

一种基于双目 SLAM 的环境建模方法

张釜恺, 芮挺, 杨成松, 王东

(陆军工程大学 野战工程学院, 江苏 南京 210007)

摘要:利用计算机视觉对环境进行建模是机器人环境感知相关研究的一项重要工作。提出了一种基于双目 SLAM 的建模方法。该方法的主要思想是利用双目 SLAM (Simultaneous Localization and Mapping, 同步定位与地图构建) 技术获取相机位姿, 并利用双目相机直接获取场景的深度图, 然后将二者融合成场景的点云地图。由于双目相机直接获取的深度图像误差点较多, 因此提出了一种基于视线约束的点云滤波方法, 来消除场景点云中的误差点。经过实验验证了建模的效果, 以及视线约束点云滤波的有效性。

关键词: 双目视觉; 视觉 SLAM; 环境建模; 视线约束; 点云滤波

中图分类号: TP24

文献标识码: A

DOI: 10.19358/j.issn.2096-5133.2020.01.009

引用格式: 张釜恺, 芮挺, 杨成松, 等. 一种基于双目 SLAM 的环境建模方法 [J]. 信息技术与网络安全, 2020, 39(1): 50-54.

An environment modeling method based on stereo SLAM

Zhang Fukai, Rui Ting, Yang Chengsong, Wang Dong

(College of Field Engineering, Army Engineering University, Nanjing 210007, China)

Abstract: Modeling the environment using computer vision is an important task in the research of robot environment perception. This paper proposes a modeling method based on stereo SLAM. The main idea of this method is to use the stereo SLAM (Simultaneous Localization and Mapping) technology to obtain the camera pose, use the stereo camera to directly obtain the depth map of the scene, and then merge the two into the point cloud of the scene. Since there are many error points in the depth images directly acquired by stereo camera, this paper proposes a point cloud filtering method based on line-of-sight constraint to eliminate the error points in the scene point cloud. The experimental results verify the effectiveness of the modeling and the effectiveness of the line-of-sight constraint point cloud filtering.

Key words: stereo vision; visual SLAM; environment modeling; sight line constraint; point cloud filtering

0 引言

对于在未知环境中的自主导航, 近年来的研究热点是同步定位与地图构建, 即 SLAM 技术^[1]。SLAM 技术^[2]是指机器人从一个未知的位置出发, 利用自身的传感器建立周围地图, 并利用已经建立的地图实现自身的定位。对场景地图的构建或环境的建模是机器人 SLAM 的一项重要任务。对未知环境的建模一方面有利于对场景的理解, 进而指导机器人的行动; 另一方面, 在对未知环境的勘探、侦察等任务中, 环境整体的建模比单纯的视频信息能够给人提供更加直观的信息, 更便于后续的使用。

为了快速获取环境整体的模型, 本文采用深度相机直接获取深度图像, 进行建模。深度相机的原

理^[3]目前有三类, 分别为结构光原理、飞行时间测距法 (Time of Flight, ToF) 原理和双目视觉原理。相对于基于结构光或 ToF 原理的深度相机, 双目相机不需要主动向外界发送信息, 不容易受到阳光中红外线的干扰, 而且探测距离较远。然而, 虽然双目相机理论上可以获取其图像中所有像素的深度, 但是在实际使用中, 由于像素匹配误差的存在, 双目相机获取到的深度存在很多误差较大的点。

针对机器人的建模需求和双目相机建模尚存在的问题, 本文提出了一种新的基于双目相机 SLAM 的环境建模方法, 将双目 SLAM 获取的相机位姿与双目相机获取的深度图像相结合, 快速地构建场景模型。针对双目相机的误差问题, 提出了一

种新的基于视线约束的点云滤波方法对误差进行控制。

1 理论分析

相比较于单目相机,双目相机的最大优势是可以通过双目立体视觉获取图像中各像素距相机的真实距离,即像素深度。在理想情况下,通过双目 SLAM 技术,系统可以获取相机各关键帧的空间位置和姿态,再结合深度图像,系统就可以获得各关键帧图像中各像素的真实空间位置。将所有关键帧中像素的空间位置集合到一起,自然就获得了场景的稠密模型。这就是基于双目 SLAM 的环境建模的基本原理^[4]。

