

doi: 10.7690/bgzd.2022.04.019

## 面向快速室内视觉定位的 ORB-SLAM2 算法

陈栩睿<sup>1</sup>, 王波<sup>2</sup>, 骆云志<sup>3</sup>, 王钤<sup>3</sup>, 孙梧雨<sup>1</sup>

(1. 中国兵器装备集团自动化研究所有限公司特种计算机事业部, 四川 绵阳 621000;

2. 陆军装备部驻重庆地区军事代表局驻成都地区第一军代室, 成都 611930;

3. 中国兵器装备集团自动化研究所有限公司武器装备信息与控制技术创新中心, 四川 绵阳 621000)

**摘要:** 针对随机采样一致性 (random sample consensus, RANSAC) 算法在快速去除错误匹配时因随机性导致算法效率较低的问题, 提出一种采用顺序采样一致性 (progressive sample consensus, PROSAC) 算法来改进 ORB-SLAM2 框架中的错误匹配删除方法。通过利用特征点的匹配质量对特征点进行预排序, 减少图像匹配过程中的迭代次数; 提出基于最大化割归一化割算法 (normalized cuts and image segmentation, Ncut) 的全局 BA 分段优化算法, 以降低计算复杂度。通过数据集验证, 结果表明: 优化后的即时定位与地图构建 (simultaneous localization and mapping, SLAM) 系统在保持绝对轨迹和相对位姿误差同 ORB-SLAM2 基本一致的情况下, 相同图像的错误匹配去除的效率提升了 50%, 证明了该算法的有效性。

**关键词:** 视觉 SLAM; 特征匹配; 全局 BA 优化

**中图分类号:** TP391 **文献标志码:** A

## ORB-SLAM2 Algorithm for Fast Indoor Visual Localization

Chen Xurui<sup>1</sup>, Wang Bo<sup>2</sup>, Luo Yunzhi<sup>3</sup>, Wang Qian<sup>3</sup>, Sun Wuyu<sup>1</sup>

(1. Department of Specialty Computer, Automation Research Institute Co., Ltd. of China South Industries Group Corporation, Mianyang 621000, China; 2. No. 1 Military Representative Office in Chengdu, Military Representative Bureau of Army Equipment Department in Chongqing, Chengdu 611930, China;

3. Weapon Equipment Information and Control Technology Innovation Center, Automation Research Institute Co., Ltd. of China South Industries Group Corporation, Mianyang 621000, China)

**Abstract:** To solve the problem of low efficiency of random sample consensus (RANSAC) algorithm in fast removing error matching, a sequential sample consensus (PROSAC) algorithm is proposed to improve the error matching deletion method in orb-slam2 framework. By using the matching quality of feature points to pre sort the feature points, the number of iterations in the process of image matching is reduced. A global BA subsection optimization algorithm based on maximum cut Ncut algorithm is proposed to reduce the computational complexity. Through the data set verification, the results show that the optimized simultaneous localization and mapping (SLAM) system improves the efficiency of error matching removal of the same image by 50% under the condition that the absolute trajectory and relative pose errors are basically consistent with ORB-SLAM2, which proves the effectiveness of the algorithm.

**Keywords:** visual SLAM; feature matching; global BA optimization

### 0 引言

即时定位与地图构建 (SLAM) 是指在陌生环境下, 利用机器人携带的传感器, 在构建当前环境地图的同时, 定位自身位置和姿态的过程<sup>[1]</sup>, 可分为激光 SLAM 与视觉 SLAM 2 类<sup>[2]</sup>。目前, 由于图像处理技术的发展以及视觉传感器价格低廉的优势, 视觉 SLAM 正成为移动机器人领域的研究热点。

2015 年, MUR-ARTAL 等<sup>[3]</sup>提出了一个具有跟踪、建图和回环检测 3 个线程的单目视觉 SLAM 系统 ORB-SLAM, 并于 2017 年在此基础上提出了一种基于特征点的 SLAM 系统 ORB-SLAM2<sup>[4]</sup>, 但其

特征追踪匹配过程存在误差, 并且累积误差会直接影响系统的整体精度。为提高视觉 SLAM 前端特征点匹配的精确性, 通常在特征匹配过程中加入随机采样一致性 (RANSAC) 算法<sup>[5-6]</sup>, 以去除错误匹配的 ORB 特征点。

