR 领域的重要技术方法。稠密、半稠密地图可以获得比特征点更为丰富的形状、轮廓等深度信息,目前因地图精度有限、复杂地图信息难以利用等原因,半稠密、稠密地图尚不能对导航精度改善发挥有效作用,但对未来视觉语义化的导航应用有重要意义,在此将基于特征点的稀疏地图方法作为主要分析对象。

立体视觉具有距离测量能力,相对单目视觉具有原理上的优势,目前立体视觉导航研究主要集中在应用方面,重点解决处理速度与导航精度的问题。单目视觉因不具备距离直接测量能力,在无外

界测量辅助或先验知识情况下,不具备距离无尺度模糊恢复能力。立体视觉因实际应用受到基线长度的限制,无法实现超过一定阈值距离的有效测量,退化为单目视觉。此外立体视觉对摄像机参数、基线长度的稳定性有很高的要求,也需要更高的处理能力。因此,取得单目视觉在原理、方法上的突破意义重大、更具挑战性,这也是目前 Davison 等国内外知名学者专注于单目视觉系统研究的主要原因。

2 光流法导航的基本原理

为有效描述动物在运动过程中视觉的变化,Gibson 定义的光流是因观测者与环境间的相对运动而在观测者视觉中明显表现出来的运动模式。

光流计算通常是建立在以下假设基础之上的:1)亮度一致性假设,给定点在不同的图像中具有相同的亮度;2)灰度梯度一致性假设,给定点在不同图像中具有相同灰度梯度;3)平滑性假设,相邻点位移不会剧烈变化;4)小幅运动假设,图像采集具有足够帧率,相邻帧仅存在小幅运动。

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (1)$$

$$I_x V_x + I_y V_y + I_t = 0 \quad (2)$$

光流的亮度不变假设如式(1),其中 I 为图像,通过泰勒展开、整理可以得到速度表达式(2), I_x 、 I_y 为图像沿 x 、 y 向灰度的偏导数, I_t 为图像对时间偏导数, V_x 、 V_y 分别为沿 x 及 y 方向的速度。下面仅对经典的光流计算原理进行简要说明。

Lucas-Kanade 法^[4],假定 p 点小领域内点具有相同的位移,利用相邻多 m 点正交方向速度一致性假设,在式(3)基础上利用最小二乘法获得速度估计。该方法对图像噪声有抑制作用,无迭代计算、效率高。在梯度变化不明显区域,因计算原理限制,无法求解,该方法通常是针对图像特征点进行计算的。

$$\begin{aligned} I_x(q_1)V_x + I_y(q_1)V_y &= -I_t(q_1) \\ &\vdots \\ I_x(q_m)V_x + I_y(q_m)V_y &= -I_t(q_m) \end{aligned} \quad (3)$$

Horn-Schunk 法^[5],基于图像相同点亮度一致的假设,认为图像中的光流平滑连续变化,在亮度一阶泰勒展开基础上,利用式(4)所示误差罚函数约束,进行迭代求解,获得全局、稠密运动信息。但因一阶泰勒展开的作用,要求图像位移小于 1 个像素,实际应用中采用模糊化、下采样等方式应对较

大位移影响。

$$E = \iint [(I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2)] dx dy \quad (4)$$

针对 Horn-Schunk 类算法因线性化带来的不足,文献[24]虽没有进行线性化,但算法复杂,仅能通过迭代的方式进行解算。为获得局部解算中良好的噪声抑制能力以及全局稠密光流,文献[25]提出了局部-全局组合方法(combined local-global methods)。Bruhn 等^[26]提出了可以有效实现局部、全局光流相结合组合检测方案。

因光流计算能够获得相对运动信息,从而实现障碍物探测、运动物体识别、运动速度估计、相对旋转判定等^[27-28],可以被广泛应用于机器人、无人机导航与自动驾驶系统中, Centeye 公司已经开发出低分辨率下可每秒计算达数千次的芯片级光流传感器^[29],并满足 2% 对比度、照度 1lux 下应用。

