

# 一种融合单目信息的 RGB-D SLAM 优化方法

马树军, 王英蕾, 金铁铮, 白昕晖  
(东北大学 机械工程与自动化学院, 辽宁 沈阳 110819)

**摘 要:** 由于 RGB-D 传感器本身存在的局限性, 其在实际应用中会有深度信息不存在或者不完整的情况, 这使得多数依赖于深度信息工作的 RGB-D SLAM 系统的精度会受到影响. 提出了一种融合单目信息的 RGB-D SLAM 优化方法. 通过对深度图像进行分析, 从深度信息的存在性、准确性和离散性分别提出三个判断条件, 作为选择单目或者 RGB-D 处理方式的依据. 该系统使得用 RGB-D 相机工作的 SLAM 系统更具精确性与鲁棒性. 最后通过运行数据集将系统与其他 SLAM 系统进行比较, 实验结果验证了本系统的可行性.

**关 键 词:** RGB-D SLAM; 单目信息; 融合; 判断机制; 优化  
**中图分类号:** TP 274      **文献标志码:** A      **文章编号:** 1005-3026(2020)06-0841-06

## A RGB-D SLAM Optimal Method for Fusing Monocular Information

MA Shu-jun, WANG Ying-lei, JIN Tie-zheng, BAI Xin-hui  
(School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China.  
Corresponding author: MA Shu-jun, E-mail: mashujun@me.neu.edu.cn)

**Abstract:** Owing to the limitations of RGB-D sensor, there will be some situations where the depth information does not exist or is incomplete in the practical application, which will affect the accuracy of RGB-D SLAM system that mostly depends on the depth information. An RGB-D SLAM optimization method is proposed to fuse monocular information. Through the analysis of depth image, three judging conditions are put forward respectively from the existence, accuracy and discreteness of depth information, which can be used as the basis of the selection of monocular or RGB-D processing mode. This new system makes the SLAM system with RGB-D camera more accurate and robust. Finally, the system is compared with other SLAM systems by running data sets, and the experimental results verify its feasibility.

**Key words:** RGB-D SLAM; monocular information; fusion; judgment mechanism; optimization

RGB-D 相机可以同时采集彩色图像和深度图像, 可以提供丰富的场景信息并且减小在 SLAM 系统中的计算难度, 基于 RGB-D 相机的 SALM 系统一直是视觉 SLAM 领域的热门研究问题<sup>[1]</sup>. 但当相机运动速度过快、场景没有足够的结构或者存在干扰光源时, 会有深度信息不存在和不可靠的情况; 而且 RGB-D 相机的有效工作范围为 0.6~4 m, 当其不在有效范围内工作时深度会变得离散和稀疏, 整体深度的不确定性成指数增长<sup>[2]</sup>. 不精确的深度估计会引起较大的线性化误差, 影响估计精度, 甚至发散<sup>[3]</sup>. 所以一直依赖深度信息工作的 SLAM 系统必然会受到存在误差的深度信息的影响.

为解决 RGB-D SLAM 存在的问题, 有学者提出融合其他传感器的方法. RGB-D 相机和激光雷达融合<sup>[4]</sup>, 激光雷达可以提供精度更高、范围更大的数据支持. 但是激光雷达价格昂贵并且几何尺寸和重量都比较大, 因此在小型的室内移

动机器人上并不适用. RGB - D 相机和 IMU 惯导模块融合<sup>[3,5]</sup>,利用 IMU 传感器提供加速度和角速度测量,在深度信息丢失或没有特征信息的情况下,仍然可以对位姿进行估计,使系统具有很好的鲁棒性并且不需时刻保持低速以防 RGB - D 相机跟丢. 但 IMU 存在漂移误差,而且增加了对系统计算能力的要求. RGB - D 和单目进行融合<sup>[6]</sup>,当图像传入系统,该方法依次对单独的彩色图像和彩色与深度组合的图像分别进行计算,然后从二者中选择最优的方式. 这种融合方法使系统在深度缺乏的情况下具有更好的精确性. 但系统对每帧都分别进行两种方式的处理会过多占用系统的计算资源.