基于以上研究, 笔者基于 ORB-SLAM2 跟踪、建图和回环检测 3 个线程的框架, 进行顺序采样一致性 (PROSAC)<sup>[7]</sup>快速误匹配剔除优化算法的分析, 将 PROSAC 算法用于消除图像间的错误匹配, 以实现更高效的实时相机定位与跟踪。针对全局 BA 中参与优化的地图点过多, 导致计算量过大的问题,

收稿日期: 2022-01-20; 修回日期: 2022-02-18

作者简介: 陈栩睿 (1997—), 女, 山东人, 从事视觉 SLAM 研究。E-mail: 1163438417@qq.com。

采用归一化割算法 (Ncut)<sup>[8]</sup>对地图点进行分段全局 BA, 以获得更加高效的全局优化。

### 1 系统结构框架

笔者基于 ORB-SLAM2 框架进行分析, 主要包括跟踪、建图、回环检测、全局 BA 优化这 4 个线程。在特征匹配线程中, 采用 PROSAC 算法去除特征匹配误差; 同时, 对全局 BA 优化线程使用基于归一化割 Ncut 算法的优化方法。总体框架如图 1 所示。

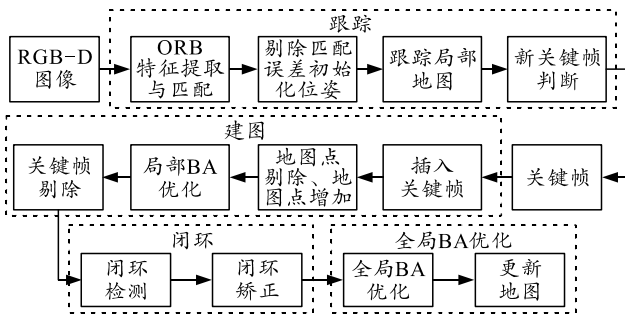


图 1 视觉 SLAM 系统框架

首先, 进行 ORB 特征的提取与匹配, 并采用优化算法去除错误匹配、初始化位姿; 然后, 跟踪局部地图, 优化当前帧的位姿, 在局部建图线程中, 删除不满足约束测试的地图点; 最后, 剔除重复的关键帧, 以降低计算复杂度。在全局 BA 线程中, 重新检测到回环时会终止此次优化, 并更新关键帧与地图, 开启新的优化进程<sup>[9]</sup>。

### 2 图像误匹配剔除 PROSAC 算法

特征匹配是视觉 SLAM 的重要组成部分, RANSAC 是计算机视觉领域中常用的优化匹配对的算法。针对 RANSAC 采样的随机性会导致算法效率低的问题, 采用优先选取质量高的数据的思想, 只需从高质量的数据子集中抽取样本, 不需要抽取所有的数据。

假设采样数据集的内点数大于外点,  $N$  个数据集点用  $\mu_N$  表示, 根据质量函数  $q(u)$ ,  $\mu_N$  中的 2 个采样点  $\mu_i$ 、 $\mu_j$  按降序排列:

$$\mu_i, \mu_j \in \mu_N : i < j \rightarrow q(u_i) < q(u_j) \quad (1)$$

$P(u_i)$  为该点的采样概率, 对  $P(u_i)$  和  $q(u_i)$  的关系做如下假设:

$$q(u_i) \geq q(u_j), \text{ 即 } i < j \rightarrow P(u_i) \geq P(u_j) \quad (2)$$

然后用 RANSAC 算法从  $\mu_N$  中提取大小为  $m$  的  $T_N$  个样本, 记为  $\{M_i\}_{i=1}^{T_N}$ 。并按照采样质量降序排列为  $\{M_{(i)}\}_{i=1}^{T_N}$ , 即

$$i < j \rightarrow q(M_{(i)}) \geq q(M_{(j)}) \quad (3)$$

如果按照  $M_{(i)}$  的顺序进行采样, 则优先提取高质量的点, 然后再提取质量低的外点。记  $U_n$  为有  $n$  个高质量内点的集合, 记  $T_n$  为  $\{M_i\}_{i=1}^{T_N}$  中包含  $U_n$  中的高质量点的平均数:

$$T_n = T_N \binom{N}{M} / \binom{N}{m} = T_N \prod_{i=0}^{m-1} \frac{n-i}{N-i} \quad (4)$$

