
论文阅读-《BAD SLAM: Bundle Adjusted Direct RGB-D SLAM》 (2019CVPR, ETH)-No.1

Author of This Article Analysis: 魏鑫燊; Time: 2019-07-10

Open Source

1. 数据集
2. BAD SLAM system

Motivation

1. 目前 SLAM 领域中直接法的位姿精度大都比 ORB-SLAM2（特征法）差，作者认为这是由于直接法容易被目前数据集中的几个因素影响：卷帘快门（rolling shutter）、图像和深度信息的同步问题、深度信息中的误差。这些问题从硬件层面更易于解决，所以作者提出了一个新的数据集，概括来说，作者认为 ORB-SLAM2 并不真的就比直接法好，而是目前的数据集存在一些问题；
2. 目前认为 Full BA 用到直接法中是不能做到实时的，已有的系统使用了一些近似的手段来处理这一问题（比如基于位姿图的 BA 优化），作者提出了一个新的用于直接法的 full BA，在单个 GPU 上达到了实时运行。

Contribution

1. 一个新的应用于直接法中的快速 BA 算法，在新的数据集上达到了同类系统最好效果；
2. 一个新的数据集，使用了同步全局快门相机（synchronized global shutter cameras）；
3. 给出了基于该数据集的在线排行榜，同时开源了提出的 BAD SLAM 源码。

Content

3D 环境表示方法

Surfel: 一个有向的圆（an oriented disc），用于表示环境地图，图像上每 4×4 个像素点的区域可生成一个 surfel。包括：一个三维中心点的坐标 P_s ，一个平面法向量 n_s ，一个半径 r_s ，一个视觉描述子（标量） d_s 。

相比于 mesh(from 《Photometric bundle adjustment for dense multi-view 3D modeling》CVPR2014): Surfel 在 BA 过程中不存在额外的拓扑更新；相比于 Voxel(from 《KinectFusion: Real-time dense surface mapping and tracking》2011): surfel 能够在任意尺度上表示很小的平面（这部分还并不是很懂）。

特征匹配

surfel 与图像中像素点的匹配过程仅使用了 surfel 的中心点和平面法向量，当以下条件满足时匹配成功：

1. surfel 中心点的投影点处的像素点有深度值（深度相机在一些尖角处常常没有深度输出）；
2. 像素点深度与 surfel 中心点深度足够相似；
3. surfel 平面法向量与深度图上得到的法向量（这里使用有限差分来计算以像素点为中心的的法向量），以及像素点与相机中心的方向足够相似。

目标函数

surfel 的优化存在病态（ill-conditioned）问题，当视觉描述子和几何信息变化不明显时（比如一面白墙），即输入数据的约束能力不足时，surfel 会优化到错误的数值，解决方案是限制优化方向（仅沿平面法向量方向优化），优化的目标函数为：

$$C(K, S) = \sum_{k \in K} \sum_{s \in S_k} (\rho_{Tukey}(\sigma_D^{-1} r_{geom}(s, k)) + \omega_{photo} \rho_{Huber}(\sigma_p^{-1} r_{photo}(s, k)))$$

目标函数中包含两部分： r_{geom} 和 r_{photo} 。分别表示几何误差和光度误差。其中限制优化方向的部分在于几何误差优化部分：

$$r_{geom}(s, k) = (Tn_s)^T (\pi^{-1}(\hat{\pi}(TP_s)) - TP_s)$$

其中的 n_s 为 Surfel 的平面法向量，即误差是沿着法向量方向计算的。即将 Surfel 的中心点投影到当前帧，再使用当前帧中深度信息反投影回去，计算这个三维点与原三维点的距离，而且这个误差距离是沿着 Surfel 法向量方向的那部分距离。

剩下的光度误差计算较为简单，在 Surfel 边缘选两个点 s_1 和 s_2 ，保证 $s_1 P_s$ 和 $s_2 P_s$ 相互垂直，然后计算这两个点投影到平面之后到 P_s 投影点之间的两个灰度梯度值，这个值将作为光度信息，它与 Surfel 视觉描述子的距离即为光度误差。

优化过程

优化过程的实时性问题的解决：当使用一些二阶方法——如高斯牛顿法——来优化以上 surfel 和帧位姿时，即使采用了舒尔补提升优化速度，优化过程仍然过慢，作者给出了优化策略的算法伪代码：

Algorithm 1 Surfel-based alternating direct BA scheme

```
1: for all keyframes do Create missing surfels
2: for  $i \in [1, \text{max\_iteration\_count}]$  do
3:   Update surfel normals
4:   Optimize surfel positions and descriptors
5:   if  $i = 1$  then Merge similar surfels
6:   Optimize keyframe poses
7:   Optimize camera intrinsics (optionally)
8:   if no keyframe moved then break
9: for all keyframes which moved in the last loop do
10:   Merge similar surfels
11: Delete outlier surfels; Update surfel radii
```

这个优化策略中对降低计算时间起作用的有以下几个点：1. surfel 的法向量没有参与优化，只是对全部对应的观测值生成法向量做了一个平均；2. 三维中心点和视觉描述子优化过程中限定了优化方向（仅沿法向量方向），提高了收敛速度；3. 优化过程中相同 surfel 会做融合，这会减少计算量。

数据集

1. 使用全局快门相机而非卷帘快门相机；
2. 使用一个双目主动红外相机来获取深度，这个相机与 RGB 相机完全同步，保证深度图和 RGB 图像的完全同步。

Experiments

	clean		async		rs		async & rs	
	avg.	med.	avg.	med.	avg.	med.	avg.	med.
BundleFusion [7]	0.34	0.22	1.10	1.14	1.10	1.02	1.48	1.40
DVO SLAM [30]	0.32	0.23	2.33	0.72	5.10	1.37	4.94	1.39
ElasticFusion [69]	1.11	0.90	1.98	1.17	2.70	1.77	3.19	2.52
ORB-SLAM2 [44]	0.47	0.30	0.60	0.40	3.25	1.57	3.49	1.55
Ours	0.15	0.02	0.40	0.21	0.99	0.87	1.01	0.98

其中 clean 表示同步并使用全局快门的数据，async 表示全局快门+异步数据，rs 表示卷帘快门数据。这里说明了直接法对深度数据和 RGB 数据是否同步更为敏感。

Conclusion

个人觉得这份工作对领域内的意义在于分析了影响直接法 RGBD-SLAM 精度的重要因素——深度数据与 RGB 数据的同步以及 RGB 数据中像素的同步曝光。可以回看 ORB-SLAM 论文中对精度超过 LSD-SLAM 的分析：直接法 SLAM 中光度一致的假

设导致其特征匹配的基线不可太长（跑太远两帧之间的光度不一致就无法匹配了），这影响了重建的精度。目前看来这一因素可能并不是影响精度的主要因素。这种类型的工作厘清了一些领域内的重要问题，颇有价值，类似的还有 CVPR2019 的

《Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving》，这篇包含开源代码的工作说明了纯视觉的三维目标检测是可以达到和激光雷达类似精度的，之前的问题是三维点表示方法上存在的问题，而不是视觉重建本身的缺陷。

至于论文中提出的这个“FULL BA”的价值有待商榷，其实验给出的运行时间数据依旧不小（在 GPU 上），且其这部分的核心贡献是一个包含众多策略的优化算法，鲁棒性存疑。

转载请注明原地址，魏鑫燊的博客：<http://slowlythinking.github.io> ,谢谢！