光流法是基于亮度不变等假设的,需要高帧频图像采集保证相邻帧满足计算要求,在图像移动较大情况下会失去计算合理性,不能利用不连续图像获得导航信息,此外光照的实时变化会影响计算精度。绝大多数光流计算是建立在单目视觉基础上的,其速度信息具有尺度模糊性。因此光流法更多应用于速度连续测量、避障等情况,虽然马里兰大学研究团队提出了较为有效的光流导航方案^[9-10],但因仅利用光流不能实现利用历史图像进行平差处理,对误差累积的抑制能力不强,在实现完整视觉导航功能方面相对 VO、VSLAM 尚有差距。

3 VO 的基本工作原理

多视图几何是 VO、VSLAM 工作的基本原理,在此对透视几何基本内容进行简要说明。

3.1 摄像机透视投影及视觉导航信息

相机坐标系中坐标为 $X_c(x_c, y_c, z_c)$ 的空间点,经线性相机透视后在图像坐标系中投影坐标为 (u, v) ,其转换关系用齐次坐标表示如下

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & r & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \mathbf{K} \mathbf{X}_c \quad (5)$$

其中, λ 为比例因子, α_x, α_y 分别是 u 轴与 v 轴的尺度因子, r 是 u 轴与 v 轴的不垂直因子,通常为 0, u_0, v_0 为光学中心。当世界坐标系与摄像机坐标系间存在旋转 $\mathbf{R}_{3 \times 3}$ 及平移 $\mathbf{t} = (t_x, t_y, t_z)^T$ 关系时,

世界坐标 $\mathbf{X}_w(x_w, y_w, z_w)$ 与摄像机坐标系的转化关系为

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \mathbf{M} \mathbf{X}_w \quad (6)$$

极线约束方程如下

$$\mathbf{p}_r^T \mathbf{K}_r^{-T} [\mathbf{t}]_{\times} \mathbf{R} \mathbf{K}_l^{-1} \mathbf{p}_l = 0 \quad (7)$$

$$[\mathbf{t}]_{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ t_x & -t_y & 0 \end{bmatrix}$$

其中, $\mathbf{p}_l, \mathbf{p}_r$ 为双目情况下左右摄像机或单目情况下不同时刻图像中匹配的点对,其齐次坐标为 $\mathbf{p}_l = [u_l, v_l, 1]^T$, $[\mathbf{t}]_{\times}$ 为平移向量的反对称阵。则点对空间约束关系表示如下:

$$\mathbf{F} = \mathbf{K}_r^{-T} [\mathbf{t}]_{\times} \mathbf{R} \mathbf{K}_l^{-1} \quad (8)$$

$$\mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R} \quad (9)$$

可见 \mathbf{E} 包含了转动与平移信息,称为本质矩阵,有 5 个自由度,秩为 2,有 2 个相等不为零奇异值。故仅依据 2 幅可匹配的图像即可获得摄像机间的相对运动信息,在相机内部参数 \mathbf{K} 已知情况下可以使用 5 点法得到 \mathbf{R}, \mathbf{t} ,从而获得相机的相对运动信息。

3.2 VO 的通用处理方法

VO 有立体、单目视觉等不同的实现方法, Scaramuzza 等^[21]对各类 VO 的基本计算过程总结如图 1 所示。

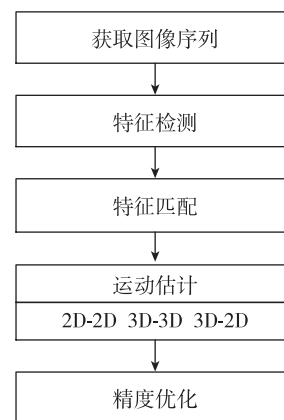


图 1 VO 的处理过程

Fig. 1 The processing of VO

稳定的图像特征可以作为图像中的地标点,从而可以利用 3.1 节中的极限约束关系得到 2 幅图像间

的相对运动信息,但成功匹配特征众多且存在误匹配情况,需要利用 RANSAC 进行误匹配剔除处理。因 VO 有基于立体、单目视觉等不同的实现方式,在相对运动估计中可使用不同计算方法,Scaramuzza 将其归纳为 2D-2D、3D-3D 以及 3D-2D 三类。

