

一种改进的语义视觉里程计优化算法

何 杰

北京邮电大学, 计算机学院, 北京

收稿日期: 2022年3月1日; 录用日期: 2022年3月30日; 发布日期: 2022年4月6日

摘 要

为了优化移动机器人的视觉SLAM过程中的里程计估算问题, 利用视觉图像的语义信息优化使得机器人在进行地图构建过程中的位姿估计更为准确, 本文提出了一种改进的视觉里程计优化算法。该算法基于语义视觉里程计算法, 提出了一种依赖语义特征不变性对视觉里程计进行约束优化的算法。通过构建由语义分割网络检测结果的语义重投影误差损失函数, 并对此函数进行优化求解, 从而对视觉SLAM中的里程计实现更高精度的优化。经过实验, 结合文中所提出的里程计优化算法的视觉SLAM算法在数据集和实际环境中的里程计误差均有优化。

关键词

移动机器人, 同步定位与建图, 语义分割

An Improved Semantic Visual Odometry Optimization Algorithm

Jie He

School of Computer, Beijing University of Posts and Telecommunications, Beijing

Received: Mar. 1st, 2022; accepted: Mar. 30th, 2022; published: Apr. 6th, 2022

Abstract

In order to optimize the odometry estimation problem in the visual SLAM process of mobile robots, the semantic information optimization of visual images makes the robot's pose estimation in the process of map construction more accurate. In this paper, an improved visual odometry optimization algorithm is proposed. Based on the semantic visual odometry algorithm, this algorithm proposes a constrained optimization algorithm for visual odometry relying on the invariance of semantic features. By constructing the semantic reprojection error loss function detected by the se-

mantic segmentation network, and optimizing this function, the odometer in visual SLAM can be optimized with higher precision. After experiments, the visual SLAM algorithm combined with the odometer optimization algorithm proposed in this paper has optimized the odometer errors in both the dataset and the actual environment.

Keywords

Mobile Robot, Simultaneous Localization and Mapping, Semantic Segmentation

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

SLAM (simultaneous localization and mapping, 同步定位与建图)技术指的是通过传感器感知周边环境并通过估计传感器位置的变化来构建环境地图的技术。移动机器人的定位与导航,无人驾驶中的记忆泊车等功能均依赖于 SLAM 技术。视觉 SLAM 技术指的就是以视觉传感器作为主要数据输入的 SLAM 技术,如单目、双目、深度摄像机等。

视觉 SLAM 中,用于获取传感器位置变换的部分一般为视觉里程计(也被称为前端)和后端优化部分。其中,前端通过图像特征在单帧或多帧之间的变化来初步估算出相机的位姿变化。而后端优化部分则使用多帧之间的变化进行位姿约束优化,从而得到一个相对精确的位置变化信息[1]。

一般用于维持后端优化进行里程计优化的依据为图像中的不变特征点,而这些特征点的提取和匹配均依赖于传统的图像处理算法,如像素亮度对比等。随着深度学习在图像处理算法方面的应用,现有的算法已经可以通过先验模型对图像中的像素点做出更为准确的界定,且这些像素点在不同的图片中均能保持同类界定,如能够将图片中的像素点均做出分类的语义分割模型。

因此,如何利用语义分割网络,在多帧图片之间维持对特定像素点的同一分类检测结果,并以此为依据提供对相机位置变换的约束优化,从而与现有的里程计优化算法结合,为现有的视觉 SLAM 提供更为准确的里程计优化结果,将会是本文的主要研究内容。

2. 相关工作

视觉 SLAM 算法按照传感器的不同组合分为了不同的发展方向,如单目、双目以及多传感器融合等。而在语义 SLAM 算法方面,其发展方向主要由语义信息与传统 SLAM 的结合点来进行区分,一般主要集中于语义地图构建以及语义视觉里程计方面。

2.1. 语义 SLAM

语义 SLAM 指的是通过在传统的 SLAM 系统中添加语义信息从而实现优化现有的 SLAM 流程或者是添加额外功能的技术。早期的研究者们认为语义 SLAM 即是在对物体进行三维重构时精确恢复出物体的细节及语义。如 Valentin 等[2]提出了使用 CRF (Conditional random field, 条件随机场)为物体添加语义信息的方式。此后的研究者提出依靠语义信息以辅助 SLAM 系统中的一些模块,如 Zhao 等[3]使用了语义信息优化了 SLAM 算法中的视觉里程计。构建语义地图同样也是语义 SLAM 的研究目标之一,如