通过研究,本文基于 ORB SLAM 系统<sup>[7-8]</sup>提出一种 RGB - D 和单目信息融合的优化方法,该方法可以通过判断深度图像的信息,有选择性地进入 RGB - D SLAM 或者 Mono SLAM.

1 系统框架

本文提出的融合单目信息的 RGB - D SLAM 优化方案如图 1 所示,系统通过 RGB - D 相机采集场景信息进行 SLAM 工作. 其中利用 RGB 图像进行特征提取与匹配,并进行初始位姿估计. 对深度图,本文提出一个基于深度的判断机制,从深度信息的存在性、准确性和离散性三个方面来判断深度图像是否具有足够的信息来进行位姿优化. 若满足,则将深度图像输入,则系统是采用包含深度信息的三维信息对当前帧的初始位姿进行优化,此为典型的 RGB - D SLAM 方式;若没有足够的深度信息,则该帧的深度图像不输入系统,系统采用二维信息对当前帧的初始位姿进行优化,此为 Mono SLAM 方式. 系统还包含回环检测和全局优化,最后将优化的位姿作为输出.

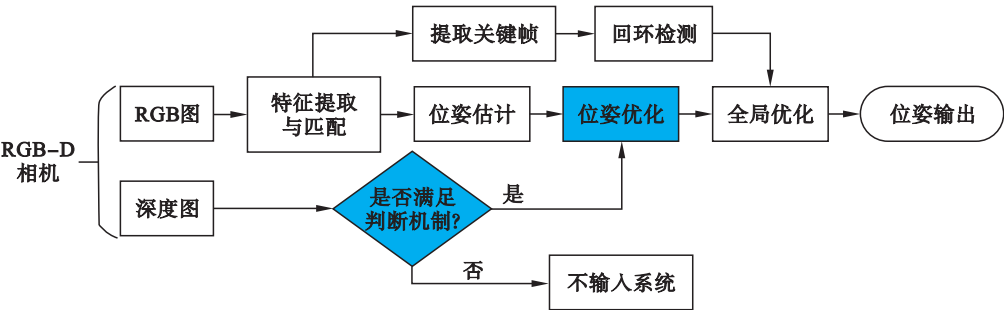


图 1 优化算法流程图  
Fig. 1 Optimization algorithm flow chart

2 系统详述

2.1 特征提取

系统在前端使用的是 ORB (oriented FAST and rotated BRIEF)特征提取算法. ORB 算法包含改进的 FAST (features from accelerated segment test) 特征检测算法和 BRIEF (binary robust independent elementary features) 特征描述算法<sup>[9]</sup>.

FAST 算法检测图像像素点的灰度值,并与周围点灰度值进行对比,若两者存在较大差异,则认为该像素点是特征点,可用式(1)来描述该过程. 在 ORB 特征算法中针对 FAST 角点不具备方向性和尺度的弱点,添加了对尺度和旋转的描述.

$$M = \sum_{x \in \text{circle}(p)} |I(x) - I(p)| > t. \quad (1)$$

式中: $I(p)$  为圆心的灰度值; $I(x)$  为圆周的灰度值; $t$  为对比阈值.

BRIEF 是一种二进制编码的描述算子,用于对已检测到的特征点进行描述. 其工作原理就是在关键点  $B$  周围随机选取两个像素点  $a, b$  作为一对,通过式(2)对比两个像素点的灰度值得到一组  $k$  位的二进制字符串. 其中  $k$  为随机选取的点对数,在本文中  $k = 256$ . 图像经过 BRIEF 算法处理之后,图像中的每个特征点都对应一个 256 位的二进制字符串用于特征点之间的匹配.

$$\tau(B;a,b) = \begin{cases} 1, I(a) < I(b); \\ 0, I(a) \geq I(b). \end{cases} \quad (2)$$

2.2 位姿估计

在位姿估计中,若系统处于正常状态,有两种位姿估计的方式:若上一帧跟踪成功并且相邻两帧有足够的匹配点对,则通过上一帧的位姿和变化的速度来估计当前帧的位姿;若上一帧跟踪失败或者相邻两帧没有足够的匹配点对,则系统需要通过关键帧来估计位姿. 若处于丢失状态,则系统需要通过重定位的方式来估计当前帧的位姿.

