doi: 10.7690/bgzdh.2023.04.018

面向动态物体场景的视觉 SLAM 方法

罗　彤[1]，骆云志[1]，沈明川[2]，陈　伟[1]
（1. 中国兵器装备集团自动化研究所有限公司信控中心，四川 绵阳 621000；
2. 陆装驻成都地区军代室，成都 611930）

摘要：针对动态场景移动物体的运动估计问题，对视觉同步定位与地图构建(simultaneous localization and mapping，SLAM)方法进行分析。现有动态 SLAM 算法可分为基于几何的方法和基于深度学习的方法 2 类，重点介绍一种基于几何方法——基于点云相关性的动态 SLAM 方法；总结目前 SLAM 技术面临的挑战，展望动态 SLAM 针对未来战争的发展潜力与发展方向。结果表明，该研究可促进 SLAM 技术在机器人导航中的应用。

关键词：动态场景；动态 SLAM；点云相关性
中图分类号：TP242.6^{+}2　　**文献标志码：**A

Visual SLAM Method for Dynamic Object Scenes

Luo Tong[1], Luo Yunzhi[1], Shen Mingchuan[2], Chen Wei[1]
(1. *Weapon Equipment Information and Control Technology Innovation Center*, *Automation Research Institute Co*., *Ltd. of China South Industries Group Corporation*, *Mianyang* 621000, *China*;
2. *Military Representative Office in Chengdu*, *Army Equipment Department*, *Chengdu* 611930, *China*)

Abstract: Aiming at the problem of motion estimation of moving objects in dynamic scenes, the simultaneous localization and mapping (SLAM) method is analyzed. The existing dynamic SLAM algorithms can be divided into 2 categories: the geometry-based method and the deep learning-based method. This paper focuses on a geometry-based method, the dynamic SLAM method based on the correlation of point clouds, summarizes the challenges faced by the current SLAM technology, and looks forward to the development potential and direction of the dynamic SLAM for future warfare. The results show that the research can promote the application of SLAM technology in robot navigation.

Keywords: dynamic scenes; dynamic SLAM; point cloud correlations

0　引言

近年来，因为摄像机的低成本和低质量，基于视觉的运动估计方法，包括视觉里程计(visual odometry，VO)[1]和视觉同步定位与映射(visual SLAM，VSLAM)[2-3]，在机器人导航中发挥了重要作用。上述方法仅需输入图像提供六自由度就能进行运动估计，但其对静态世界的假设严重限制了适用情景。在现实中，基于静态世界假设的方法会受到视场(field of view，FOV)中出现的运动物体影响，甚至失败。包含移动物体的场景称为动态环境。根据移动物体所占据的 FOV 面积，动态环境可分为轻微动态环境和高度动态环境。由于在微动态环境中，视场只有一小部分被运动目标覆盖，传统的鲁棒估计方法如随机样本一致性(random sample consistency，RANSAC)[4]方法和鲁棒加权函数[5-6]可消除运动目标的大部分影响。相反，如果视场的大部分被运动目标覆盖，则运动目标的观测量多于静态场景的观测量，将导致鲁棒估计方法失败。传统的 VO 和 VSLAM 假设世界是静态的，在实际应用中应用有限，如何消除这些运动物体的影响成为一个重要课题。

学者利用几何信息和语义信息等多种类型的先验信息，解决动态环境下移动物体引起的运动估计失效问题。几何信息，即利用同一个刚体运动物体上的映射点独立于静态场景的运动一致性。S. Lee 等[7]提出了《动态环境下基于刚体运动模型的鲁棒实时 RGB-D 视觉里程计》，利用运动一致性来确定运动对象上的点。现实世界也包括非刚性的物体，不是任何移动对象上的所有映射点都能用这种方法确定。除了几何信息外，基于深度学习的方法(如动态 SLAM：在动态场景中的追踪、建图和修复[8])从训练集中学习语义信息作为先验信息，直接在图像上从静态场景中分离出可疑的运动物体。尽管这些方法可提供更精确的结果，但不能识别不存在于训练集中的未知对象，而且计算资源消耗也非常巨

收稿日期：2022-12-30；**修回日期：**2023-01-29
作者简介：罗　彤(1996—)，男，湖北人，工程师，从事视觉 SLAM 和点云相关性研究。E-mail: a200809102007@qq.com。

大，通常很难在嵌入式环境中实时运行；因此，在未知动态环境下，VSLAM 和 VO 能否提供健壮、准确的导航信息仍是一个疑问。

上述方法尽管可在静态环境中提供优异的性能，但目前基于视觉的运动估计方法在环境太具挑战性时(例如在高度动态环境中)往往会失败[9-10]。现有的鲁棒估计方法仅解决了轻度动态环境中运动物体的部分干扰。科学家们提出了许多方法来处理中度和高度动态环境中的运动物体干扰，可分为基于几何的方法和基于深度学习的方法 2 种主要类型。笔者重点对基于几何的方法进行介绍。

