文章编号:1001-5078(2017)10-1271-05 ·红外技术及应用·

基于视觉里程计的单目红外视频三维重建

陈博洋[1,2],孙韶媛[1,2],叶国林[1,2],赵海涛[3]
(1. 东华大学信息科学与技术学院,上海 201620;2. 东华大学 数字化纺织服装技术教育部工程研究中心,上海 201620;
3. 华东理工大学信息科学与工程学院,上海 200237)

摘 要:针对红外视频的特点,提出了一种基于直接法和稀疏法视觉里程计的单目红外视频三维重建方法。该方法首先通过对红外热像仪标定获得热像仪内参,然后构建直接法和稀疏法视觉里程计模型,视觉里程计前端执行帧管理和点管理的任务,利用滑动窗口并借助高斯-牛顿迭代对总光度误差进行优化,计算出直接法和稀疏法视觉里程计模型所依赖的所有变量,完成定位热像仪和建图的任务。通过实验证明了该方法能够实时实现对单目红外视频进行三维重建。

关键词:红外视频;三维重建;直接法和稀疏法视觉里程计
中图分类号:TP391.41 **文献标识码**:A **DOI**:10.3969/j.issn.1001-5078.2017.10.015

3D reconstruction of monocular infrared video based on visual odometer

CHEN Bo-yang[1,2], SUN Shao-yuan[1,2], YE Guo-lin[1,2], ZHAO Hai-tao[3]
(1. College of Information Science and Technology, Donghua University, Shanghai 201620, China; 2. Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China;
3. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: Aiming at the characteristics of infrared video, a three-dimensional reconstruction method of monocular infrared video based on direct and sparse visual odometer is proposed. In this method, the internal parameters of the thermal imager are obtained by the calibration of the infrared camera, and then the direct and sparse visual odometer model is constructed. Visual odometer front-end performs frame management and point management tasks. The total photometric error is optimized by the sliding window and the Gauss-Newton iteration, and all the variables that the direct and sparse visual odometer model needs are calculated. Finally, the tasks of locating the thermal imager and plotting are fullfilled. Experimental results show that the proposed method can achieve the three-dimensional reconstruction of monocular infrared video in real time.

Key words: infrared video; 3D reconstruction; direct sparse odometer

1 引 言

三维重建是运用图像处理等相关技术,使二维数据还原出三维信息,形成三维立体表面的先进技术,并且随着计算机技术广泛应用于生产生活等[1]。

目前在可见光领域,单目视觉三维重建主要有

基金项目:国家自然科学基金项目(No. 61375007);上海市科委基础研究项目(No. 15JC1400600)资助。
作者简介:陈博洋(1991-),男,硕士,主要研究方向为红外图像处理与模式识别。E-mail:chenboyang100@163.com
通讯作者:孙韶媛(1974-),女,教授,主要研究方向为夜视机器视觉。E-mail:shysun@dhu.edu.cn
收稿日期:2017-03-06

运动推算结构(structure frommotion,SFM)[2]和同时定位与地图构建(simultaneous localizationand mapping,SLAM)[3-4]两个分支。运动推算结构是通过在多幅图像序列中检测匹配特征点集,使用数值方法恢复场景三维结构的一种方法。同时定位与建图是机器人从未知环境的未知地点出发,在运动过程中通过重复观测到的地图特征定位自身位置和姿态,再根据自身位置增量式的构建地图,从而达到同时定位和地图构建的目的。Klein 等[5]于 2007 年提出基于特征点法的 PTAM,它的基本思想是通过检测匹配特征点并最小化重投影误差以优化相机位姿,并生成稀疏的点云,之后 Engel 等[6]提出了一套基于直接法的视觉测量系统,该系统后扩展为 LSD-SLAM[7],该算法不采取特征点检测与匹配的策略,直接对像素操作,能够生成复现场景的半稠密点云。

