

基于 Kinect 的 SLAM 方法

李海洋¹, 李洪波², 林颖¹, 刘贺平¹, 杨东方²

(1. 北京科技大学 自动化学院, 北京, 100083;

2. 清华大学 计算机科学与技术系, 智能技术与系统国家重点实验室, 北京, 100084)

摘要: 利用逆深度参数化方法研究 Kinect 摄像机在未知环境下的同时定位与地图构建问题。利用 SURF 特征提取和扩展卡尔曼滤波实现摄像机的位置和环境地图的更新, 从而提高算法的精确度和鲁棒性; 在此基础上有效融合 Kinect 摄像机采集的图像深度信息, 加快滤波算法的收敛速度。实验结果证明了所提方法的有效性和可用性。

关键词: 逆深度; SLAM; 扩展卡尔曼滤波; SURF 特征

中图分类号: TP302.1

文献标志码: A

文章编号: 1672-7207(2013)S2-0028-05

Method of SLAM with Kinect

LI Haiyang¹, LI Hongbo², LIN Ying¹, LIU Heping¹, YANG Dongfang²

(1. School of Automation and Electrical Engineering, University of Science and Technology Beijing,
Beijing 100083, China;

2. State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science & Technology,
Tsinghua University, Beijing 100084, China)

Abstract: A method for Kinect sensor was presented based on SLAM (simultaneous localization and mapping) problem via inverse depth parameterization model, where the depth of features from Kinect was employed to improve the convergence of method. To further improve the accuracy and robustness, SURF (speed up robust features) feature detection operator was adopted to extract image features, and extended Kalman filter (EKF) was used to estimate the trajectory of camera and the position of features. Experimental results are given to demonstrate the effectiveness and applicability of the proposed method.

Key words: inverse depth parameterization; SLAM; EKF; SURF extraction

同时定位与构图(Simultaneous Localization and Mapping, 简称为 SLAM)的主要目的是利用机器人自身携带的传感器获取环境信息估计机器人位置并构建其环境地图^[1-4], 从而实现机器人的自主运动。由于所用的传感器都存在着测量噪声, 因此, 定位与构图无论哪一方存在误差都会影响对方的精确度, 即存在定位与构图互相依赖的问题。传统的 SLAM 算法常采用

激光探测仪或声纳传感器等来测量环境的深度信息, 但是, 由于这些传感器的价格较贵在用于一般场合的机器人上性价比不高。随着运动恢复结构 SFM (structure from motion)理论^[5-7]研究的发展, 视觉传感器得到了人们的广泛关注并取得了一定的进展。相对于传统传感器, 视觉传感器很容易买到而且价格便宜, 另外, 采集的图像可以提供更多的环境信息。单目摄

收稿日期: 2013-03-01; 修回日期: 2013-05-02

基金项目: 国家重点基础研究发展计划(“973”计划)项目(2012CB821206); 国家自然科学基金资助项目(61004021, 61174069, 61004023); 北京市自然科学基金资助项目(4122037)

通信作者: 李洪波(1980-), 男, 辽宁朝阳人, 助理研究员, 从事网络控制与机器人研究; 电话: 010-62777704; E-mail: hbli@mail.tsinghua.edu.cn

像是通过透视投影原理, 一般采用针孔摄像机模型来将环境信息映射到摄像机的成像平面实现将环境的 3-D 坐标变换成 2-D 图像坐标^[8-10]。由于在转换过程中丢失了一个参数, 所以, 只能获取环境特征的方位信息而无法获得环境的深度信息^[11-12]。可以通过“视差”法来获取深度信息解决这个问题。目前基于逆深度参数化方法^[13]可以很好地解决这方面的问题, 并通过结合滤波算法来实现摄像机位置和环境地图的预测和更新。在地图特征的提取中, 大多数视觉 SLAM 都基于特征点提取的方法^[14], 并根据数据关联算法实现特征的匹配^[15], 提高 SLAM 的精度。最近, 微软公司推出的 Kinect 相机可以同时提供环境的彩色图像和对应的深度信息, 对 SLAM 研究来说非常方便而且性价比也较高。在基于 Kinect 的 SLAM 研究方面也引起了许多学者的广泛关注^[16]。本文作者在 Javier 等的逆深度参数化方法基础上提出采用 SURF(speed up robust features)提取算法来提高特征点的匹配精度, 在特征点初始化阶段融合 Kinect 采集的深度信息, 作为各特征点的深度加快滤波算法的收敛速度, 通过扩展卡尔曼滤波(EKF)实现摄像机的位置估计和地图更新。最后, 通过实验对比验证本文提出的方法的有效性。