1 基于几何的方法

最常用的几何策略是将动态运动元素视为必须区分和消除的噪声。基于几何的方法不需要运动对象的先验信息，比基于学习的方法处理速度更快。Zhang 等[11]提出了 FlowFusion，这是一种使用基于学习的网络来获得每帧预测光流的系统，能迅速将动态对象的区域分割出来，并完成静态背景的重构。Cheng 等[12]提出了一种基于光流的方法，该方法使用光流从使用 RGB 图像作为唯一输入提取的动态特征点中区分和消除动态特征点。静态融合[13]利用图像的概率分割来重建背景，并将其集成到加权密集优化框架中。Li 等[14]提出了关键帧中边缘点的静态加权方法，该系统显著减少了动态对象的干扰。Dai 等[15]基于静态对象应随时间呈现连续姿态变换的事实，利用点相关性来分离静态和动态地图点。

在基于几何的 SLAM 方法中，所有的补偿估计都是标准视觉 SLAM 方案中的一部分，在执行分割时不存在冗余的计算负担；因此，基于几何的方法实时性较好。基于几何的 SLAM 方法通过物体的运动来确定分割，并由高几何残差来表示。在这种情况下，无法处理移动物体暂时静止的问题，而且会混淆由运动物体引起的误差和错误匹配引起的误差。这些问题都需要后继者解决。虽然基于几何的 SLAM 方法能够在一定程度上提高 SLAM 系统的鲁棒性，但与基于深度学习的 SLAM 方法相比，它在位姿估计的精度上有所欠缺。

2 基于学习的方法

基于深度学习的点云分割方法大体上被分为基于多视图、基于体素、基于树和直接对点进行处理 4 类。

基于多视图的方法，在三位分析中最具影响力的是 MVCNN[16]。因为图像领域可通过渲染得到 3D 模型 12 个角度的图像，MVCNN 对这种图像领域进行分类、检测。这使得点云数据的结构化问题得到解决，但还有 2 个缺陷：1) 由于 2 维多视图图像只是 3 维场景的近似值，它们存在的局限性会导致几何结构的损失；2) 在大型、复杂环境中，很难为多视图投影选择充足、有效的视点。

体素化能够有效地处理原始点云的无序和非构造化问题，其中数据经过 3 维卷积进行进一步处理；但这种体系在效率上有着严重缺陷，该方法如果要保留足够的点的空间信息，那么点云数据量会变得非常庞大，算力开销很大。如果要保证效率，就必须减少分辨率及位姿估计的精度。

该方法主要处理原始点云的非结构化问题，多用于 3D 模型识别任务。因为该方法不使用 2 维或 3 维网格，减少了不良的缩放行为；但该方法不仅使数据丧失了深度信息，还造成了更高的计算成本。

PointNet[17]直接对点云数据进行处理，使用对称函数来解决点云的顺序问题，但它没有考虑到相邻点云间的局部信息。Qi 等[18]提出了 PointNet++，设计了层级结构 set abstraction 来捕获局部特征信息，并提出了 MSG/MRG 结构，适应非均匀分布的点云数据。

综上所述，大多数基于深度学习的点云方法因为依赖于深层神经网络的处理，都要将点云转换成结构化网格。这会丢失点云的深度信息，也会提高计算成本；因此，对点云数据进行直接处理的方法成为主要研究方向。

同时，虽然基于深度学习的 SLAM 方法在动态环境中有着优越的鲁棒性，但也有其限制，即不能识别不存在于训练集中的物体，当场景连续变化时，很难获得物体运动的先验信息；因此，在缺少先验信息的情况下，如何准确地估计动态环境下移动物体的运动状态，依然是一个巨大的挑战。

3 基于点云相关性的方法

3.1 基于点云相关性改进 BA 优化

Dai 等提出的动态环境下基于点相关性的 RGB-D SLAM 方法在传统的 SLAM 方法中加以改进。传统的 SLAM 方法中，集束调整(Bundle Adjustment)优化的因子图(图 1)中，静态点和动态点之间不存在约束条件。其补偿函数为：

$$J_{ba}(x)=\frac{1}{2}\sum_{i,k}e_{y,ik}(x)^{\mathrm{T}}C_{ik}^{-1}e_{y,ik}(x)。\tag{1}$$

