

一种非迭代的位姿估计方法

张一铭,吕耀文,徐熙平,吕梦凌,钟文婷
(长春理工大学光电工程学院,吉林 长春 130022)

摘要:针对深度相机运动中的位姿估计问题,提出了一种无需迭代的估计方法。首先,在二维图像上应用图像特征点提取和描述方法,完成不同视点的初始匹配。其次,选择初始匹配度量距离最小的2个特征点作为种子点。以三维空间中欧式距离与坐标系的建立无关为准则,对初始匹配进行筛选。剔除误匹配点对,进而计算运动位姿参数。最后,采用nyuv2图像数据库进行实验,验证了本文算法的可行性和正确性。实验结果表明:与传统算法相比,该方法计算效率平均提高了8倍以上,特别适用于大型场景中的同步定位和地图构建SLAM(Simultaneous Localization and Mapping)问题。

关键词:位姿估计;深度相机;距离筛选;SLAM

中图分类号:TP391 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2018.05.019

A new method for pose estimation without iteration

ZHANG Yi-ming, LÜ Yao-wen, XU Xi-ping, LÜ Meng-ling, ZHONG Wen-ting

(School of Opto-Electronic Engineering, Changchun University of Science and Technology, Changchun 130039, China)

Abstract: Aiming at the problem of estimating the motion parameters of depth camera, a new estimation method without iteration is proposed. Firstly, the initial match of different viewpoints is completed according to extraction and description method of the image feature points. Secondly, the two feature points with smallest distance from the initial matching are selected as seed points. Based on the independence of Euclidean distance and coordinate system in 3D space, the initial match is filtered to eliminate the mismatch points. Then, the motion parameters are calculated. Finally, in order to verify the feasibility and correctness of the proposed algorithm, experiments are carried out in the open database nyuv2. The results show that the proposed method can improve the computational efficiency by 8 times or more compared with the traditional algorithm, and it is especially suitable for large scale SLAM.

Key words: pose estimation; depth camera; distance filter; SLAM

1 引言

SLAM是指机器人在未知环境中从某一位置开始移动,在移动过程中根据位置估计和传感器数据进行自身的定位,同时重建增量式地图的过程^[1]。

SLAM被认为是实现全自主移动机器人的核心环节,被广泛应用于未知环境的探索、无人机的导航与路径规划、灾难现场的搜寻与营救等众多领域中^[2]。

基金项目:吉林省科技发展计划资助青年基金项目(No.20160520018JH);吉林省重点科技攻关项目(No.20170204048GX)资助。

作者简介:张一铭(1993-),男,硕士研究生,主要从事vSLAM方面的研究。E-mail:isyiming_zhang@outlook.com

通信作者:吕耀文(1987-),男,工学博士,讲师,硕士生导师,主要从事计算机视觉与机器学习方面的研究。E-mail:lvyawen2005@163.com

收稿日期:2017-09-11

一个清晰的 SLAM 系统框架大致可以分为两部分:前端和后端,前端负责提取传感器的数据并构建模型,后端根据前端的数据进行优化。对于视觉的 SLAM 而言,前端主要是指视觉里程计的实现,即根据图像求解相机位姿。而后端则是对前端的计算结果进行优化,以获得更精确的位姿估计和全局一致的地图。因此,位姿估计是 SLAM 中的首要环节,估计精度直接影响最终的定位和地图重建结果^[3-4],计算消耗的时间也将影响到 SLAM 的实时性。由此可见,对 SLAM 前端的位姿估计进行优化是非常必要的。