然而在实际的运行中,该方法存在两个主要问题:(1)双目相机获取的深度图像存在误差,甚至在一些位置误差会很大。这主要是由于室内场景的墙壁、地面、桌椅等大部分表面均十分平滑,纹理很少,不利于双目相机在左右两图像之间进行像素匹配。(2)不同关键帧的观察会存在重叠,深度图像也会存在重叠,造成一些位置被反复建模,出现大量的重复冗余点。针对第一个问题,本文提出了一种新的基于视线约束的点云滤波,其基本思想是相机与目标之间的空间不应该存在遮挡,如果有地图点存在于相机与目标之间,那这个地图点为误差点的概率就比较大。系统通过这种滤波方法可以滤除一大部分误差较大的点。针对第二个问题,采用体素滤波的方法,在一定的空间内只保留1个点来去除冗余点。

2 算法流程

本文的双目建模方法使用能够直接输出深度图像的双目相机,而不在本系统中进行双目匹配与深度图的计算,这可以节省大量的计算资源。本方法基于 ORB-SLAM2 系统的双目方案。首先将双目相机获取的深度图像转化为关键帧点云,然后依据双目 SLAM 系统获取到的关键帧位姿,将关键帧点云进行位姿变换后加入场景的点云中。由于双目 SLAM 中的关键帧位姿会进行调整,而且冗余的关键帧会被系统剔除,因此本系统在向场景点云中添加强键帧点云时,会滞后 5 个关键帧,而非在一个新的关键帧被创建时就使用它。

为了消除双目相机的误差点,本系统采用一种基于视线约束的滤波方法,将场景点云中明显的误差点进行消除。系统的框架如图 1 所示。

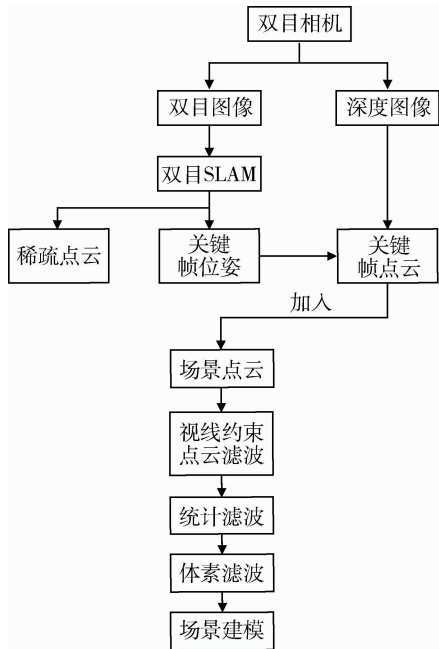


图1 基于双目 SLAM 的环境建模方法整体框架

3 双目 SLAM 模块

本方法的系统是基于 ORB-SLAM2 的双目方案^[5]建立的。ORB-SLAM2 的双目方案与前文中研究过的 ORB-SLAM 单目方案在基本思想和整体结构方面基本一致。但是在细节方面,针对双目相机的特点有一些改动,具体包括:

(1)在跟踪线程之前,对双目图像进行预处理,包括对双目图像各自提取 ORB 特征点,进行匹配后获取双目关键点。

(2)在地图初始化时,可以直接获得深度信息,不需要像单目方案一样使用运动恢复结构进行初始化。

(3)相比于单目 ORB-SLAM 方案中可能出现的尺度漂移,双目的信息将会使得尺度信息可观。并且,几何校验和位姿图优化将不再需要处理尺度漂移。

(4)由于双目相机可以获取特征点的深度,系统依据特征点的深度,以 40 倍基线距离为界将特征点分为了远、近两类。

(5)关于关键帧的插入,ORB-SLAM2 遵循单目 ORB-SLAM 中的法则,即经常插入关键帧并且剔除冗余的关键帧。远近特征点的差异为插入关键帧提供了新的条件。如果被跟踪的近特征点过少,并且当前帧将会创造新的近特征点,系统将会插入关键帧。