对于红外图像,因其有无彩色、缺乏纹理信息、对比度低、图像模糊等缺点[8],以至于目前还没有针对单目红外视频进行场景三维重建的方法。本文提出了一种基于直接法和稀疏法视觉里程计(Direct Sparse Odometry,DSO)[9]的单目红外视频三维重建方法。通过对红外热像仪标定获得热像仪内参,然后构建直接法和稀疏法视觉里程计模型,滑动窗口优化关键帧的总光度误差,利用高斯-牛顿法迭代求解出热像仪位姿和激活点的逆深度,实时生成稀疏的点云。图 1 是本文的算法思路框图。

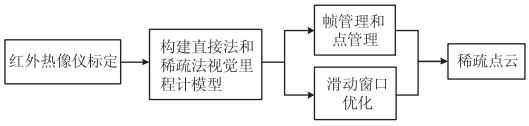

图 1 算法流程图

Fig. 1 Algorithm flow chart

2 直接法和稀疏法视觉里程计模型

2.1 红外热像仪标定

为了简单地描述红外热像仪的成像原理,将红外热像仪的成像看做是小孔成像。在算法的优化框架中,需要利用红外热像仪的内参进行变量的初始化,因此需要对红外热像仪进行标定,本文采用的是张氏标定法[10],标定板的制作精度直接影响着红外热像仪标定结果的准确性。为了获得高对比度的黑白相间的棋盘格红外图像,我们采用附加热源及隔热板的方案,附加热源选用装满温水的热水箱,隔热板选用可以隔热 1000 ℃高温的石棉板,石棉板的大小为 30 mm×30 mm,在硬纸板上打印针对普通彩色摄像头的棋盘标定板,将石棉板依次准确地固定在硬纸板上,再将硬纸板固定在装有温水的热水箱上。图 2 为红外热像仪棋盘标定板图像。

图 2 红外热像仪棋盘标定板图像

Fig. 2 Infrared imager board calibration plate image

2.2 模型制定

首先假设所有的模型都是一个把噪声 Y 作为输入并计算未知估计量 X 的概率模型,我们要寻找使得在 X 发生这个条件下 Y 最小这一事件达到最大概率的 X,可以记作 $X^* := \operatorname{argmax}_X P(Y \mid X)$。直接法是将摄像机采集的数据直接作为噪声输入,而间接法是将图像数据做处理之后提取有效信息作为模型的输入。稀疏法和稠密法的区别在于在稀疏法建立地图时所用到的点数更少,并且不会用到先验的几何知识。在直接法和稀疏法视觉里程计模型中,我们将光度误差看做噪声 Y,光度误差 E_{pj} 定义如下:

$$E_{pj} := \sum_{p \in N_p} w_p \left\| (I_j[p'] - b_j) - \frac{t_j e^{a_j}}{t_i e^{a_i}} (I_i[p] - b_i) \right\|_\gamma \tag{1}$$

其中,N_P 是以像素点 p 为中心的像素集合,参见图 3;t_i,t_j 分别是图像 I_i,I_j 的曝光时间;a_i, b_i, a_j, b_j 是光照变化函数的参数;$\| \cdot \|_\gamma$ 表示 Huber 范数;p' 表示 p 点的投影点位置,可以通过下式计算:

$$p' = \Pi_c (R \Pi_c^{-1}(p, d_p) + t) \tag{2}$$

其中:

$$\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} := T_j T_i^{-1} \tag{3}$$