McCormac 等[4]提出的结合 CNN 以及条件随机场的语义地图构建算法。

2.2. 语义视觉里程计

本文的主要研究内容为语义 SLAM 中的语义里程计优化部分。关于此部分的研究内容,典型的为 Kevin [5]等提出的使用 DPM (Deformable Parts Model, 可形变部件模型)结合传统 SLAM 做概率融合从而优化视觉 SLAM。此后, Lianos 等[6]提出的 VSO 则是使用距离变换将分割结果的边缘作为约束从而优化视觉里程计。而 Li 等[7]提出基于对目标的检测结果构建和地图构建模块融合的优化项,从而构建一个整体的优化问题。而 Stenborg 等[8]提出了依靠语义检测结果在现有的 SLAM 算法中做定位优化的思路。

3. 语义优化算法研究

在视觉 SLAM 算法中,进行里程计估计优化的依据一般为图像中的通用特征,一般为点或线等特征。通过匹配这些特征进行计算和优化。而在引入了语义分割网络后,图像在通用特征点之外还包括了语义网络检测出的像素语义信息,在进行里程计的计算与优化时,同样也可以以此为依据对算法估计出的里程计进行优化。在进行里程计的初步计算时,由于实时性和计算效率的要求,此部分并不适合加入语义分割的检测结果。而里程计的优化一般为一个频繁的迭代优化过程,其中主要的计算消耗为迭代优化部分,且执行的频率相对较低,因此在此部分引入语义分割结果对整个算法效率造成的影响较小。

3.1. 累计误差消除算法

在里程计的优化部分主要做的工作为对初步的里程计信息进行误差优化,即对于估计过程产生的累计误差进行消除。通过对一般视觉里程计的原理进行分析,无论是基于特征点法还是基于直接法,由于光照和相机角度的影响,初步进行匹配计算出的里程计信息均存在一定的误差。而随着相机的运动逐渐叠加,这种细微误差会逐渐累计,大量的累计误差积累将会使得整个算法的整体误差逐渐扩大,最终使得 SLAM 算法完全不能估计出相机所处的位置,对地图的构建与定位造成巨大的影响。

通用的累计误差消除算法为最小化重投影误差算法。其核心思想为:在相机的不同状态,对于空间中的坐标点观测信息经过转换后应能保证基本重合(误差处于可接收范围内)。即在估计出的相机运行信息下,通过计算点在当前图片中的位置与变换后图片上位置的差值并进行优化从而得出较为精确的里程计信息。对于在两张不同图片上的同一点的确定同样依赖于对图片中的通用特征进行提取。

3.2. 语义代价函数

在一般的视觉 SLAM 算法中,对于里程计的优化问题一般可以概括为一个非线性优化问题,而为了进行优化,则需要定义一个待优化求解的代价函数,该函数需要能正确衡量误差与估计值的关系。对于语义优化约束来说,本文采取类似于最小化重投影误差的方法,与其不同的是,进行误差比较的依据为图片中的语义分类信息。

$$p(S_k | T_k, X_i, Z_i = c) \quad (1)$$

如公式(1)所示,将点投影到图像 S_k 中,该点类别为 c 的概率,而 T_k 和 X_i 分别表示里程计以及点的坐标。如果类别相同,则概率较大,如果类别不同,定义公式(2)表示概率与最近的类别为 c 的点之间距离的关系, DT 用来表示这个距离。

$$p(S_k | T_k, X_i, Z_i = c) \propto e^{-\frac{1}{2\sigma^2} DT_k^{(c)}(\pi(T_k, X_i))^2} \quad (2)$$

因此,可以用公式(3)表示语义代价函数。其中 $\omega_i^{(c)}$ 表示点是类别 c 的概率。

$$\begin{aligned}
e_s(k, i) &= \sum_{c \in C} \omega_i^{(c)} \log(p(S_k | T_k, X_i, Z_i = c)) \\
&= - \sum_{c \in C} \omega_i^{(c)} \cdot \frac{1}{\sigma^2} DT_k^{(c)}(\pi(T_k, X_i))^2
\end{aligned} \tag{3}$$