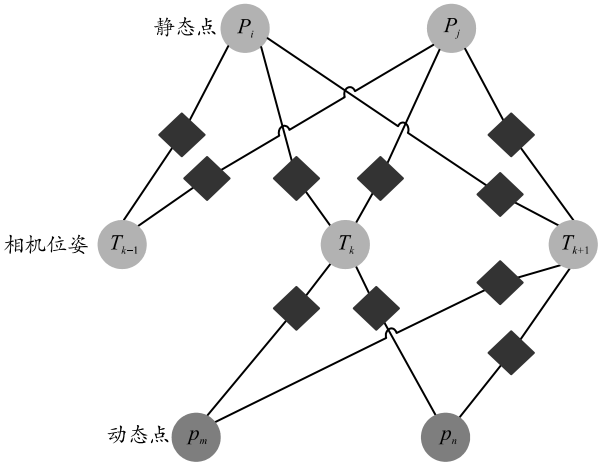
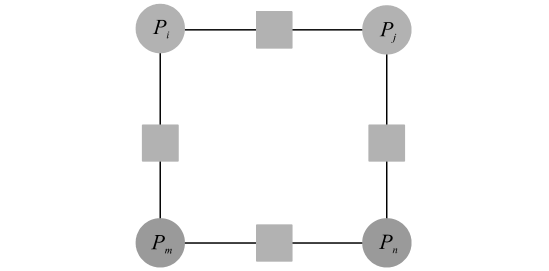


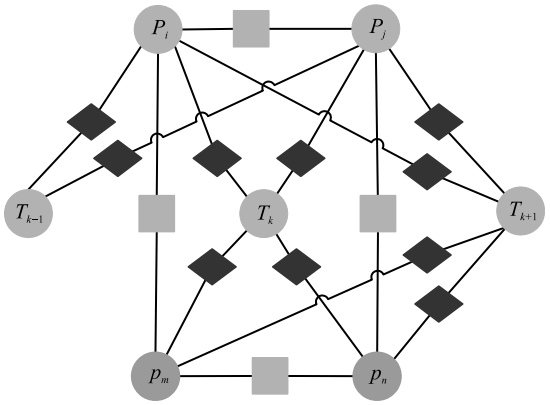
图 1　集束调整(BA)优化的因子图

图中：T_k 表示 k 时刻相机的位姿；P_i 表示第 i 个静态地图点的位置变量；P_m 表示第 m 个动态地图点的位置变量；菱形表示相机与地图点之间的约束。

BA 的计算过程中，基于点相关性的方法在前端的集束调整(bundle adjustment)优化过程中加入地图点之间的刚性约束，从而构建新的因子图如图 2 所示。



(a) 地图点之间刚性约束的因子图



(b) 传统 SLAM 中 BA 优化的因子图加入点间刚性约束后的新的因子图

图 2　因子图

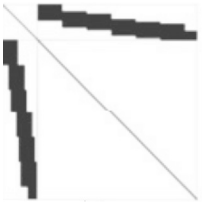
点间的刚性约束的补偿函数为：

$$J_p(x_g)=\frac{1}{2}\sum_{ij,k}e_{z,ijk}(l_{ijk})^{\mathrm{T}}C_{ijk}^{-1}e_{z,ijk}(l_{ijk})。\tag{2}$$

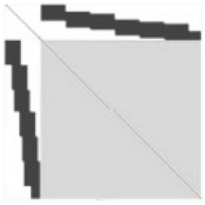
则新的因子图的补偿函数则变为：

$$J(x)=J_p(x_g)+J_{ba}(x)。\tag{3}$$

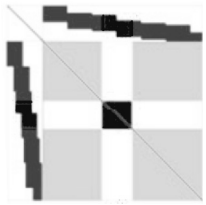
在新构建的因子图下，海森矩阵也发生了变化。其变化如图 3 所示。



(a) 传统 BA 优化的 Hessian 结构



(b) 目标函数的 Hessian 结构



(c) 剔除不一致边后的 Hessian 结构

图 3　黑森结构的示例。

图中：浅灰色块表示点相关性，深灰色块表示几何体-姿态相关性，黑色块表示相机图像中的移动对象。

从上图可以看到，当动态场景中的运动物体发生移动时，静态点和动态点之间的相对位置发生改变，动静点间刚性约束的误差较大。在优化时，剔除误差较大的点检刚性约束，就能分离动态点与静态点之间的相关性评估结果，完成分割。

3.2　使用德劳奈三角剖分降低计算复杂性

如果所有的点和点之间都要去计算它们的相关性，会是一个非常复杂的计算过程。Dai 等引入德劳奈(Delaunay)三角剖分来完成工程化修改。作为一种十分重要的预处理技术，德劳奈三角剖分并不是一种算法，只是给出一个“好的”三角网格定义。

在进行德劳奈三角剖分时，会最大化生成所有三角形的内角，避免尖锐三角形的产生。这能保证点云中一定是最近的三个点形成三角形，三角形上连接边的两点一定是近邻点，确保地图点之间的相关性有意义。在进行 BA 优化时，计算生成三角形边上两点间的相关性即可。