2D-2D 模式就是利用 2 幅平面图像信息进行 \mathbf{R} 、 \mathbf{t} 计算,则式(7)可化为

$$\begin{aligned} \tilde{\mathbf{p}}^T \mathbf{E} \tilde{\mathbf{p}}' &= 0 \\ \tilde{\mathbf{p}} &= \mathbf{K}_r^{-1} \mathbf{p}_r = [\tilde{u} \quad \tilde{v} \quad 1]^T \end{aligned} \quad (10)$$

$$\tilde{\mathbf{p}}' = \mathbf{K}_l^{-1} \mathbf{p}_l = [\tilde{u}' \quad \tilde{v}' \quad 1]^T$$

针对每一匹配点对,式(10)可以进行如下变形

$$[\tilde{u}\tilde{u}' \quad \tilde{u}\tilde{v}' \quad \tilde{u}' \quad \tilde{u}\tilde{v}' \quad \tilde{v}\tilde{v}' \quad \tilde{v}' \quad \tilde{u} \quad \tilde{v} \quad 1] \mathbf{E} = 0 \quad (11)$$

其中, $\mathbf{E} = [e_{11} \quad e_{12} \quad e_{13} \quad e_{21} \quad e_{22} \quad e_{23} \quad e_{31} \quad e_{32} \quad e_{33}]^T$, 依据 Nistér 5 点法,选择 5 对匹配点按照式(11)可构成求解关系,利用 SVD 法求解得到的本质矩阵,进而可以依据 \mathbf{E} 分解得到 \mathbf{R} 、 \mathbf{t} ,具体分解方法可参考文献[16]。

3D-3D 模式是利用不同时刻特征点的三维坐标测量结果进行本质矩阵恢复,进而获得 \mathbf{R} 、 \mathbf{t} 。3D-2D 模式则是将之前已经获得特征点三维坐标投影到当前二维图像中,然后进行的求解。由于 3D-2D 模式相对 3D-3D 利用了更多的直接测量结果,能够更好地减少计算误差,因此即使双目 VO 中也是更多采用的 3D-2D 方案。

对于单目 VO,无法获得准确的平移尺度,但可以利用不同时刻计算得到的 3D 特征点,通过式(12)得到相对尺度,其中, \mathbf{X}_i^k 为 k 时刻第 i 个特征点的三维坐标。

$$r = \frac{\|\mathbf{X}_i^{k-1} - \mathbf{X}_j^{k-1}\|}{\|\mathbf{X}_i^k - \mathbf{X}_j^k\|} \quad (12)$$

5 点法计算得到的相对运动信息会含有误差,必须进行误差抑制,否则会造成严重的误差累积。误差抑制的方法有多种,如 Pose Graph、RANSAC 以及 BA,其中 BA 是目前使用最多的方法。BA 利用距当前时刻一段时间内的图像特征点及相对运动结果进行平差,其基本处理思想如式(13)

$$\arg \min_{\hat{\mathbf{T}}^i, \hat{\mathbf{X}}_j} \sum_{i=1}^m \sum_{j=1}^n \|\hat{\mathbf{T}}^i \hat{\mathbf{X}}_j - \mathbf{p}_j^i\|^2 \quad (13)$$

其中, i 为图像编号, j 为特征点号, $\hat{\mathbf{T}}^i = \mathbf{K}^i [\mathbf{R}^i \quad \mathbf{t}^i]$ 为 i 图像对应的投影关系, $\hat{\mathbf{X}}_j$ 为特征点三维坐标, \mathbf{p}_j^i 为特征点 j 在 i 图像中的齐次坐标。BA 的