以上公式基于语义网络对目标的边界检测较为准确的情况，但是在实际环境中，由于语义分割网络对目标进行分割的不确定性，并不能完全依靠边界信息作为多帧之间不变性的依据，因此，引入其他结果对其进行约束是很有必要的。不难发现，语义网络对于边界的检测波动较大，但是对于整个目标位置与轮廓的确认则较为精确，因此可以将其作为一个约束项加入到代价函数中。定义点在图片帧中与语义网络检测类别为 c 的结果质心的距离如公式(4)所示。

$$M = \text{Center}(X_i, Z_i = c) \tag{4}$$

结合公式(3)和(4)，最终得出的语义代价函数如公式(5)所示，其中 λ 为超参数，根据目标的多少适当调整质心距离对语义代价函数的影响。

$$e_s(k, i) = - \sum_{c \in C} \omega_i^{(c)} \cdot \frac{1}{\sigma^2} DT_k^{(c)}(\pi(T_k, X_i))^2 \cdot \frac{\lambda}{(M_o - M_k)^2} \tag{5}$$

3.3. 优化与求解

公在定义了待求解的代价函数后，需要对函数进行优化求解从而优化初步的里程计信息。对于类似的复杂代价函数，一般使用 EM (Expectation Maximization, 期望最大化) 的算法进行优化[9]。即根据观测的数据估计初步参数，再根据估计出的参数计算未观测数据。之后再次进行估计，逐步迭代至算法收敛。

对于本文中的代价函数，按照 E 和 M 两个步骤去求解。在 E 步骤中，对每个地图点求解权重向量，固定点的位置和里程计。而在 M 步骤中，保持权重不变，优化点的位置和里程计。通过这种方式逐步完成对代价函数的优化。

4. 语义优化算法实现与实验

4.1. 语义约束优化算法实现

语义约束优化算法需要与现有的视觉 SLAM 进行结合，通过同时考虑重投影误差[10]以及语义误差，最终构建出对于里程计信息的联合优化。算法流程的伪代码如图 1 所示。

Algorithm 1 Optimizer

Input: $P = (P_i, i = 0, \dots, k - 1), T, \text{maxcount}, \text{Sem}$

```

1: while maxcount > 0 do
2:   for  $i = 0 \rightarrow k - 1$  do
3:      $E(P, T, \text{Sem}) += (e_{\text{base}}^i(P_i, T) + e_{\text{Sem}}^i(P_i, T, \text{Sem}))$   计算累计误差
4:   end for
5:   if  $E \leq \text{min}E$  then
6:     maxcount = 0  误差满足要求，停止优化
7:   else
8:      $T = LM(E, P, T)$   优化求解
9:     maxcount - = 1
10:  end if
11: end while

```

Figure 1. Pseudocode of semantic constraint optimization process

图 1. 语义约束优化流程伪代码

对于每一次约束优化, 传入的信息为确定进行优化的点的位置 P , 此刻的里程计数据 T , 最大迭代次数 $maxcount$ 。算法首先根据公式(5)构建待求解的累计误差, 此后在步骤 8 使用 LM (Levenberg-Marquardt, 列文伯格 - 马夸尔特) 算法进行求解[11], 并进行计数, 知道误差满足条件或者迭代次数达到最大, 则结束优化, 并使用优化后的结果更新里程计数据。

4.2. 语义约束优化算法实验

为了验证算法的优化效果, 本文使用基于 ROS (Robot Operating System, 机器人操作系统) 的机器人分别在数据集和实际环境中进行算法的实验。该机器人搭载深度摄像机并运行于室内环境。数据集的测试使用 TUM [12] 数据集, 该数据集包含使用高精度传感器记录的里程计信息。而在实际环境中的测试使用机器人的编码器和 IMU 的数据融合作为数据参照。算法的语义分割输入依赖实时的 RGBD 语义分割网络: ESANet [13]。机器人分别使用 ORBSLAM3 [14]、ORBSLAM3 + VSO 以及基于本文优化算法的 ORBSLAM3 进行里程计估计。使用相对姿态误差作为衡量的大小最终得出的实验结果如表 1 所示。

Table 1. Experiment result

表 1. 实验结果

实验对象	TUM	实际环境
ORBSLAM3	1.95 m	3.67 m
ORB SLAM3 + VSO	1.84 m	3.25 m
ORBSLAM3 + 本文语义优化	1.77 m	3.22 m
VSO 优化占比	5.64%	11.44%
本文语义优化占比	9.23%	12.26%