则有:

$$\frac{T_{n+1}}{T_n} = \frac{T_N}{T_N} \prod_{i=0}^{m-1} \frac{n+1-i}{N-i} \prod_{i=0}^{m-1} \frac{N-i}{n-i} = \frac{n+1}{n+1-m} \quad (5)$$

最终可得:

$$T_{n+1} = (n+1)/(n+1-m) T_n \quad (6)$$

这样  $T_n$  个采样集合只包含集合  $U_n$  中的高质量内点集, 而  $U_{n+1} = U_n \cup \{u_{n+1}\}$ , 则  $T_{n+1} - T_n$  个采样集合中包含数据点  $u_{n+1}$  和  $m-1$  个  $U_n$  中的数据点。通常  $T_n$  的值不是整数, 设  $T'_m = 1$ , 则  $T'_{n+1} = T'_n + |T_{n+1} - T_n|$ 。

定义增长函数为  $g(t) = \min\{n : T'_n \geq t\}$ , 在 PROSAC 中, 第  $t$  次采样集合  $M_t$  为  $M_t = \{u_{g(t)}\} \cup M'_t$ 。这里  $M'_t \subset U_{g(t)}$  为一个从  $U_{g(t)}$  集合中随机抽取的势为  $|M'_t| = m-1$  的子集。

针对视觉 SLAM 中传统算法在去除错误匹配时存在随机性而导致计算量过大的问题, 优化后的算法按照质量高低对特征点进行预排序。虽然增加了计算量, 却减小了迭代次数, 从而减少整体运算量, 以提高算法效率。

### 3 光束平差后端优化算法

全局优化时, 由于参与优化的关键帧及地图点过多, 会导致计算量偏大; 因此, 需要减少参与优化的地图点。在图像分割领域, Ncut 是一种常用的分割算法, 其主要思想是聚类, 可以将那些相关性更加紧密的图像分为一类, 然后再进行子序列的分割<sup>[8]</sup>。

笔者将 Ncut 算法应用于全局 BA 中, 通过分段的思想来减少优化的地图点数量。该算法将所有地图点分为若干个段, 每段输出一个新地图点。新地图点和关键帧将会作为全局 BA 的变量。这种思想使得参与优化的地图点数量大大减少, 降低了计算复杂度。

把带权图  $G=(V, E)$  分为 2 个不相交的集合  $A, B$ , 且  $A \cup B = V, A \cap B = \phi$ 。那么可以用 Ncut 标准计算 2 个集合的不相关程度:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)}. \quad (7)$$

其中： $assoc(A, V) = \sum_{u \in A, t \in V} W(u, t)$ ,  $assoc(B, V) = \sum_{u \in B, t \in V} W(u, t)$ ,  $cut(A, B) = \sum_{u \in A, t \in B} W(u, v)$ 。

### 4 实验及其结果分析

实验中使用的计算机是戴尔 PowerEdge T110 II 塔式服务器，运行平台为 ubuntu16.04。

TUM(technische universitaet münchen) 数据集是由慕尼黑工业大学提供的 RGBD 数据集，数据集包括由深度相机获取的 RGB 图像信息与深度图像信息，以及由 8 个高速追踪相机所构成的高精度运动捕捉系统所获取的机器人真实移动轨迹 (ground truth)<sup>[10]</sup>。

#### 4.1 特征点匹配对比实验

笔者对优化前后的算法进行对比实验，将 6 张图片分为特征点多、较多以及少 3 组，进行错误匹配去除时间对比实验，结果如表 1 所示。

表 1 优化前后算法指标比较

特征点数量	方法	匹配总时间/ms	错误匹配去除时间/ms	特征点对数	内点数
多	优化前	40.039	6.794	374	320
	优化后	96.138	47.094	374	323
较多	优化前	36.478	5.143	230	171
	优化后	80.726	42.946	230	169
少	优化前	32.117	3.215	122	74
	优化后	66.030	33.121	122	79