具体实现可以参考文献[30]。

如前所述,VO、VSLAM 同属于 SFM 方法,VO 与 VSLAM 在计算原理上是相同的,因此对 VO 不同实现方式的比较在 VSLAM 中统一进行。

4 VSLAM 的基本工作原理

按照 VSLAM 的实现方式,下面分别对滤波、BA、Pose Graph 三种方法的基本思想进行介绍,并给出比较分析。

4.1 滤波 VSLAM

滤波模型的有效工作是建立在对被估计对象变化规律、测量原理的有效描述之上的,为完成相机、地标点同时交叉定位的目的,SLAM 算法的通用滤波模型可以描述如下:

状态方程描述了相机运动及地标点位置的变化规律,如式(14)

$$\begin{bmatrix} \mathbf{X}_k^{\text{WC}} \\ \mathbf{X}_k^{\text{LM}} \end{bmatrix} = \begin{bmatrix} f(\mathbf{X}_{k-1}^{\text{WC}}, \mathbf{V}_{k-1}) \\ \mathbf{X}_{k-1}^{\text{LM}} \end{bmatrix} + \begin{bmatrix} \mathbf{W}_{k-1} \\ 0 \end{bmatrix} \quad (14)$$

其中, \mathbf{X}_k^{WC} 代表 k 时刻相机在世界坐标系中的多维位置描述向量, \mathbf{V}_{k-1} 为 $k-1$ 时刻相机的速度矢量, \mathbf{W}_{k-1} 为 $k-1$ 时刻的过程噪声, \mathbf{X}_k^{LM} 为 n 个地标点在 k 时刻位置向量, $\mathbf{X}_k^{\text{LM}} = [P_k^1 \quad P_k^2 \quad \dots \quad P_k^n]^T$,因地标点是静止的,故采用常值描述模型。

观测方程利用每个地标点在图像中的投影构建,是投影坐标与相机位置及地标点位置间的函数关系。

$$\mathbf{Z}_k = h(\mathbf{X}_k^{\text{WC}}, \mathbf{X}_k^{\text{LM}}) + n_k \quad (15)$$

实际应用中不同方法会对式(14)的状态向量有不同的选择,进而也会造成函数 f 的变化。Davison^[22]针对单目 VSLAM 首先使用了三维直角坐标与四元数相结合的七维摄像机位姿信息状态,Civera、Davison 等对地标点采用了直角坐标与角度、inverse depth 相结合的六维向量^[31],还有其他研究者使用相机姿态欧拉角描述方案,但会造成系统模型有较严重非线性。

SLAM 滤波模型是非线性的,通常采用 EKF 求解,基于 PF 的 FastSLAM 也是一种代表性方法,UKF 逐渐被研究者采纳,但更多地是应用于视觉/惯性等组合应用中^[34]。此外,目前的 VSLAM 滤波处理更多是各种标准化滤波方法的简单应用,对如何最优使用滤波器的考虑不够,未能有效发挥自适应滤波等技术对估计精度提高的潜力。

4.2 BA 优化 VSLAM

VSLAM 中的 BA 算法从原理上与 VO 中的完

全相同,仅是在具体实现中使用了不同的改进设计,在保证处理效果基础上提高 BA 处理速度是此类算法研究的重点。

1)PTAM

PTAM算法的核心内容是仅利用关键帧进行地图构建,并将地图构建与地标点跟踪在计算上进行了分离,采用2个独立的线程进行处理。在图像处理环节,针对每个关键帧都进行了4层金字塔处理,以便实现不同尺度下特征的快速搜索,采用FAST特征。在跟踪阶段,当新的图像采集完成后,使用既有相机位姿预测新位姿,以大尺度关键图粗搜索由少量特征匹配计算获得的相机初步位姿,并在此初步位姿基础上重新进行大量特征投影、匹配,进而获得高精度的位姿更新。在地图创建中,仅针对选定关键帧进行,采用了局部BA与全局BA相结合的方式进一步提高处理速度。PTAM在地图初始化阶段,利用双目视觉完成稠密3D重构,后续执行过程中可完全使用单目摄像机工作。PTAM是一种可满足实时应用的框架性方案。