5. 结论

本文研究了视觉 SLAM 中的常见问题: 即视觉 SLAM 中的视觉里程计以及后端优化问题。文中介绍了语义 SLAM 以及语义视觉里程计的发展现状, 并分析了语义里程计进行优化约束的必要条件。通过引入语义分割网络, 本文在语义视觉里程计的基础上, 提出了一种改进的语义视觉里程计优化算法。该算法通过构建语义约束代价函数, 并对此函数进行优化求解, 从而辅助优化原有视觉 SLAM 算法中的里程计。经过实际实现和测试, 证明了该算法可以与现有的 SLAM 算法结合并进行优化, 优化结果相较于作为基础的语义视觉里程计, 其对于里程计误差的优化占比拥有了 1%~2% 的提升。

致 谢

在完成本论文的过程中, 很多人在不同层面上给予了很多帮助。首先要感谢我的导师在研究期间的指导。同时, 感谢实验室其他小伙伴的交流与启发, 正是因为你们的帮助我才能完成本论文。同时, 也对参与论文评审的专家老师和参与排版的编辑们表示感谢。

参考文献

- [1] 赵洋, 刘国良, 田国会, 等. 基于深度学习的视觉 SLAM 综述[J]. 机器人, 2017, 39(6): 889-896.
- [2] Sengupta, S., Valentin, J., Warrell, J., et al. (2013) Mesh Based Semantic Modelling for Indoor and Outdoor Scenes. *CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Portland, OR, 23-28 June 2013, 2067-2074. <https://doi.org/10.1109/CVPR.2013.269>
- [3] Zhao, Z. and Chen, X. (2016) Building 3D Semantic Maps for Mobile Robots Using RGB-D Camera. *Intelligent Ser-*

- vice Robotics*, **9**, 297-309. <https://doi.org/10.1007/s11370-016-0201-x>
- [4] McCormac, J., Handa, A., Davison, A., *et al.* (2017) SemanticFusion: Dense 3D Semantic Mapping with Convolutional Neural Networks. *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, 29 May-3 June 2017, 4628-4635. <https://doi.org/10.1109/ICRA.2017.7989538>
- [5] Bowman, S.L., Atanasov, N., Daniilidis, K., *et al.* (2017) Probabilistic Data Association for Semantic SLAM. 2017 *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, 29 May-3 June 2017, 1722-1729. <https://doi.org/10.1109/ICRA.2017.7989203>
- [6] Lianos, K.N., Schnberger, J.L., Pollefeys, M., *et al.* (2018) VSO: Visual Semantic Odometry. *European Conference on Computer Vision (ECCV)*, **11208**, 246-263. https://doi.org/10.1007/978-3-030-01225-0_15
- [7] Li, P., Qin, T. and Shen, S. (2018) Stereo Vision-Based Semantic 3D Object and Ego-Motion Tracking for Autonomous Driving. *European Conference on Computer Vision (ECCV)*, **11206**, 664-679. https://doi.org/10.1007/978-3-030-01216-8_40
- [8] Stenborg, E., Toft, C. and Hammarstrand, L. (2018) Long-Term Visual Localization Using Semantically Segmented Images. 2018 *IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, 21-25 May 2018, 6484-6490. <https://doi.org/10.1109/ICRA.2018.8463150>
- [9] 陈昱皓, 彭道刚, 王志萍, 等. 基于 EM-ORB 算法的移动机器人 SLAM 系统研究[J]. 电气传动, 2020, 50(5): 67-74.
- [10] 周单, 董秀成, 张帆, 等. 基于自适应重投影误差单目位姿优化算法[J]. 激光与光电子学进展, 2019, 56(2): 111-118.
- [11] 白方. 基于约束最优化的 SLAM 问题的解法和非线性分析[D]: [硕士学位论文]. 沈阳: 东北大学, 2019.
- [12] Schubert, D., Goll, T., Demmel, N., *et al.* (2018) The TUM VI Benchmark for Evaluating Visual-Inertial Odometry. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, 1-5 October 2018, 1680-1687. <https://doi.org/10.1109/IROS.2018.8593419>
- [13] Seichter, D., Khler, M., Lewandowski, B., *et al.* (2020) Efficient RGB-D Semantic Segmentation for Indoor Scene Analysis. 2021 *IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, 30 May-5 June 2021, 13525-13531. <https://doi.org/10.1109/ICRA48506.2021.9561675>
- [14] Campos, C., Elvira, R., Rodríguez, J.J.G., *et al.* (2020) ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM. *IEEE Transactions on Robotics*, **37**, 1874-1890. <https://doi.org/10.1109/TRO.2021.3075644>