重定位的方式是从包含关键帧的库中找到与之最相似的关键帧,并进行匹配. 通过 EPNP( efficient perspective-n-point)算法计算当前相机坐标系下的匹配点坐标,再利用 ICP( iterative closest point)算法进一步求解当前帧的位姿.

2.3 深度判断机制

为了评估相机在各方面的适用性,本文进行了一系列测试来描述不同情况下深度数据的质量.

实验一:通过控制搭载 Kinect 的移动小车在实验室和楼道各采集 200 张图片,对每张图片进行是否检测到深度数据的测试( $d > 0$ ). 同时,该实验也对 TUM<sup>[10]</sup>数据集内的多个数据集进行了测试. 实验结果见图 2,该图表示的是深度信息低于 70% 的图像占数据集总数的百分比.

前 6 个数据均来自与 TUM 数据集,后 2 个分别是采集楼道和实验室的图像集. 统计结果说明,不管是数据集还是实际应用中,都会遇到深度信息采集不完全的情况,而这样的场景信息必然会影响 SLAM 的鲁棒性.

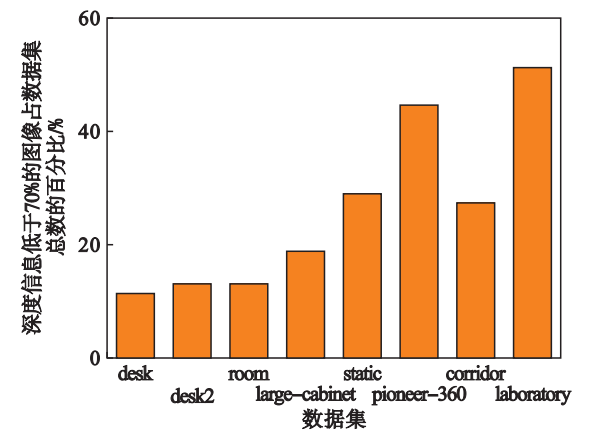


图 2 深度信息分析  
Fig. 2 Deep information analysis

实验二:对 Kinect 距离准确性进行测试. 使 Kinect 面向普通墙壁,与墙壁的距离由 50 cm 增加至 440 cm,每次的增量为 30 cm,在每一个节点中采集 50 张图像,为排除地板和图像边角误差的干扰,划定一个 160 × 240 的网格区域来进行测度检测,如图 3 中蓝色图框所示.



图 3 实验样本图  
Fig. 3 Pictures of experimental samples

图 4 展示了本次实验的结果,分别计算了每一组图像距离排序的前 50% 和前 80% 像素点所对应的距离与真实距离的误差百分比. 从结果可以看出,测量距离为 0.8 ~ 3.8 m 时结果比较理想,当距离过近或过远时,整个结果会出现较大误差.

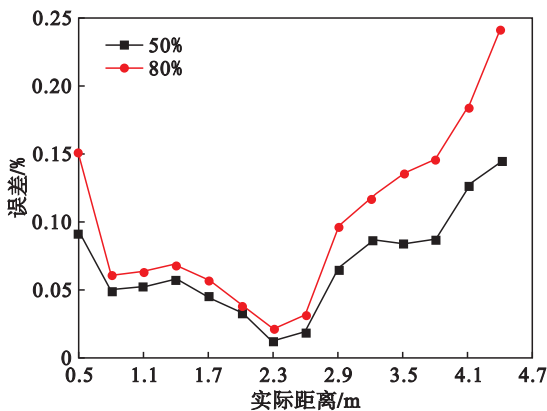


图 4 前 50% 和 80% 数据与真实距离的误差百分比  
Fig. 4 The percentage of error between the first 50% and 80% data and the true distance

实验三:提出猜想,即使有足够的深度信息,但是深度特征过度分散,也会影响 SLAM 的精度. 针对以上,本文对 TUM 数据集进行了实验. 如图 5 所示,在TUM两个数据集内分别提取两