2)ORB-SLAM

ORB-SLAM^[32]是建立在PTAM处理框架之上的改进实时处理算法。ORB-SLAM针对PTAM的不足进行了有针对性的改进,首先使用了具有尺度、旋转不变性的ORB特征,除满足跟踪需求外,具有场景识别能力;场景识别处理可独立工作,在跟踪异常情况下进行相机初步重定位;使用Covisibility Graph、Pose Graph进行了地图闭环处理。因ORB-SLAM强调了闭环处理,并使用了词袋进行场景识别,在KITTI测试集的封闭路段取得很好的导航效果,但在非重复路段定位效果仍有待提高。

4.3 Pose Graph 优化 VSLAM

Pose Graph优化也是SLAM研究中重要的一种优化方法,通过关键帧之间位姿关系约束进行地图优化^[33],其基本原理可用式(16)说明

$$\sum_i \sum_j \|C_i - T_{i,j} C_j\|^2 \quad (16)$$

C_i 为第*i*帧图像时摄像机在世界坐标系的位姿矩阵, $T_{i,j}$ 为*i,j*两时刻摄像机位姿矩阵间的转移矩阵,在得到各时刻摄像机位姿信息后,通过式(16)最小化可达到地图优化的目的,通常也是在最近时间窗内进行优化的。

Engle等^[19]提出的LSD-SLAM是目前使用Pose Graph较成功的算法,同时也使用了EKF、关

键帧、地图跟踪与创建分离等技术,其目的是在获得相机位姿的同时对场景进行大规模三维重构,所有工作都实时完成,该方法能够直接部署到手机上作为VO实时工作。

LSD-SLAM建立在一种Semi-dense VO^[18]方法之上,其一个重要特点是直接利用图像亮度信息进行深度计算,相对图像特征而言能够使用更多的图像信息。其三维重构是针对图像中梯度较大的区域进行,因此是一种半稠密重构。深度信息估计中使用了EKF,通过预测、校正提高深度图的准确性。

4.4 优化方法的分析

滤波VSLAM是通过滤波计算提高相机及特征点的估计精度,EKF、UKF、PF等算法都是采用连续递推方式,具有序列化合理利用图像特征空间约束的作用,但是解算会存在以下不足:滤波精度受视觉观测误差不确定性的影响明显;EKF的计算量随特征点数量的增加指数上升;缺乏利用时间窗口内多幅关键帧内图像特征进行联合约束的处理方式,在滤波器受干扰下误差的抑制能力不足。BA优化VSLAM有效利用了时间窗内观测及计算信息进行综合约束,能够达到良好的平差效果,但是计算量过大,需要采用PTAM等快速处理方法才能实现实时处理。Pose Graph优化VSLAM在优化方式上仅是使用了序列位姿关系的一致性,相对BA约束条件有所弱化,会导致优化结果最优性有所降低。

Strasdt^[21]针对滤波及BA两种方法,通过对单目、双目VSLAM的分析及大量仿真,得出了增加特征点数量而不是关键帧对精度提高最有效的结论,这种情况下BA的效率要优于滤波VSLAM。对于BA与Pose Graph,Mur-Artal^[32]在ORB-SLAM、PTAM及LSD-SLAM性能对比试验中发现从导航精度角度出发,ORB-SLAM略优于PTAM,LSD-SLAM精度较差,主要原因是LSD-SLAM采用了Pose Graph优化,从侧面反映了BA与Pose Graph的差别。

5 视觉导航技术未来发展

纯视觉导航研究取得了重要的进展,并较好地实现了精度与处理速度的大幅度提高,特别是突破了单目视觉在应用中的多种局限性,已经初步形成了理论体系并能够支撑实际应用,但在无先验知识、非往复