在使用深度相机进行位姿估计时,常见的方法是求解 n 线透视 (PnL)^[5] 或者 n 点透视 (PnP) 问题,其中 PnP 求解较为常见^[6-7]。PnP 计算时要求参与计算的点都是正确的,但是初始图像特征匹配经常存在误匹配。通常采用 RANSAC (Random Sample Consensus) 迭代算法^[6] 剔除误匹配,这种方法可靠性强、精度高、鲁棒性强,具有较好的剔除误匹配点的能力。但 RANSAC 依赖迭代运算提高精度,所以计算代价很高。尤其是在数据量大、特征点匹配错误率高时,时间消耗会呈指数级增长。Felix Endres 等人^[8] 在 RANSAC 算法随机抽样时,引入了三点空间距离相等的筛选准则,在一定程度上减少了 RANSAC 算法的计算量,但是算法仍需迭代。

为此,本文以三维空间中欧式距离与坐标系的建立无关为准则,结合初始图像匹配中特征点描述的距离信息,提出了一种无需迭代的新方法,剔除大量错误特征点,进行直接求解。从而提高求解精度,减少求解时间。

本文首先介绍了求解深度相机位姿变化的数学基础,包括 PnP 算法和采用 RANSAC 剔除误匹配求解 PnP 的算法;其次详细论述和说明了本文的改进方法,然后通过实验,从时效和准确性两个方面对改进方法和传统 RANSAC 算法进行了对比分析。最后,对全文做了归纳和总结。

2 深度相机的位姿变化估计过程

2.1 PnP 算法

相机运动过程中,利用两帧具有部分景物重叠的图像可以求解相机的位姿变化。如图 1 所示,设空间中一点 $P(x, y, z)$ 在两幅图像 F1、F2 中的投影图像的齐次坐标分别为: $p_1(u_1, v_1, 1)$ 和 $p_2(u_2, v_2,$

1)。第二帧图像相对于第一帧图像位姿变化由旋转矩阵 R 和位移矢量 t 描述。相机的内参数为 K 。所以:

$$\begin{aligned} d_1 p_1 &= KP \\ d_2 p_2 &= K(RP + t) \end{aligned} \quad (1)$$

其中, d_1 、 d_2 是两个图像点由深度相机获取的已知投影深度。记归一化相机坐标为:

$$\begin{aligned} x_1 &= d_1 K^{-1} p_1 \\ x_2 &= d_2 K^{-1} p_2 \end{aligned} \quad (2)$$

将式(2)代入式(1)得到:

$$x_2 = R x_1 + t \quad (3)$$

R 和 t 有 6 个自由度,一对匹配点可以得到关于 R 和 t 的两个约束方程,因此获取三个以上的匹配点就可以利用对应点的关系求解相机的位移矢量和旋转矩阵,即得到了相机的位姿变化。实际计算中,在多于三对特征点时,由于特征点的定位误差和深度图像的采集误差,匹配点对并不能完全满足公式(3)。因此在根据式(3)求解 R 和 t 的过程中需要将求解转化为优化问题,优化方程为:

$$\min \sum_{c=1}^n \| R x_1^c + t - x_2^c \| \quad (4)$$

其中, x_1^c 、 x_2^c 分别是图像 1 和图像 2 的归一化相机坐标,由式(2)计算得到, n 为参与计算的特征点对数。

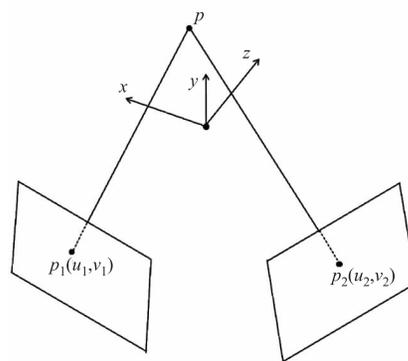


图 1 相机成像示意图

Fig. 1 Schematic diagram of camera imaging

2.2 RANSAC 求解 PnP

首先由深度相机获取两帧图像为 F1、F2。在这两帧图像中分别提取 SIFT 特征^[9]。SIFT 特征点由关键点和描述子组成,关键点记录了特征点的位置坐标,描述子由关键点周围像素的梯度方向直方图生成。匹配 SIFT 特征时,采用描述子的