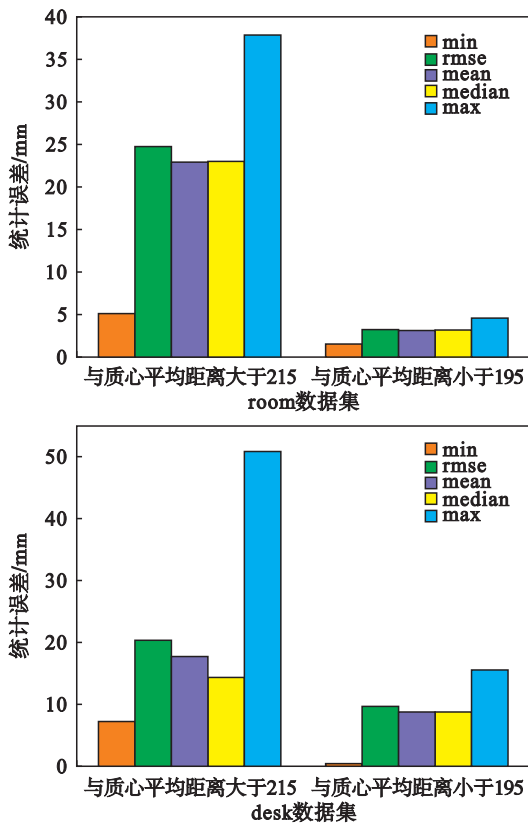


图 5 距离质心不同像素距离数据运行结果对比  
Fig. 5 Comparisons of running results of distance centroid with different pixel distances

组数据,一组为与质心的像素距离大于 215 和一组与质心像素距离小于 195,对这两组数据进行实验.实验结果表明,在其他条件相似的情况下,距离过大会导致误差增大.

针对以上实验,本文从三个方面提出三个模型作为深度选择机制:

情况一:深度信息不存在.

$$d > 0. \tag{3}$$

情况二:深度信息不可信.

$$\gamma_1 \leq d_{\text{real}} \leq \gamma_2. \tag{4}$$

情况三:深度信息过度分散.

$$\alpha = \frac{1}{n} \sum_{i=1}^n d(p_i - P). \tag{5}$$

$$P = \frac{1}{n} \sum_{i=1}^n (p_i). \tag{6}$$

其中: $d$  为像素点的深度值; $d_{\text{real}}$  为像素点的真实距离. $P$  为深度图像的质心,由式(6)计算得到.模型(3)和(4)是对单一的像素点进行判断,最后还要满足式(7),对整张图片所有像素点的判断.

$$n_c/n < \gamma_d. \tag{7}$$

$n_c$  为满足式(3)和(4)的像素点的数量; $n$  为整张图像像素点的总和,本文中涉及到的  $n = 480 \times 640$ .阈值的选取通过实验得到: $\gamma_1$  和  $\gamma_2$  分别为 0.8 m 和 3.8 m, $\alpha$  取 215, $\gamma_d$  在保证系统处理速度的情况下取数据集总量的 15%.

2.4 位姿优化

在这一过程中本文采用光束平差法(BA 算法)优化表征当前帧位姿的旋转矩阵  $R$  和平移向量  $t$ ,以及最小化世界坐标系下匹配的 3D 点云  $X^i \in R^3$  和  $X_{(\cdot)}$  的重投影误差:

$$\{R, t\} = \arg \min_{R, t} \sum_{i \in X} \rho(\|X_{(\cdot)}^i - \pi_{(\cdot)}(RX^i + t)\|_{\Sigma}^2). \tag{8}$$

式中: $\rho$  是 Huber 的 cost 函数, $\Sigma$  是协方差矩阵. $X_{(\cdot)}^i$  表示特征点的坐标向量; $\pi_{(\cdot)}$  为投影函数.其中当系统只使用彩色图像时,特征点坐标用  $X_M = (u, v)$  表示, $X_M^i \in R^2$ ,投影函数  $\pi_M$  用式(10)表示;系统同时使用彩色和深度图像时,特征点坐标用  $X_R = (u, v, u_d)$  表示, $X_R^i \in R^3$ , $u_d$  由式(9)计算而来.投影函数  $\pi_R$  用式(11)表示.

$$u_d = u_L - \frac{f_x b}{d}, \tag{9}$$

$$\pi_M \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \end{bmatrix}, \tag{10}$$

$$\pi_R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \\ f_x \frac{X-b}{Z} + c_x \end{bmatrix}. \tag{11}$$

式中: $b$  为结构光相机和红外摄像机之间的基线; $f_x, f_y, c_x, c_y$  为相机的内部参数.