4 基于视线约束的点云滤波

基于视线约束来进行点云滤波的基本思想,脱胎于八叉树地图中的反向传感器模型^[6]。八叉树地图是一种描述三维模型的数据结构。八叉树地图将空间分为许多正方体立体元,而每一个立体元可以继续细分成8个小的立体元,直到达到设定的最小立体元尺寸。它们之间的关系又被称为父节点与子节点。最小的立体元大小即为八叉树的分辨率。在八叉树地图中,如果一个父节点下面的8个子节点全部处于占据或未被占据的状态,该父节点就不会被展开,只占据一个节点的存储空间。用八叉树地图对三维模型进行描述的最大优势是其数据量小,但是模型对于场景的描述较为粗略。

反向传感器模型^[7]是八叉树地图中用于描述各立体元被占用情况的模型。在此模型中,如图2所示,黑色是障碍物的位置,传感器 P 中心与障碍物连线上的立体元为灰色。在八叉树占有率地图的构建过程中,每次有新的深度图像输入时,都会将灰色的立体元被占有的概率降低,而将黑色立体元被占用的概率提高。

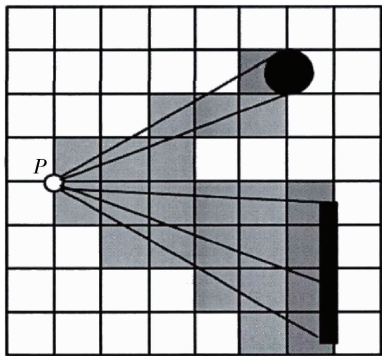


图2 八叉树地图中的反向传感器模型

反向传感器模型的基本思想是相机与目标之间的空间不应该存在遮挡,如果有地图点存在于相机与目标之间,那么这个地图点为误差点的概率就比较大。基于这种思想,本文针对点云模型,提出了基于视线约束的点云滤波方法。如图3所示,相机的位置为 P ,黑色点 Q 为深度图像上某一像素点对应的空间位置。根据相机投影模型,点云中位于灰色区域的点在图像上的投影位置为相同的像素。那么该区域中的点,如果距离相机的距离小于点 Q 的深度,就有很大概率为误差点。

该滤波的具体操作方法如下:

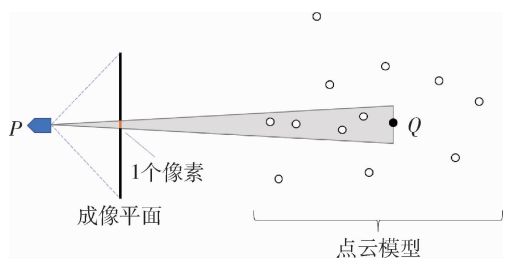


图3 基于视线约束的点云滤波

若系统创建了一个序号为 j 的新关键帧 F_j ,在向场景点云中加入该关键帧点云之前,首先依据其变换矩阵 T_j ,将场景点云中的点投影到该关键帧上。对于场景点云中序号为 i ,世界坐标为 P_i 的一点 p_i ,它在该关键帧上的投影点坐标 $Q_{i,j}$ 为:

$$Q_{i,j} = \frac{1}{Z_{i,j}} \mathbf{K} T_j P_i = [u_{i,j}, v_{i,j}] \quad (1)$$

其中, $Z_{i,j}$ 表示点在相机坐标系下的 Z 坐标, $\mathbf{K}$ 表示相机的内参矩阵。如果点投出了图像的边缘就略过该点。然后,获取该关键帧对应的深度图像在投影点的深度 $d(u_{i,j}, v_{i,j})$,如果深度图像在该点上没有深度也略过。

对于场景地图中的每一个点 p_i ,如果该点在关键帧 F_j 上的投影点在图像范围内,而且关键帧深度图像在该点有深度值 $d(u_{i,j}, v_{i,j})$,定义该点在此帧的不可靠度 $e_{i,j}$ 为:

$$e_{i,j} = \begin{cases} \frac{d(u_{i,j}, v_{i,j})}{Z_{i,j}} & Z_{i,j} < d(u_{i,j}, v_{i,j}) \\ 0 & Z_{i,j} \geq d(u_{i,j}, v_{i,j}) \end{cases} \quad (2)$$