1 系统模型

本文采用逆深度参数化方法, 并结合扩展卡尔曼滤波算法实现各参数的估计。

1.1 状态方程

摄像机的全状态向量表示如下^[13]:

$$\mathbf{X} = (X_v^T, y_1^T, y_2^T, \dots, y_n^T)^T \quad (1)$$

其中: $\mathbf{X}_v = (r^W, \mathbf{q}^{WC}, v^W, w^W)^T$ 为摄像机的运动向量。相应的状态更新方程为

$$\mathbf{X}_v = \begin{bmatrix} r_{k+1}^{WC} \\ q_{k+1}^{WC} \\ v_{k+1}^W \\ \omega_{k+1}^C \end{bmatrix} = \begin{bmatrix} r_k^{WC} + (v_k^W + V_k^W)\Delta t \\ q_k^{WC} \times q((\omega_k^C + W^C)\Delta t) \\ v_k^W + V^W \\ \omega_k^C + W^C \end{bmatrix} \quad (2)$$

其中: r^W 代表摄像机在世界坐标系下的位置; \mathbf{q}^{WC} 表示用四元数表示的摄像机旋转向量, $\mathbf{q}^{WC} = (q_0, q_x, q_y, q_z)$; v^W 和 w^W 分别表示摄像机的线速度和角速度; $q((\omega_k^C + W^C)\Delta t)$ 是旋转向量 $(\omega_k^C + W^C)\Delta t$ 的四元数表示; y_i 表示特征点在世界坐标系下的位置。在逆深度模型下, 该坐标由 6 个参数表示, 即

$$y_i = (x_i, y_i, z_i, \theta_i, \phi_i, \rho_i)^T = (x_i, y_i, z_i)^T + \frac{1}{\rho_i} \mathbf{m}(\theta_i, \phi_i) \quad (3)$$

其中: (x_i, y_i, z_i) 是摄像机第 1 次观测到特征点时的位置坐标; $\mathbf{m}(\theta_i, \phi_i)$ 是单位方向向量; $\mathbf{m}(\theta_i, \phi_i) = (\cos \phi_i \sin \theta_i, \sin \phi_i, \cos \phi_i \cos \theta_i)^T$, θ_i 和 ϕ_i 分别为摄像机在世界坐标系下的偏转角和俯仰角; ρ_i 为沿摄像机光束方向光心到特征点之间的距离 d_i 的倒数, 即 $\rho_i = 1/d_i$, 如图 1 所示。

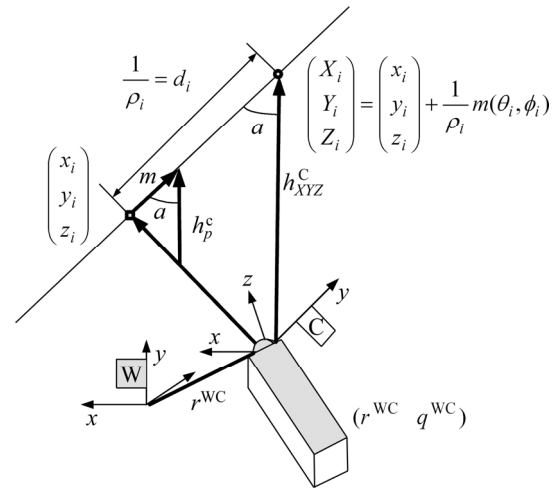


图 1 逆深度参数化模型

Fig.1 Inversed depth parameterization model

1.2 测量方程

根据图 1 所示, 当摄像机移动到一个新的位置 r^{WC} 和新的旋转方向 q^{WC} 时, 从摄像机到同一特征点 y_i 的方向向量 h^C 可如下表示^[13],

$$h^C = (h_x^C, h_y^C, h_z^C)^T = R^{CW} \left\{ (x_i, y_i, z_i)^T + \frac{1}{\rho_i} \mathbf{m}(\theta_i, \phi_i) - r^{WC} \right\} = R^{CW} \left\{ \rho_i \{ (x_i, y_i, z_i)^T - r^{WC} \} + \mathbf{m}(\theta_i, \phi_i) \right\} \quad (4)$$