2.5 全局优化

系统利用关键帧之间建立的约束完成闭环检测的工作.为了更好地闭合回环,最后通过全局 BA 优化来实现误差的消除.

3 实验研究

所有实验均在 Ubuntu 14.04 环境下完成.实验数据来自于 TUM RGB-D 数据集<sup>[10]</sup>.

本文用优化的系统运行 TUM 数据集,来验证整个系统的优劣.本文以 teddy 数据集展示运行结果.图 6 中展示了数据集运行的对比结果.通过实验结果可以看出,本系统运行得到的轨迹结果与真实的轨迹十分贴合.通过运行 EVO 算法评估真实轨迹和估计轨迹,可以得到不同时间轨迹的绝对误差、均方误差、中位数和中值等参数指标以及不同坐标轴的轨迹误差.图 7 展示了评估结果.

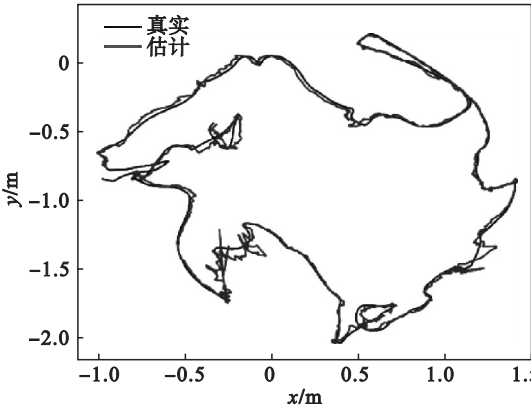


图 6 估计轨迹和真实轨迹对比图  
Fig. 6 Comparison of estimated trajectory and real trajectory

本文还通过 EVO 工具对本文系统和 ORB-SLAM 运行相同数据集时得到的运行结果进行了对比.如图 8 所示,从运行结果可以看出,本文优化的系统相比 ORB-SLAM 系统的精度更高一些.