欧氏距离 d 作为两幅图像中特征点的相似性判断标准。利用 FLANN (Fast Library for Approximate Nearest Neighbors)^[10] 方法对两个点集中的元素进行匹配,找到图像 F1 中特征点 p_1^c 在图像 F2 中的最佳的匹配点 p_2^c 。在这些匹配的特征点对中含有很多错误的匹配,本文剔除大于 $4d_{\min}$ (匹配对的最小欧式距离),以提高正确匹配点的比例。初步筛选后,依然会存在误匹配点,因此需要采用 RANSAC 做进一步筛选。假设剩余 n 组初始匹配点为:

$$P_1 = \{p_1^1, p_1^2, p_1^3, \dots, p_1^n\} \in F_1$$

$$P_2 = \{p_2^1, p_2^2, p_2^3, \dots, p_2^n\} \in F_2$$

则利用 RANSAC 算法剔除误匹配点,求解 PnP 的过程为:

①在集合 P_1 、 P_2 中随机抽取 3 对匹配的特征点。

②根据方程(3)求解 R 和 t 。

③遍历集合中其余点是否符合以上模型,即计算 $\|R x_i^c + t - x_i^c\|$ 是否小于设定的阈值。统计符合以上模型的点数,即为内点数。

④重复以上步骤,直到达到最大迭代次数或获得所需内点数。

⑤选择内点数最多的模型,用式(4)求解位姿参数。

3 本文方法

经过对两幅图像 F1、F2 中的 SIFT 特征点进行匹配,目的是找到空间中同一点在两幅图像中的投影 p_1^c 和 p_2^c 。若匹配出现错误,其在深度图像中相对某一正确匹配的点 p_1^n 和 p_2^n 的绝对距离必然发生变化。如图 2 所示,种子点为 p_1^n 和 p_2^n ,正确匹配的特征点 p_1^c 和 p_2^c 到种子点的空间绝对距离 l_1^n 和 l_2^n 相等;而错误的匹配 p_1^c 和 p_2^c 到种子点的距离 l_1^{2n} 和 l_2^{2n} 不会相等(极端的对称情况因为发生的概率很低,暂时不予考虑)。由于观测误差的存在,可以设定一个阈值进行筛选。以 p_1^x 和 p_2^x 为例,定义匹配准则:

$$l_{xn} = |l_1^{2n} - l_2^{2n}| / (l_1^{2n} + l_2^{2n}) \quad (5)$$

显然,在无测量误差时,若 p_1^x 与 p_2^x 为正确匹配点,则 $l_{xn} = 0$ 。因此,可设定阈值 T ,若 $l_{xn} < T$ 则认为 p_1^x 和 p_2^x 匹配正确,否则认定匹配错误。由此可见,可以利用正确匹配的种子点剔除误匹配点,但是需要保证种子点匹配是完全正确的。

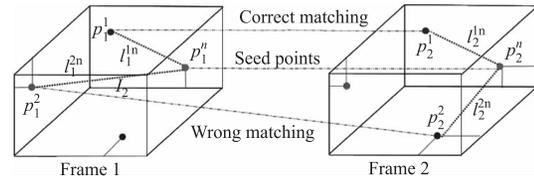


图 2 三维空间中的特征点

Fig. 2 The distribution of features in the 3D space

本文的方法基于两个假设条件:

(1)特征点描述子的距离越近,正确匹配的概率越高。

(2)当选择的种子点匹配错误时,根据 $l_{xn} < T$ 筛选留下的匹配特征点会很少。因为正确的匹配点与错误匹配点的距离不会相等,而错误匹配的特征点与错误种子点具有相同欧式距离的几率很低。

本文筛选错误匹配点的算法流程如下:

①选择描述子矢量距离 d 最小的 2 对特征点 $\{p_1^1, p_2^1\}$ 和 $\{p_1^2, p_2^2\}$ 作为种子点。

②遍历初始匹配点集,按式(5)分别计算特征点对相对于种子点 $\{p_1^1, p_2^1\}$ 和 $\{p_1^2, p_2^2\}$ 的距离匹配度量值。得到集合 $L_1 = \{l_1^{11}, l_1^{21}, l_1^{31}, \dots, l_1^{n1}\}$ 和 $L_2 = \{l_2^{12}, l_2^{22}, l_2^{32}, \dots, l_2^{n2}\}$ 。

③设定阈值 T 对 L_1 和 L_2 进行筛选,并对筛选后的集合求交集,若交集的元素数量小于初始点集的 10%,则重复步骤①选取新的种子点,并对筛选集合两两求交集。直到交集点数目大于 10%,或者种子点数等于 4。则输出最大交集点,进入步骤④。

④在交集点中,任意选择一对特征点作为种子点,对匹配点按几何距离度量再做一次筛选。如果筛选后的点数小于 90%,则舍弃该对种子点,再次选择。

⑤将最终筛选得到的特征点代入公式(4)求解 R, t 。

需要说明的是:

(a)在实际的算法中,建议 T 的取值为 0.05 ~ 0.15 之间,在本文的实验中选取 $T = 0.1$ 。

(b)步骤③是为了验证种子点是否选择正确,实验结果表明,如果种子点选取个数超过 4 个,则此时初始匹配点中存在极高的误匹配率,RANSAC 和本文算法都不能得到正确的结果。

(c)步骤④是为了避免出现两个不同点距离种子点距离相同的极端情况。即在空间中与两个种子点具有相同距离的点集是一个圆,而不是唯一确定

的一个点。当加入第三个内部种子点时,可以唯一确定正确的匹配点,这与 GPS 的绝对定位原理相同。

4 实验结果与分析

实验的计算机配置: Intel (R) Core (TM) i5 - 3210M CPU 2.20GHz, 2GB, Ubuntu 虚拟机下。实验图像数据来自 nyuv2 公开数据库 (http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html) 中 Kinect 深度相机采集的图像。Kinect 提供了两幅图像: RGB 图像和深度图像, 本文采用文献[11]中的方法对 RGB 图像和深度图像进行配准。

实验中, RANSAC 算法采用 OpenCV 中的 solvePnPRansac() 函数实现。其中的三个主要输入参数: 限制的最大迭代次数: 100; 阈值: 1.0; 最小内点数: 100。这里的迭代次数和内点数量限制都是根据实践经验设置的, 如果二者选取过小, 算法收敛于正确结果的概率会降低。由于 100 个内点在本文实验中无法得到满足, 故实验基本上都是迭代了 100 次, 平均时耗稳定。

实验中选择三组图像, 分别为相机位姿变化较小、适中和较大三种情况。分别应用 RANSAC 算法和本文算法对这三组图像进行处理。实验图像分别如图 3、图 4 和图 5 所示。