运动情况下,同样存在不可避免的导航误差累积问题,且成像干扰、图像中断会导致测量精度迅速下降甚至不可用。视觉导航与卫星导航、惯性导航等其他导航方式相比,体现出更多的互补性,因此组合导航仍然是最为有效的解决途径。在众多组合模式中,视觉/惯性组合具有良好的互补性、自主性,普及推广优势明显,实践证明惯性信息的引入,可以大幅度降低算法的复杂度、提高性能^[20,33,35-37]。面向应用的视觉/惯性组合有2方面的问题亟待解决:1)针对移动平台具有较高惯性测量精度的组合应用,因惯性测量对短时间内运动变化的测量精度远高于视觉,应重点研究可被短时间内视觉测量敏感的惯性误差表现形式,否则失去组合的必要性;2)针对低成本组合应用,重点解决低成本惯性器件高测量误差下的BA加速、视觉/惯性滤波组合,使紧凑、小型化嵌入式组合系统达到因时间造成的累积误差近似于零的水平,并具备可靠的模块化应用能力。

VO、VSLAM 虽然目前已经能够通过滤波、PTAM 实现实时计算,地图在向稠密构建方向发展,并已能够初步实现空间的稠密实时三维重构^[38],但对于导航应用,基于鲁棒特征的最简化稀疏地图构建、BA 快速实现等方面的研究可大幅度提高视觉及其组合导航系统的鲁棒性、可实现性,意义十分重要。此外因地标点具有绝对空间位置的不变性,存在仅利用其不变性约束、回避地标点具体位置解算的简化、高精度处理可能,该方向研究的成果可从根本上降低计算量,使视觉及视觉组合导航应用能够高可靠性地部署到低成本硬件平台上。

场景自动识别已经成为视觉导航一个重要研究方向^[39],深度学习等智能化算法在场景稠密深度图实时构造辅助下,使场景深入理解成为可能,从而可以克服目前导航地图只包含无意义的空间坐标信息的不足,赋予场景中全部或主要对象明确的语义,就可以在语义层次上实现导航,在现有导航系统能力上增加基于环境理解的宏观导航精度保障措施,是未来视觉导航技术的必然发展方向。

参考文献

- [1] 张广军. 机器视觉[M]. 北京:科学出版社,2005.
- [2] Morevec H P. Towards automatic visual obstacle avoidance[C]. International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc. 1977:584-584.
- [3] Franz M O, Mallot H A. Biomimetic robot navigation [J]. Robotics and Autonomous Systems, 2000, 30(1): 133-153.
- [4] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision [C]. International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc. 1981:674-679.
- [5] Horn B K P, Schunck B G. Determining optical flow [J]. Artificial Intelligence, 1980, 17(1-3): 185-203.
- [6] Mukawa N. Optical-model-based analysis of consecutive images [J]. Computer Vision and Image Understanding, 1997, 66(1): 25-32.
- [7] Izzo D, Weiss N, Seidl T. Constant-optic-flow lunar landing: Optimality and guidance [J]. Journal of Guidance, Control, and Dynamics, 2012, 34(5): 1383-1395.
- [8] Ohnishi N, Imiya A. Dominant plane detection from optical flow for robot navigation [J]. Pattern Recognition Letters, 2006, 27(9): 1009-1021.
- [9] Hyslop A, Humbert J S. Wide-field integration methods for autonomous navigation in 3-D environments [C]. AIAA Guidance, Navigation and Control Conference and Exhibit, 2008:1-18.
- [10] Hyslop A M, Humbert J S. Autonomous navigation in three-dimensional urban environments using wide-field integration of optic flow [J]. Journal of Guidance, Control, and Dynamics, 2010, 33(1): 147-159.
- [11] Longuet-Higgins H C. A computer algorithm for reconstructing a scene from two projections [J]. Nature, 1981, 293(5828):133-135.
- [12] Harris C G, Pike J M. 3D positional integration from image sequences [J]. Image and Vision Computing, 1988, 6(2): 87-90.
- [13] Jung S H, Taylor C J. Camera trajectory estimation using inertial sensor measurements and structure from motion results [C]. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2001:732-737.
- [14] Nützi G, Weiss S, Scaramuzza D, et al. Fusion of IMU and vision for absolute scale estimation in monocular SLAM [J]. Journal of Intelligent & Robotic Systems, 2011, 61(1): 287-299.
- [15] Scaramuzza D, Fraundorfer F. Tutorial: visual odometry [J]. IEEE Robotics and Automation Magazine, 2011, 18(4): 80-92.
- [16] Nistér D. An efficient solution to the five-point relative pose problem [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6): 756-770.