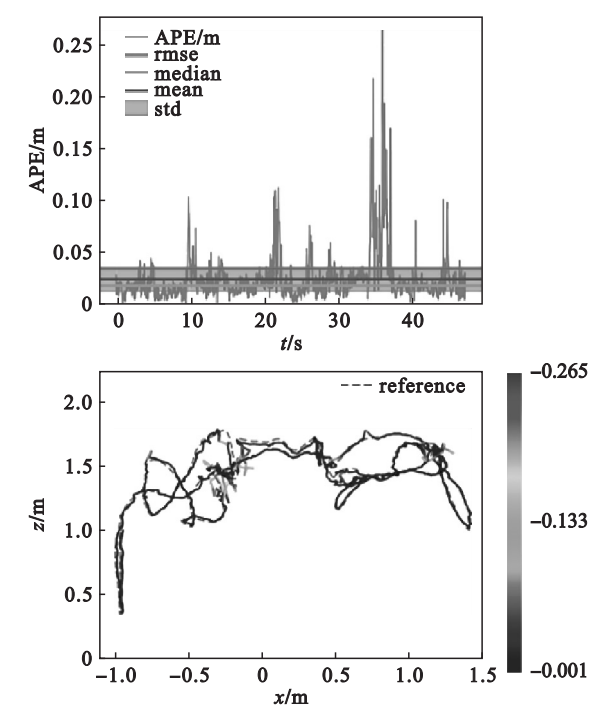


图 7 teddy 数据集分析结果

Fig. 7 Analysis result of teddy data set

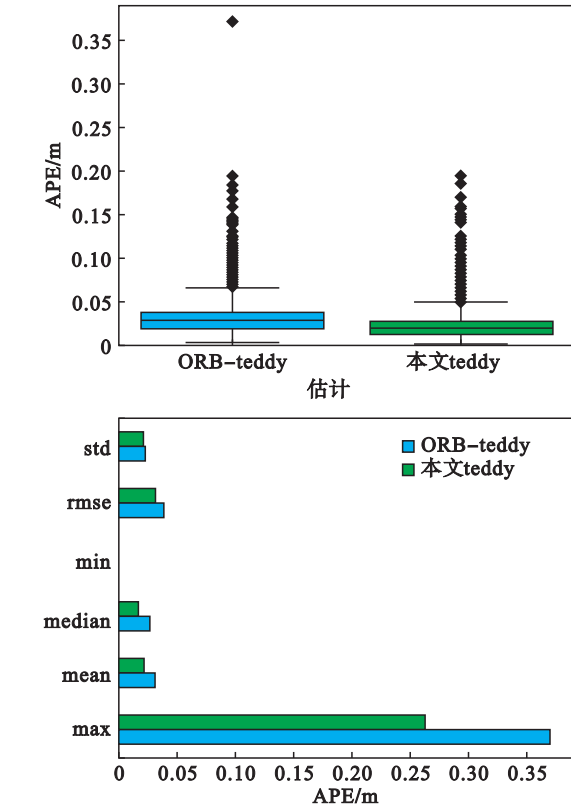


图 8 teddy 数据集对比结果

Fig. 8 Comparison of teddy data set

TUM 数据集共包含 6 个分类 39 个图像序列,本文将优化的系统与 ORB SLAM<sup>[7]</sup> 分别进行对比. 由于篇幅有限,只选取其中具有代表性的数

据进行展示. 图 9 为两种算法运行数据集得到的均方误差的对比结果;图 10 为两种算法运行数据集所需时间的对比. 其中共有 15 组序列,1 ~ 6 对应的序列分别为 rgbd\_dateset\_freiburg1 模式的 desk, floor, room, rpy, xyz 和 teddy; 7 ~ 9 对应 rgbd\_dateset\_freiburg2 模式的 xyz, pioneer\_360 和 pioneer\_slam2 序列; 10 ~ 15 对应的序列是 rgbd\_dateset\_freiburg3 模式下的 structure\_texture\_near, nostructure\_texture\_far, sitting\_static, sitting\_xyz, walking\_xyz 和 teddy. 从图中可以看出,在多数情况下,本文改进的优化系统具有更好的精确性和高效性.

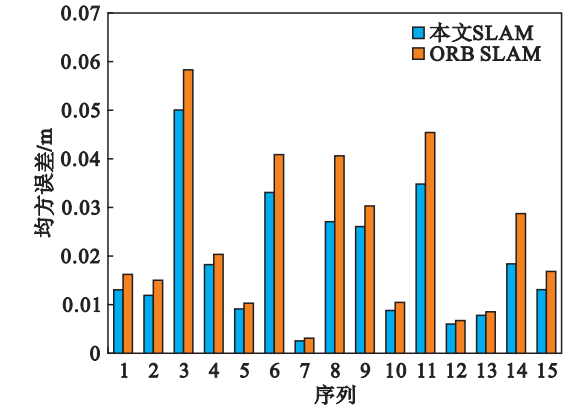


图 9 两种算法均方误差对比

Fig. 9 RMSE comparison of two algorithms

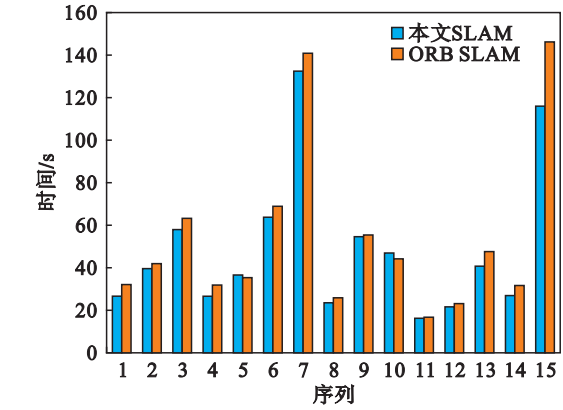


图 10 两种算法时间对比

Fig. 10 Time comparison of two algorithms

## 4 结 语

本文提出一种融合单目信息的 RGB - D SLAM 优化系统. 该算法可以更好地规避误差较大的深度图像给系统造成的影响. 通过实验验证了改进算法的可行性与精确性.

(下转第 851 页)