使用德劳奈三角剖分能够在保证点间相关性精

度的同时，尽可能地减少点间相关性的计算量。

这种基于点间相关性的分割方法提高了 SLAM 系统在动态环境中的鲁棒性和精确性；同时，它也具有一般性，能够适用于广义视觉传感器。

4 结束语

未来，在基于几何方法的 SLAM 方法研究中，还可扩展到单目、立体和光探测及测距系统中的传感器。由于不同传感器的噪声模型有所不同，而准确的噪声模型在保证传感器的有效信息方面发挥着至关重要的作用；因此，在多传感融合的 SLAM 方面还有很多工作要做。此外，还可进一步提高该方法的性能。例如，可增量实现图构造，以避免重复创建图并降低复杂性。

综上所述，将基于深度学习的计算机视觉系统、自然语言处理、语义知识库与基于几何的推理系统相结合的 SLAM 方法不仅可行，而且将成为今后的发展趋势。目前，由于所必需的工具大多是独立开发的，集成和接口规范将成为一个更加突出的主题，这方面的研究进展可促进机器人在 SLAM 系统内实现环境推理。为实现自动驾驶汽车、协作机器人和服务机器人安全有效的导航和工作，只识别动态物体是不够的。如果机器人在 SLAM 系统中能够对场景中的动态物体(尤其是人类)进行推理并预测其行为或意图，将推动机器人技术进步；因此，将几何方法与深度学习相结合，研究混合几何信息与语义信息的 SLAM 方法将是未来的研究趋势。

参考文献：

[1] SCARAMUZZA D, FRAUNDORFER F. Visual Odometry [Tutorial][J]. Robotics & Automation Magazine, IEEE, 2011, 18(4): 80−92.

[2] FUENTESP J, RUIZA J, RENDÓNM J M. Visual Simultaneous Localization and Mapping: A Survey[J]. Artificial Intelligence Review, 2015, 43(1): 55−81.

[3] THRUN S, LEONARD J J. Simultaneous Localization and Mapping[J]. Springer Handbook of Robotics, 2008, 10: 871−889.

[4] FISCHLER M A, BOLLES R C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography-ScienceDirect[J]. Readings in Computer Vision, 1987, 24(6): 726−740.

[5] KERL C, STURM J, CREMERS D. Robust Odometry Estimation for RGB-D Cameras[C]//Robotics and Automation (ICRA), 2013 IEEE International Conference on. IEEE, 2013.

[6] MACTAVISH K, BARFOOT T D. At all Costs: A Comparison of Robust Cost Functions for Camera Correspondence Outliers[C]//Conference on Computer & Robot Vision. IEEE Computer Society, 2015: 62−69.

[7] LEE S, SON C Y, KIM H J. Robust Real-time RGB-D Visual Odometry in Dynamic Environments via Rigid Motion Model[J]. arXiv, 2019, 1907: 08388v1.

[8] PALAZZOLO E, BEHLEY J, LOTTES P, et al. ReFusion: 3D Reconstruction in Dynamic Environments for RGB-D Cameras Exploiting Residuals[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS). IEEE, 2019.

[9] CADENA C, CARLONE L, CARRILLO H, et al. Past, Present, and Future of Simultaneous Localization and Mapping[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309−1332.

[10] SAPUTRA, Muhamad R U, MARKHAM, et al. Visual SLAM and Structure from Motion in Dynamic Environments: A Survey[J]. ACM computing surveys, 2018, 51(2): 1−36.

[11] ZHANG T, ZHANG H, LI Y, et al. FlowFusion: Dynamic Dense RGB-D SLAM Based on Optical Flow[C]//International Conference on Robotics and Automation. IEEE, 2020.

[12] CHENG J Y, SUN Y X, MENG Q H. Improving monocular visual SLAM in dynamic environments: an optical-flow-based approach[J]. Advanced Robotics, 2019, 33(12): 576−589.

[13] SCONA R, JAIMEZ M, PETILLOT Y R, et al. StaticFusion: Background Reconstruction for Dense RGB-D SLAM in Dynamic Environments[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.

[14] LI S, LEE D. RGB-D SLAM in Dynamic Environments using Static Point Weighting[J]. IEEE Robotics & Automation Letters, 2017, 2(4): 2263−2270.

[15] DAI W, ZHANG Y, LI P, et al. RGB-D SLAM in Dynamic Environments Using Point Correlations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 373−389.

[16] GE L, HUI L, YUAN J, et al. Robust 3D Hand Pose Estimation in Single Depth Images: From Single-View CNN to Multi-View CNNs[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016.

[17] QI C R, SU H, MO K, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.

[18] QI C R, LI Y, HAO S, et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS, 2017.