其中,Π_c 表示投影矩阵;Π_c^{-1} 表示重投影矩阵;d_p 表示 p 点的逆深度[11];R 表示旋转矩阵;t 表示平移向量;T_i,T_j 分别表示第 i 帧和第 j 帧热像仪的位姿。另外权重系数 w_p 如下:

$$w_p := \frac{c^2}{c^2 + \left\| \nabla I_i[p] \right\|_2^2} \tag{4}$$

其中，c 为常量；$\nabla I_i|p|$ 表示 p 点为中心的图像梯度。

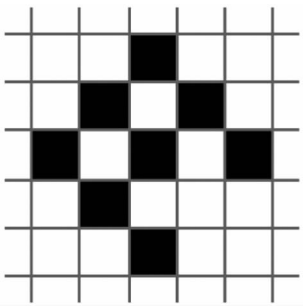

图3 模式 N_P

Fig. 3 Pattern N_P

总结上面的内容，误差 E_{pj} 取决于以下变量：(1)该点的逆深度值 d_p；(2)热像仪内参；(3)所涉及的热像仪位姿 T_i，T_j；(4)光照变化函数参数 a_i，b_i，a_j，b_j。

对于两帧的一个像素点的光度误差我们用上面的方法定义，待优化的所有帧和所有点的总光度误差给出如下：

$$E_{photo} := \sum_{i\in F}\sum_{p\in P_i j\in obs(p)}\sum E_{pj} \tag{5}$$

式中，i 遍历所有集合中的图像帧；p 遍历图像帧 i 的所有地图点；j 遍历所有帧中能够观测到点 p 的对应点。这样求解一个极大似然估计的问题转化成了一个无约束的非线性优化问题，通过最小化 E_{photo} 求解出它所依赖的所有变量。

2.3 帧管理

总是保留最多 $N_f = 7$ 个关键帧，新来的每帧图像，只和最新的关键帧比较，追踪热像仪位姿，之后新来的这帧图像被用于产生一个新的关键帧或者被丢弃，如果被用于产生关键帧，在所有的关键帧进行总光度误差优化之后，该关键帧在满足一定条件将被边缘化。下面是帧管理的步骤：

步骤一：新帧跟踪。当新的关键帧被创建时，所有激活的地图点投影到该帧上，从而创建半稠密地图。当新帧产生时，仅相对最近关键帧进行图像直接配准，从而完成跟追踪，其中会使用到多尺度图像金字塔和恒速运动模型。

步骤二：创建关键帧。以下三种情况会创建新的关键帧：(1)在视角变化时会创建新的关键帧；(2)在相机平移导致遮挡或遮挡去除时会创建新的关键帧；(3)在曝光时间显著变化时会创建新的关键帧。

步骤三：边缘化关键帧。假设 $I_1,\cdots,I_n$ 是已被激活的关键帧集合，其中 I_1 是最新的关键帧；I_n 是最旧的关键帧。边缘化策略是：(1)总是保留最新的两个关键帧 I_1 和 I_2；(2)该关键帧的少于5%的点在关键帧 I_1 中被观测到，则边缘化该关键帧；(3)如果激活的关键帧多于 N_f，则边缘化“距离分数” $s(I_i)$ 最大的关键帧(不包括 I_1 和 I_2)，公式为：

$$s(I_i) = \sqrt{d(i,1)}\sum_{j\in[3,n]\setminus\{i\}}(d(i,j)+c)^{-1} \tag{6}$$

其中，i，j 为关键帧的序号；n 为关键帧的总数；$d(i,1)$ 是关键帧 I_i 和 I_1 间的欧几里得距离；$d(i,j)$ 是关键帧 I_i 和 I_j 间的欧几里得距离；c 为常数。采用距离分数是为了保证激活的关键帧在3D空间中的良好分布，使更多的关键帧靠近最近的关键帧。

2.4 点管理

点目标是在优化时始终保持固定数量 N_p 的激活点(使用 $N_p = 2000$)，在3D空间和激活帧中均匀分布。首先，识别每个新关键帧中的 N_p 个候选点。候选点不立即添加到优化框架中，而是在后续帧中单独追踪，生成粗略的深度值估计以用于初始化。当需要将新点添加到优化框架中时，激活一些候选点(来自优化窗口中的所有帧)，添加到优化框架中。注意，在每个帧中选择 N_p 个候选点，但只保留所有激活帧中的 N_p 个激活点。具体点管理的步骤分为以下三步：