图 3 实验 1: 小位姿变化

Fig. 3 Experiment 1: Small pose change

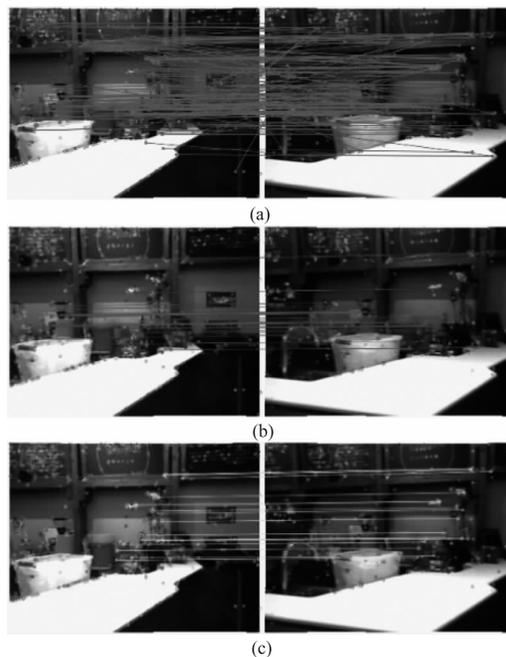


图 4 实验 2: 适中位姿变化

Fig. 4 Experiment 2: Moderate posture change

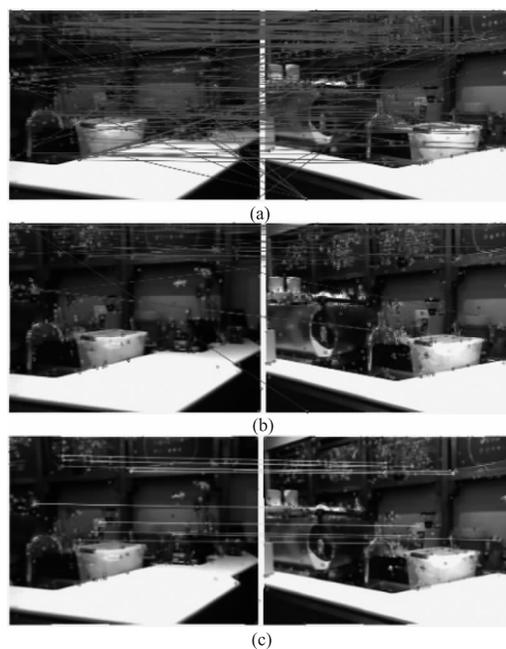


图 5 实验 3: 较大位姿变化

Fig. 5 Experiment 3: Large pose change

图 3~5 中, (a) 为对 FLAAN 算法匹配结果进行初步筛选的结果, 实验一和实验二选择 $4d_{\min}$, 实验三由于 $4d_{\min}$ 筛选后匹配点对太少, 因此选择 $10d_{\min}$ 筛选。图中连接线的两端为两幅视图之间的匹配点; (b) 为 RANSAC 算法的筛选结果; (c) 为本文算法筛选结果。

三组实验的运行时间和计算结果如图表 1 所

示。表中的真实值(True value)为经过 SLAM 系统后端优化后的结果,Inliers 为算得到的内点数量,运行时间为程序运行 200 次的平均统计结果。相机位移的测量误差(Transl. Error)为位移矢量(Transla-

tion vector)和真实值的矢量距离,将三个欧拉角(Euler angle,角度制)Pitch、Yaw 和 Roll 看作一个向量,计算该向量和真实值的矢量距离即为相机旋转的测量误差(Rot. Error)。

表 1 实验结果

Tab. 1 Results of experiment

Experiments	Methods	Inliers	Runtime/ms	Transl. vector/m	Euler angle/($^{\circ}$)	Transl. Error	Rot. Error
1	True value			[0.020,0.022,-0.003]	[0.559,1.420,-1.385]		
	Ransac	16	27.849	[0.040,0.006,-0.016]	[0.639,1.521,-1.433]	0.036688	0.616125
	Proposed algorithm	19	3.392	[0.046,0.006,-0.024]	[0.559,1.420,-1.385]	0.029636	0.668919
2	True value			[0.097,0.001,-0.008]	[1.439,10.736,0.925]		
	Ransac	12	31.591	[0.088,0.002,-0.015]	[1.012,10.938,0.918]	0.011146	0.472804
	Proposed algorithm	12	2.453	[0.088,0.002,-0.015]	[1.012,10.938,0.918]	0.011146	0.472804
3	True value			[0.105,0.016,-0.052]	[1.930,16.083,-0.059]		
	Ransac	24	30.649	[1.723,0.668,-0.016]	[10.444,47.429,-19.461]	1.745865	37.8343
	Proposed algorithm	15	2.573	[0.139,0.028,-0.046]	[1.107,15.274,-1.520]	0.037098	1.861073