当特征点处于无穷远时, 方向向量可以近似为 $h^C = R^{CW} \mathbf{m}(\theta_i, \phi_i)$ 。

根据针孔摄像机模型原理, 特征点的三维坐标可以描述成

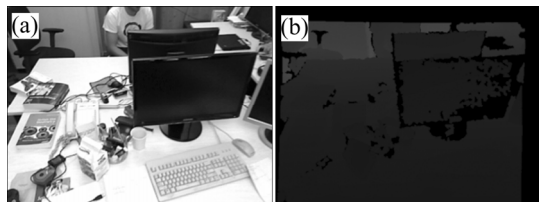
$$h = (u, v)^T = \left(u_0 - \frac{f_x}{d_x} \frac{h_x^C}{h_z^C}, v_0 - \frac{f_y}{d_y} \frac{h_y^C}{h_z^C} \right)^T \quad (5)$$

其中: (u_0, v_0) 是摄像机光心坐标; f_x 和 f_y 是摄像机在水平和垂直方向上的焦距; d_x 和 d_y 分别为每一个像素在水平和垂直方向上的物理尺寸, 这些值都是通过对摄

像机标定而得到的。

1.3 Kinect 获取特征点的深度信息

Kinect 传感器由 IR 发射器、IR 相机和标准 RGB 相机组成,通过 IR 发射器和 IR 相机对空间点做三角测量,可以获得对应 RGB 图像以及每个像素点的深度信息,见图 2。



(a) 彩色图像; (b) 深度图像

图 2 Kinect 采集的彩色图像和深度图像

Fig.2 RGB image and depth image from Kinect

利用 Kinect 深度信息可以求得摄像机光心到特征点的距离,如图 3 所示。 P 点为 Kinect 摄像机坐标系下的特征点,其三维坐标为 $P(x_c, y_c, z_c)$,其中 z_c 为 Kinect 相机采集到的特征点深度信息,深度可以通过 Kinect 的深度图像中对对应点的像素值获得,几何意义为特征点到摄像机光心所在平面的垂直距离。根据图 3 的投影原理图可知:利用三角关系可以求出特征点的逆深度值,即特征点到摄像机光心距离的倒数 $1/d$,其中 $d^2 = x_c^2 + y_c^2 + z_c^2$,在算法运行过程中需要添加并初始化新特征点时可以直接将计算得到的特征点的逆深度值代入 EKF 滤波循环中,由于 EKF 滤波对初始值的依赖性很大,所以,通过 Kinect 采集的深度值可以很大程度上提高 EKF 的收敛速度。

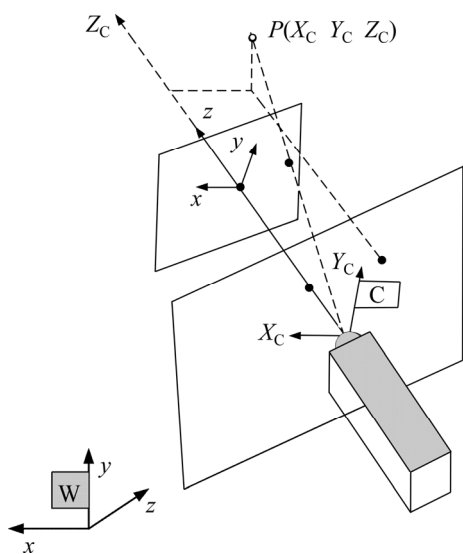


图 3 特征点在摄像机的投影原理

Fig.3 Principle of camera projection

2 特征提取算法

Harris 角点提取方法主要是基于图像梯度提取角点检测算子,并结合归一化互相关实现两幅图像中角点的匹配,该方法的缺点是算法鲁棒性不强,而且,角点提取的多少以及角点质量过分依赖于阈值的设定。而 SURF 特征提取则可以有效地克服上面的缺点,也是本文选择 SURF 提取的原因。SURF 算法的特征点提取和匹配结果如图 4 所示。



图 4 SURF 特征点匹配

Fig.4 Feature extraction and match with SURF

3 实验结果分析

采用 Computer Vision Group 网站提供的 Kinect 数据集作为实验数据,该数据集中包含了 Kinect 摄像机采集的彩色和深度图像以及摄像机运动轨迹数据,其中摄像机轨迹数据通过 8 个高速跟踪传感器获得,精确度高,该数据可以近似为摄像机真实的运动轨迹值。数据集中包含了多种环境下采集到的数据,结合本实验特点本文采用 `rgbd_dataset_freiburg1_xyz` 数据集作为实验数据,实验结果如图 5 和图 6 所示。