由上表可知，2 种算法获得内点数基本相同，但采用优化后的匹配总时间与错误匹配去除时间都大幅减少，使用 PROSAC 算法将大大提高特征匹配线程的效率。

原始匹配如图 2 所示，优化后的处理效果如图 3 所示。



图 2 优化前匹配结果



图 3 优化后的匹配结果

#### 4.2 特征误匹配剔除定位实验

运行上述数据集，对优化前后的算法运行时间进行比较，结果如表 2 所示。可见，优化后运行时间明显减少。

表 2 改进前后的时间比较

数据集	原始运行时间/ms	改进后的运行时间/ms
Fr1_desk	36.17	20.48
Fr1_desk2	42.63	23.28

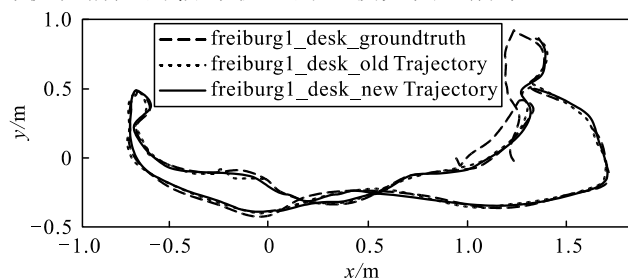
采用绝对轨迹误差 (absolute trajectory error, ATE) 和相对姿态误差 (relative attitude error, RPE) 来比较估计轨迹与真实轨迹之间的误差<sup>[11]</sup>。绝对轨迹误差直接计算相机位姿的真实值与视觉 SLAM 系统的估计值的差，适用于评估视觉 SLAM 系统的性能，相对姿态误差用于计算 2 个相同时间戳上的位姿变化的差异，适用于估计视觉 SLAM 系统的漂移<sup>[12]</sup>。

$$ATE_{ALL} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\log(T_{gt,i}^{-1} T_{esti,i})^\vee\|_2^2}. \quad (8)$$

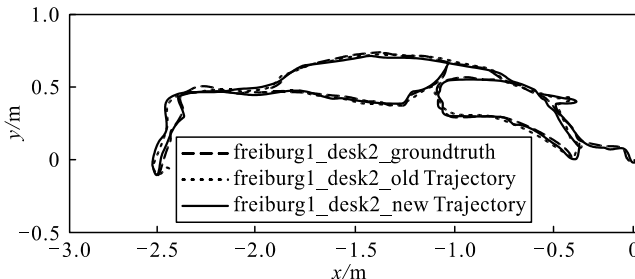
其中： $T_{esti,i}$  为估计轨迹； $T_{gt,i}$  为真实轨迹。考虑  $i$  时刻到  $i+\Delta t$  时刻的运动，那么相对位姿可以定义为：

$$RPE_{ALL} = \sqrt{\frac{1}{N-\Delta t} \sum_{i=1}^{N-\Delta t} \|\log(T_{gt,i}^{-1} T_{gt,i+\Delta t})^{-1} \log(T_{esti,i}^{-1} T_{esti,i+\Delta t})^\vee\|_2^2}. \quad (9)$$

运行数据集 fr1\_desk 和 fr1\_desk2 的真实轨迹与优化前后的估计轨迹的比较如图 4 所示。



(a) fr1\_desk 真实轨迹与优化前后的估计轨迹对比



(b) fr1\_desk2 真实轨迹与优化前后的估计轨迹对比

图 4 真实轨迹与估计轨迹之间的误差比较

根据实验结果，分别计算出优化前后 2 组数据集的均方根误差和中值，数据指标如表 3 所示。

表 3 各项指标比较

数据集	方法	均方根误差/m	中值/m
Fr1_desk	PROSAC	0.018 121	0.011 180
	RANSAC	0.018 243	0.012 141
Fr1_desk2	PROSAC	0.026 472	0.019 013
	RANSAC	0.026 707	0.019 813

由上表可知,优化前和优化后的视觉 SLAM 系统在定位效果及精度上差别很小,但优化后的视觉 SLAM 运行时间大大缩短,效率有明显提高。

### 4.3 全局 BA 优化算法实验

采用基于归一化割 Ncut 的全局 BA 优化算法,以 Fr1\_desk 数据集为例,结果如图 5 所示。

```

18 parts
1, 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
2, 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41
3, 42 43 44 45 46 47 48 49 50 51 ... 63 64 65 66 67 68 69 70 71 72 73 74 75
4, 76 77 78 79 80 81 82 83 84 85 ... 97 98 99 100 101 102 103 104 105 106
5, 107 108 109 110 111 112 113 ... 142 143 144 145 146 147 148 149 150 151
6, 152 153 154 155 156 157 158 ... 187 188 189 190 191 192 193 194 195 196
7, 197 198 199 200 201 202 203 ... 229 230 231 232 233 234 235 236 237 238
8, 239 240 241 242 243 244 245 ... 261 262 263 264 265 266 267 268 269 270
9, 271 272 273 274 275 276 277 ... 301 302 303 304 305 306 307 308 309 310
10, 311 312 313 314 315 316 317 ... 362 363 364 365 366 367 368 369 370 371
11, 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388
...
18, 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 |
    
```