不可靠度 $e_{i,j}$ 表示该点在此关键帧观察下的不可靠程度。如果该点在此帧相机坐标系下的 Z 坐标 $Z_{i,j}$ 小于该点的深度 $d(u_{i,j}, v_{i,j})$,就说明该点可能在双目相机观察到的景物与相机之间,而两者的差距越大,说明该点在该帧的不可靠程度越高。

设 E_i 为该点的总不可靠度。总误差度 E_i 等于该点在所有关键帧上的误差度之和:

$$E_i = \sum_j e_{i,j} \quad (3)$$

最后,在获取了所有点的总不可靠度之后,将所有不可靠度大于5的点都去除,即完成了基于视线约束的点云滤波。

在视线约束滤波结束后,再对点云进行统计滤波以消除离群点,进行体素滤波^[8]以消除冗余点,即在一定的空间内,点云模型只保留1个点来去除冗余点。最终得到了场景的稠密点云模型。

5 实验与分析

为了对本方法进行实验验证,使用双目摄像头获取数据,并将本文的系统运行在小型笔记本电脑上,其处理器为 Intel i5-4200H,内存 4 GB,搭载 Ubuntu 16.04 系统。双目摄像头使用美国 Stereo-labs 公司生产的 ZED 双目摄像头(图 4),它的有效距离可达 20 m,并且能够以 30 f/s 的帧率输出 1 080 p 分辨率的深度图像,或者以 100 f/s 的帧率输出 800 × 480 分辨率的深度图像,可以满足本系统的实时性需求。实验中,ZED 双目相机获得的部分深度图像与对应的 RGB 图像如图 5 所示。

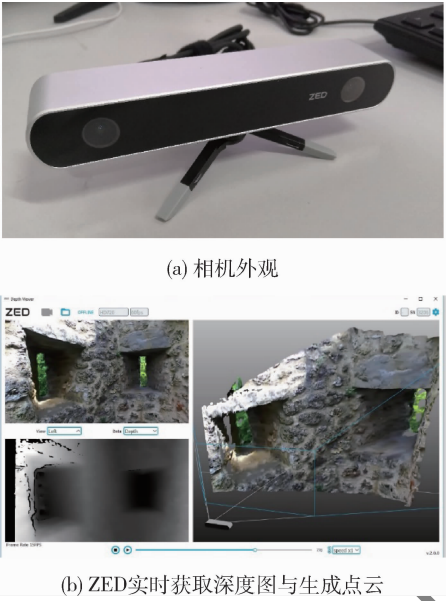


图 4 ZED 双目相机输出

在教室的场景(图 6)中,使用本文的方法对场景进行建模还原。首先,在不使用视线约束点云滤波,只使用统计滤波与体素滤波的情况下,得到的场景点云模型如图 7 所示。

在使用约束点云滤波之后,建模的效果如图 8 所示。

模型中间的空缺是双目相机没有拍到的部分。总体上看,本文的方法将房间的结构以及主要陈设进行了较为准确的还原,而且使用视线约束点云滤波去除了大量的误差点。但是在一些细节部分,受限于双目相机的精度,出现了波浪状的线条,建模效果仍有提升空间。