3.1 匹配精度和收敛速度分析

图 5 和图 6 所示均为第 230 帧图像时的实验结果。图 5(a)和 6(a)中椭圆代表正确匹配的点,三角形代表被系统拒绝的点,正方形代表无法被匹配的点。图 5(b)和 6(b)所示箭头表示未匹配点的方向,细长的椭圆表示特征点估计的不确定性程度,越小表示越接近真实值,当不确定程度减小时箭头变为椭圆。从数据集中选出第 60 帧到第 280 帧的图像,该过程是一个摄像机旋转运动过程。

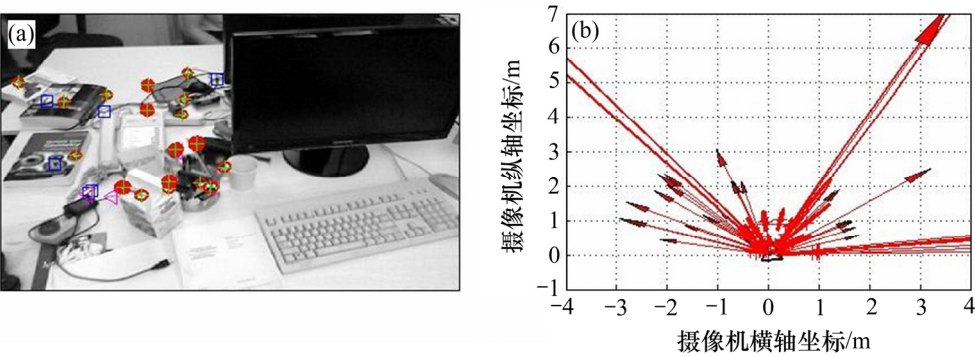
由图 5 和图 6 可以看出:在单纯利用逆深度法的情况下存在许多无法匹配上的特征点,这是由于 Harris 角点特征固有缺陷导致的,而使用 SURF 算法匹配的特征点要精确得多,而且不需要太多的匹配点

就可以完成定位功能。另外, 在摄像机运动的过程中会有新特征点被检测到, 对新特征点需要进行初始化处理, 图 6 所示为融合了 Kinect 采集的特征点深度信息后的结果, 与直接逆深度法相比 EKF 滤波的收敛速度有了明显的提高。

3.2 定位精度分析

表 1 所示为在不同帧时, 逆深度法和本文提出的改进方法的摄像机运动轨迹估计值与摄像机真实轨迹值的比较。

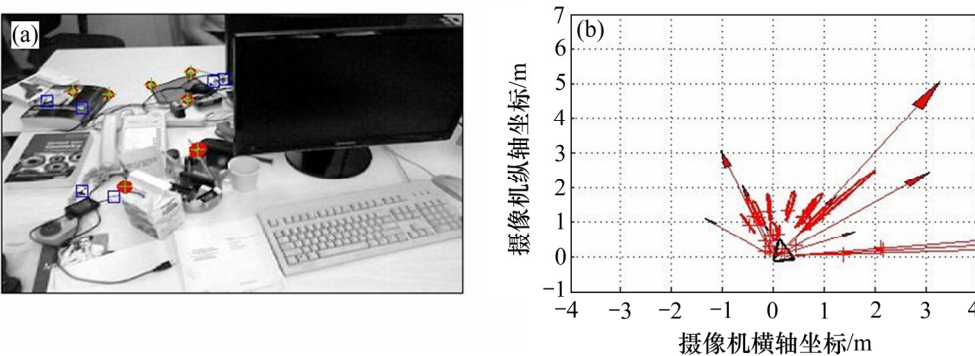
从表 1 可以看出: 在第 70 帧时摄像机轨迹的真实值为(1.168 2, 0.623 4, 1.422 8), 相应的逆深度法估计值为(0.584 2, 0.294 9, 0.915 6), 其对应的均方误差为 0.840 4; 计算得改进的逆深度法估计值为(0.825 3, 0.893 3, 1.586 7), 其对应的均方误差为 0.466 1, 计算可得改进的方法要比直接采用逆深度法在精确度上提升了 44.5%。同时又分别对比了另外 3 个时刻的真实值与估计值。第 90 帧时直接逆深度法与改进法的均方误差分别为 0.471 7 和 0.274 6; 第 130 帧时分别为