由表 1 可知,当相机位姿变化不大时(实验 1 和实验 2),本文算法与 RANSAC 获取基本一致的计算精度,但是在计算效率上,本文算法明显优于 RANSAC,计算效率提高了 8 倍以上。在实验 1 中,两种算法得到了不一样的内点数,其主要原因是因为 RANSAC 的投影误差参数与本文算法筛选的距离度量值,不具有完全等价的几何意义,导致某些点在本文算法中认为是内点,而在 RANSAC 算法则认为是在外点。从最终结果看,这些点并不太影响最终的估计结果。

在实验 3 中,由于相机位姿变化较大,初始匹配中外点比例较大,RANSAC 算法出现了错误,其主要原因是因为 solvePnP($\text{Ransac}()$)函数筛选时,是从图像点出发计算投影误差是否符合模型,而图像点对应的三维空间点并不唯一,因此出现如图 3(b)所示的错误。而本文则直接应用空间点的三维距离对外点进行筛选,取得了比较好的效果。故针对深度相机的视觉 SLAM 而言,本文提出的方法在一定程度上增强了系统鲁棒性。

5 结论

针对深度相机的姿态估计,本文提出了一种根据空间绝对距离筛选特征匹配的方法。将特征点描述匹配的距离信息,应用于后续的错误匹配筛选中。由此避免了迭代运算,提高了算法的运行速度。实验结果表明,和传统算法相比,运

行时间提高了 8 倍以上。另外,由于场景中两点的距离与相机的转动和平移无关,因此,当相机转动角度较大时,算法的鲁棒性和准确性也优于传统的 RANSAC 迭代算法。最后,本文中提出的空间绝对距离筛选的方法,也可很容易推广到其他深度获取设备中,如双目相机、RGB_D 摄像机和激光雷达等。

参考文献:

- [1] M Kaess, A Ranganathan, F Dellaert. iSAM: Incremental smoothing and mapping [J]. IEEE Trans. on Robotics, 2008, 24(6): 1365-1378.
- [2] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: part I [J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99-110.
- [3] WU Mengqi, LI Zhongwei, ZHONG Kai, et al. Adaptive point cloud registration method based on geometric features and photometric features [J]. Acta Optica Sinica, 2015, 35(2): 229-236. (in Chinese)
伍梦琦, 李中伟, 钟凯, 等. 基于几何特征和图像特征点云自适应拼接方法 [J]. 光学学报, 2015, 35(2): 229-236.
- [4] Carlone L, Aragues R, Castellanos J A, et al. A fast and accurate approximation for planar pose graph optimization [J]. International Journal of Robotics Research, 2014, 33(7): 965-987.

- [5] LI Xin,ZHANG Yueqiang,LIU Jinbo,et al. A direct least squares method for camera pose estimation based on straight line segment correspondences [J]. Acta Optica Sinica,2015,35(6):203-213. (in Chinese)
李鑫,张跃强,刘进博,等. 基于直线段对应的相机位姿估计直接最小二乘法[J]. 光学学报,2015,35(6):203-213.
- [6] Fischler M A,Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. Communications of the Acm,1981,24(6):381-395.
- [7] LI Xin, LONG Gucan, LIU Jinbo, et al. Accelerative orthogonal iteration algorithm for camera pose estimation [J]. Acta Optica Sinica,2015,35(1):0115004. (in Chinese)
李鑫,龙古灿,刘进博,等. 相机位姿估计的加速正交迭代算法[J]. 光学学报,2015,35(1):0115004.
- [8] Endres F,Hess J,Engelhard N,et al. An evaluation of the RGB-D SLAM system [C]. IEEE International Conference on Robotics and Automation. IEEE,2012:1691-1696.
- [9] Lowe D G, Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision,2004,60(2):91-110.
- [10] Plimpton S. Fast parallel algorithms for short-range molecular dynamics [J]. Journal of Computational Physics, 1995,117(1):1-19.
- [11] Lachat E,Macher H,Mittet M A, et al. First experiences with kinect v2 sensor for close range 3d modelling [J]. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015, XL-5/W4(5):93-100.