图 5 数据集分段结果

由上图可知,依据帧之间可视地图点的数量差异,在差异较大的地方作为分段点,将序列分为 18 段,该数据集原始需优化的位姿为 564 个,通过优化过程需要求解的位姿变为 18 个;因此,该优化思想大大降低了待优化位姿的数目,降低了计算复杂度。

如表 4 所示,无论数据集有无回环,关键帧获取数量基本保持一致,并且优化后的算法消耗时长明显减少。

表 4 分段策略的耗时对比

数据集	是否分段	消耗时间/s	关键帧获取数量	数据集帧数量
Fr1_desk	是	5.018	564	2 338
	否	9.207	573	2 338
Fr1_desk2	是	6.135	620	2 428
	否	11.022	626	2 428

将基于归一化割 Ncut 的全局 BA 优化算法加入上文优化后的算法中,进行定位实验,并与传统的 ORB-SLAM2 进行比较。分别计算出 2 组数据集在不同方法下的均方根误差与中值,数据指标如表 5 所示。

表 5 各项指标比较

数据集	是否分段	均方根误差/m	中值/m
Fr1_desk	是	0.018 321	0.013 180
	否	0.018 014	0.012 843
Fr1_desk2	是	0.026 472	0.019 013
	否	0.025 117	0.018 482

由上表可知,优化前和优化后的视觉 SLAM 系统在定位效果及精度上差别很小,但优化后的视觉 SLAM 运行时间大大缩短,效率大大提高。

## 5 结论

基于 ORB 特征进行轨迹估计与位姿估计时,快速有效的错误匹配消除过程十分重要。笔者基于 ORB-SLAM2 框架,采用 PROSAC 算法消除错误匹配,优化后的 SLAM 系统在保持绝对轨迹和相对位姿误差同 ORB-SLAM2 基本一致的情况下,相同图像的错误匹配去除效率提升了 50%。另外,提出基于归一化割 Ncut 算法的全局 BA 分段优化算法来解决全局 BA 中计算量过大的问题,使优化后的算法相较于传统算法有着更高的效率。

## 参考文献:

- [1] 庄严,王伟,王珂,等. 移动机器人基于激光测距和单目视觉的室内同时定位和地图构建[J]. 自动化学报, 2005(6): 113-121.
- [2] 权美香,朴松昊,李国. 视觉 SLAM 综述[J]. 智能系统学报, 2016, 11(6): 768-776.
- [3] 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.
- [4] 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, 33(5): 1255-1262.
- [5] 方贤勇. 图像拼接技术研究[D]. 杭州: 浙江大学, 2005.
- [6] 唐管政,唐大全,谷旭平. 复杂纹理环境的光流/惯性组合导航方法[J]. 兵工自动化, 2021, 40(6): 26-31.
- [7] 赵璐璐,耿国华,李康,等. 基于 SURF 和快速近似最近邻搜索的图像匹配算法[J]. 计算机应用研究, 2013, 30(3): 921-923.
- [8] 席秋波. 基于 Ncut 的图像分割算法研究[D]. 成都: 电子科技大学, 2010.
- [9] 邓琳蔚,王治,张梦梦,等. 一种基于光束平差法的相机标定方法[J]. 兵工自动化, 2020, 39(2): 8-13.
- [10] 周明超. 基于 ORB-SLAM2 改进的八叉树地图构建[J]. 电脑知识与技术, 2018, 14(34): 236-237.
- [11] 王召东,郭晨. 一种动态场景下语义分割优化的 ORB-SLAM2[J]. 大连海事大学学报, 2018, 44(4): 121-126.
- [12] 高翔,张涛,颜沁睿,等. 视觉 SLAM 十四讲: 从理论到实践[M]. 北京: 电子工业出版社, 2017: 33-38.