- [17] Nistér D, Naroditsky O, Bergen J. Visual odometry [C]. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2004; 652-659.
- [18] Schöps T, Engel J, Cremers D. Semi-dense visual odometry for AR on a smartphone [C]. 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)- Media, Arts, Social Science, Humanities and Design. IEEE, 2014; 145-150.
- [19] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM [C]. European Conference on Computer Vision. Springer International Publishing, 2014; 834-849.
- [20] Tardif J P, George M, Laverne M, et al. A new approach to vision-aided inertial navigation [C]// 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2010; 4161-4168.
- [21] Strasdat H, Montiel J M M, Davison A J. Real-time monocular SLAM; Why filter? [C]. 2010 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2010; 2657-2664.
- [22] Davison A J. Real-time simultaneous localisation and mapping with a single camera [C]. Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2003; 1403-1410.
- [23] Klein G, Murray D. Paralleltracking and mapping for small AR workspaces [C]. IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2007; 225 - 234.
- [24] Brox T, Bruhn A, Papenbergh N, et al. High accuracy optical flow estimation based on a theory for warping [C]. European Conference on Computer Vision. Springer Berlin Heidelberg, 2004; 25-36.
- [25] Brox T, Bregler C, Malik J. Large displacement optical flow [C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009; 41-48.
- [26] Bruhn A, Weickert J, Schnörr C, et al. Lucas/Kanade meets Horn/Schunck; Combining local and global optic flow methods [J]. International Journal of Computer Vision, 2005, 61(3): 211-231.
- [27] Chao H, Gu Y, Gross J, et al. A comparative study of optical flow and traditional sensors in UAV navigation [C]. American Control Conference (ACC). IEEE, 2013; 3858-3863.
- [28] Chao H, Gu Y, Napolitano M. A survey of optical flow techniques for robotics navigation applications [J]. Journal of Intelligent & Robotic Systems, 2014, 73(1): 361-372.
- [29] <http://www.centeye.com/technology/vision-chips/>
- [30] Hartley R, Zisserman A. Multiple view geometry in computer vision. 2nd ed [M]. Cambridge University Press, 2003.
- [31] Civera J, Davison A J, Montiel J M M. Inverse depth to depth conversion for monocular slam [C]. 2007 IEEE International Conference on Robotics and Automation. IEEE, 2007; 2778-2783.
- [32] 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.
- [33] Tiefenbacher P, Schulze T, Rigoll G. Off-the-shelf sensor integration for mono-SLAM on smart devices [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2015; 15-20.
- [34] Olson E, Leonard J, Teller S. Fast iterative alignment of pose graphs with poor initial estimates [C]. Proceedings of the 2006 IEEE International Conference on Robotics and Automation. IEEE, 2006; 2262-2269.
- [35] Zhou H, Ni K, Zhou Q, et al. An SfM algorithm with good convergence that addresses outliers for realizing mono-SLAM [J]. IEEE Transactions on Industrial Informatics, 2016, 12(2): 515-523.
- [36] Huang R, Tan P, Chen B M. Monocular vision-based autonomous navigation system on a toy quadcopter in unknown environments [C]. 2015 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2015; 1260-1269.
- [37] Indelman V, Gurfil P, Rivlin E, et al. Real-time vision-aided localization and navigation based on three-view geometry [J]. IEEE Transactions on Aerospace and Electronic Systems, 2012, 48(3): 2239-2259.
- [38] Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time [C]. IEEE International Conference on Computer Vision. IEEE Computer Society, 2011; 2320-2327.
- [39] Lowry S, Sünderhauf N, Newman P, et al. Visual-place recognition: A survey [J]. IEEE Transactions on Robotics, 2016, 32(1): 1-19.
- [40] Michael M, Sebastian T, Daphne K, et al. FastSLAM 2.0; An improved particle filtering algorithm for simultaneous localization and mapping that provably converges [C]. Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAD). 2003; 1151-1156.