不使用视线约束点云滤波的模型与使用视线约束点云滤波的模型在房门附近部分的细节对比如图 9 所示。



图 5 ZED 双目相机输出图像

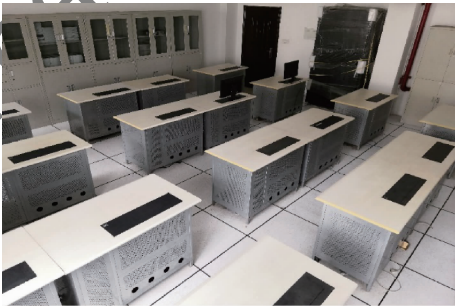


图 6 教室环境

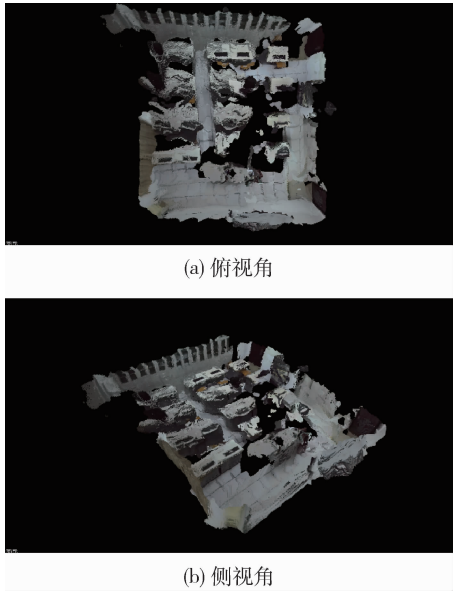
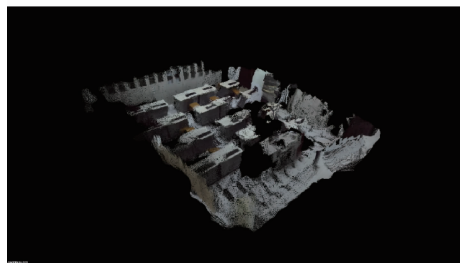


图 7 不使用视线约束点云滤波的场景点云模型



(a) 俯视角



(b) 侧视角

图8 使用视线约束点云滤波的场景点云模型



(a) 不使用



(b) 使用

图9 不使用与使用视线约束点云滤波的模型细节对比

从图9中可以发现,由于使用了视线约束点云滤波,圈起来的两处较为明显的误差点均得到了有效消除。这证明了本文使用的视线约束点云滤波的有效性。

6 结论

为了获取场景模型,本文使用双目相机,提出了一种基于双目SLAM的环境建模方法。该方法使用双目相机直接获取深度图像,并且根据双目SLAM系统获取的关键帧位姿,对深度图像进行融合,生成场景点云模型。针对双目相机深度图像在建模中的误差问题,本文提出了一种基于视线约束

的点云滤波方法。本文先依次对系统的理论进行分析,研究了系统框架与双目SLAM模块,并重点对提出的视线约束点云滤波方法进行了讲解。最后通过实验验证了该滤波策略的可行性。

参考文献

- [1] LIU H M, ZHANG G F, BAO H J. A survey of monocular simultaneous localization and mapping[J]. Journal of Computer-Aided Design & Computer Graphics, 2016.
- [2] 邸凯昌, 王文辉, 赵红颖, 等. 视觉SLAM技术的进展与应用[J]. 测绘学报, 2018, 47(6): 84-93.
- [3] 张翠红. 基于Kinect深度相机的室内三维场景重构[D]. 大连: 大连理工大学, 2013.
- [4] 余小欢. 基于双目立体视觉的微小型无人机的室内三维地图构建系统的设计与研究[D]. 杭州: 浙江大学, 2014.
- [5] MUR-ARTAL R, TARDOS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017: 1-8.
- [6] 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): 242-243.
- [7] HOOVEF A, OLSEN B D. A real-time occupancy map from multiple video streams[C]. Proceedings of the IEEE International Conference on Robotics and Automation, 1999, 3: 2261-2266.
- [8] TANG F F, LIU J N, ZHANG X H, et al. A voxel-based filtering algorithm for DTM data extraction in forest areas[J]. Journal of Beijing Forestry University, 2009, 57(3): 55-59.

(收稿日期: 2019-10-24)

作者简介:

张釜恺(1995-), 男, 硕士研究生, 主要研究方向: 计算机视觉、SLAM。

芮挺(1972-), 通信作者, 男, 博士, 教授, 主要研究方向: 计算机视觉、模式识别、人工智能。E-mail: 785344305@qq.com。

杨成松(1982-), 男, 博士, 讲师, 主要研究方向: 计算机视觉、模式识别。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部
中国电子信息产业集团有限公司第六研究所