步骤一：候选点选择。所选取点的策略是：(1)在图像中良好分布；(2)相对它们周围的环境具有较高的图像梯度。通过将图像分成 32×32 块可以获取一个区域自适应梯度阈值。对于每一个图像块，规定它的阈值为 $\bar{g}+g_{th}$，其中，$\bar{g}$ 是该块中所有像素的强度梯度的中值；g_{th} 为全局常数(使用 $g_{th}=7$)。为了在整幅图像获得一个均匀的分布，将图像分成 32×32 块，针对每一块区域如果有大于区域自适应阈值的点，选择区域具有最大像素梯度的点，如果该区域没有大于区域自适应阈值的点，舍弃该区域。

步骤二：候选点追踪。在后续帧中沿着候选点的极线进行离散化搜索，追踪候选点，从而最小化光度误差。从最佳匹配结果中，计算深度值及其方差，其用于约束后续帧的搜索空间。这种跟踪策略的灵感来自 LSD－SLAM 算法。一旦点被激活，计算的深度值仅用于初始化。

步骤三：候选点激活。在一组旧点被边缘化之后，需要激活候选点以替换它们。同样，目标是保持地图点在整个图像中均匀分布。为此，首先将所有

激活点投影到最近的关键帧上。然后激活与现有的地图点距离最大的候选点(在第二和第三次候选点选择时创建的候选点需要更大的距离阈值)。

2.5 滑动窗口优化

直接法和稀疏法视觉里程计模型公式给出了总光度误差公式(5)。帧管理给出了待优化的所有关键帧。点管理给出了关键帧中的激活点。利用滑动窗口并借助高斯 - 牛顿法最小化总的光度误差可以迭代求解出总光度误差所依赖的所有变量。

将所计算的高斯 - 牛顿系统做如下定义:

$$H = J^T W J \qquad b = -J^T W r \tag{7}$$

其中, W 是包含权重的对角矩阵; r 是残差向量; J 是 r 的雅克比。

如图 3 所示,每一个点为能量值贡献 $|N_P| = 8$ 个残差,为了简单表示,接下来只考虑一个残差 r_k 和它所对应的雅克比 J_k。在优化或者边缘化过程中残差总是从当前状态开始计算的:

$$r_k = r_k(x(\zeta_0))$$

$$= (I_j[p'(T_i, T_j, d, c)] - b_j) - \frac{t_j e^{a_j}}{t_i e^{a_i}}(I_j[p] - b_i) \tag{8}$$

$$x(\zeta_0) := e^{\hat{x}} \cdot \zeta_0 \tag{9}$$

其中, $(T_i, T_j, d, c, a_i, a_j, b_i, b_j) := x(\zeta_0)$ 残差所依赖的当前状态变量; ζ_0 表示初始的状态; x 为更新值。计算雅克比 J_k 的公式为:

$$J_k = \frac{\partial r_k((\delta + x)(\zeta_0))}{\partial \delta} \tag{10}$$

它可以分解为:

$$J_k = [\underbrace{\frac{\partial I_j}{\partial p'}}_{J_I} \underbrace{\frac{\partial p'((\delta + x)(\zeta_0))}{\partial \delta_{geo}}}_{J_{geo}}, \underbrace{\frac{\partial r_k((\delta + x)(\zeta_0))}{\partial \delta_{photo}}}_{J_{photo}}] \tag{11}$$

其中, I_j 为第 j 个关键帧; p' 为关键帧的一个像素点; δ_{geo} 表示几何参数; δ_{photo} 表示光度参数; J_I 表示图像对像素的导数; J_{geo} 表示像素对几何参数的导数; J_{photo} 表示残差对光度参数的导数。