(a) Harris 特征点提取结果; (b) 地图特征不确定性

图 5 直接逆深度方法

Fig.5 Unified inverse depth parameterization



(a) SURT 特征点提取结果; (b) 地图特征不确定性

图 6 所提方法的实验结果

Fig.6 Experimental result with proposed method

表 1 实验结果比较

Table 1 Comparative results

图像帧号	实际值/m	逆深度法		改进法	
		估计值/m	均方差/m	估计值/m	均方差/m
70	(1.168 2, 0.623 4, 1.422 8)	(0.584 2, 0.294 9, 0.915 6)	0.840 4	(0.825 3, 0.893 3, 1.586 7)	0.466 1
90	(1.121 3, 0.632 1, 1.369 7)	(0.762 3, 0.385 0, 1.189 3)	0.471 7	(1.384 6, 0.709 9, 1.373 5)	0.274 6
130	(1.085 4, 0.644 6, 1.323 6)	(0.995 4, 0.542 8, 0.363 4)	0.969 8	(1.112 4, 0.903 0, 1.085 9)	0.351 2
200	(1.408 3, 0.637 5, 1.736 4)	(1.071 7, 0.623 4, 0.426 2)	1.352 8	(1.494 9, 0.537 9, 1.200 9)	0.551 5

0.969 8 和 0.351 2; 第 200 帧时分别为 1.352 8 和 0.551 5。

从结果可以看出: 虽然改进的逆深度法的估计值稍有些波动, 但精确度要远高于直接使用逆深度法的估计值。

4 结论

提出了一种基于 Kinect 的 SLAM 方法。所提方法基于扩展卡尔曼滤波算法实现了摄像机的位置估计和地图特征的更新。在此框架下, 采用逆深度参数化和 SURF 特征提取方法来提高了视觉 SLAM 的鲁棒性, 利用 Kinect 传感器的采集的深度信息提高了滤波的收敛速度。实验结果证明了所提方法的有效性和可用性。在后续的工作中, 将进一步致力于提高所提方法的实时性和精确性。

参考文献:

- [1] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052–1067.
- [2] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping[J]. IEEE Robotics and Automation Magazine, 2006, 13(2): 99–110.
- [3] Dissanayake G, Newman P, Clark S, et al. A solution to the simultaneous localization and map building (SLAM) problem[J]. IEEE Transactions on Robotics and Automation, 2001, 17(3): 229–241.
- [4] 迟健男, 徐心和. 移动机器人即时定位与地图创建问题研究[J]. 机器人, 2004, 26(1): 92–96.
CHI Jannan, XU Xinhe. Research on simultaneous localization and mapping of mobile robot[J]. Robotics, 2004, 26(1): 92–96.
- [5] Tomasi C, Kanade T. Shape and motion from image streams under orthography: A factorization method[J]. International Journal of Computer Vision, 1992, 9(2): 137–154.
- [6] Sturm P, Triggs T. A Factorization based algorithm for multi-image projective structure and motion[C]// Pro ECCV. Cambridge, UK, 1996: 709–720.
- [7] Morita T, Kanade T. A sequential factorization method for recovering shape and motion from image streams[J]. IEEE Transactions on Pattern and Analysis and Machine Intelligence, 1997, 19(8): 858–867.
- [8] Davison A J, Murray D W. Simultaneous localization and map-building using active vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 865–880.
- [9] Rusinkiewicz S, Levoy M. Efficient variants of the icp algorithm[C]// Proc of the 3rd Int Conf on 3-D Digital Imaging and Modeling. Quebec City, Que: IEEE, 2001: 145–152.
- [10] Jin H, Favaro P, Saototo S. Real-time 3d motion and structure of point features: A front-end system for vision-based control and interaction[C]// Proc of IEEE Int Conf on Computer Vision and Pattern Recognition (CVPR). Hilton Head Island, SC, 2000: 778–779.
- [11] Lemaire T, Lacroix S, Sola J. A practical 3d bearing-only slam algorithm[C]// Proceedings of IROS. Alberta, Canada, 2005: 2449–2454.
- [12] Cummins M, Newman P. Fab-map: Probabilistic localization and mapping in the space of appearance[J]. The International J. Robotics Research, 2008, 27(6): 674–665.
- [13] Montiel J M M, Civera J, Davison A J. Inverse depth parameterization for monocular SLAM[J]. IEEE Transactions on Robotics, 2008, 24(5): 923–945.
- [14] Bay H, Ess A, Tuytelaars T, et al. SURF: Speed up robust features[J]. Computer Vision and Image Understanding (CVIU), 2008, 110(9): 346–359.
- [15] Neira J, Tardos J. Data association in stochastic mapping using the joint compatibility test[J]. IEEE Transactions on Robotics and Automation, 2001, 17(6): 890–897.
- [16] The Willow Garage website. Available[EB/OL]. [2010–01–01]. <http://www.sidekick.co.il/>

(编辑 杨幼平)