在计算 J_{geo} 和 J_{photo} 时,借助雅克比第一次估计[12-13],之后迭代求解出相机位姿和点的逆深度,完成定位和建图的任务。

3 实 验

3.1 红外热像仪标定以及红外视频三维重建

对红外热像仪标定首先将热像仪固定,平移和旋转标定板,录制一个红外视频,然后对红外视频进行抽帧选取 20 张具有不同位姿且对比度较高的标定板图像,将其输入到 matlab 标定工具箱进行标定,需要手工标出边缘的四个角点,程序会自动提取其余角点并计算出热像仪内参。标定结果为:像素焦距 f_c = [1636.10735 1516.22703]、主点坐标 cc = [206.06274 344.97944]、畸变系数 k_c = [-0.21301 5.85797 0.01173 -0.03941 0.00000]。

实验平台系统为 Ubuntu14.04,cpu 为 Inter(R) Core(TM) i3 - 4160,内存大小为 8GB。我们用 FLIR PathFindIR 系列第一代产品采集一段东华大学图书馆的视频,然后将视频抽帧得到连续的图像序列,并将图像序列重命名为规定的格式,然后将标定获得的红外热像仪内参写入 camera.txt 文件。因为红外热像仪成像不受光照变化的影响,所以在程序运行时我们将程序运行模式设置为无光度文件运行,程序运行完毕之后三维重建的实验结果如图 4 所示。第一行从左到右依次为红外视频的第 1 帧、第 101 帧、第 201 帧、第 301 帧,第二行为两个视角三维重建的点云。

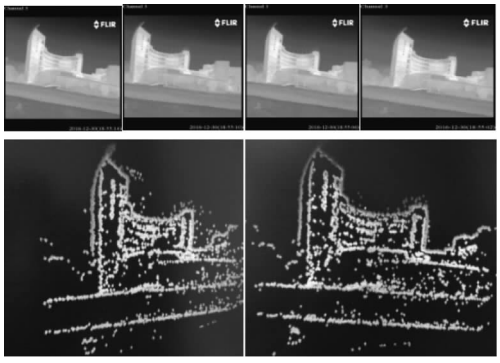

图 4 不同视角的点云图

Fig. 4 Point cloud of different perspectives

3.2 结果分析

红外热像仪标定的重投影误差为 err = [1.11672 1.68977]。由最后的误差可以看出标定的误差在 1 到 2 个像素级,标定结果是比较准确的。在点云重建过程中,算法对视频的处理的可以达到实时,处理速度达到每秒 52 张关键帧,远远超过了我们录制的红外视频帧率每秒 24 帧。另外为了避免偶然因素对实验结果的影响,我们将逆序的图像序列作为输入图片集,最后计算出的三维点云和图 4 中的点云结果基本一致,验证了该算法的灵活性。由于直接法没有检测匹配图像中的特征,所以位姿跟踪的鲁棒性比较差,对于帧间运动过大的场景容易出现跟丢的情况。图 4 不同视角的点云图可以很好地反映出图书馆的结构,但是表征道路的点云只是两条稀疏的线,这是由于稀疏法只是选取了图像

强度梯度较高的点作为三维重建的对象。另外由于同一个像素点可以被观测多次,三维空间中的一个点可能对应地图中的多个具有不同逆深度的点,所以从一些角度看点云可能是发散的。

4 结 论

本文采用直接法和稀疏法视觉里程计恢复单目红外视频中的场景结构。该算法将不进行图像的特征点检测与匹配,而是直接利用灰度值构建待优化的模型,通过视觉里程计前端完成对关键帧和激活点的筛选,最后利用高斯 - 牛顿法滑动窗口优化总光度误差求解出相机的位姿和空间点的逆深度,完成实时三维重建的任务。虽然现在的点云结果没有达到稠密化,但综合来看该算法的实时性较好,并且能够很好地重建单目红外视频场景的结构,将具有较好的应用前景。

参考文献:

[1] WU Tong, FU Zhongli. 3Dreconstruction technology and its military application[J]. National Defense Science & Technology, 2015, 36(1): 31 - 34. (in Chinese)
吴彤,傅中力. 三维重建技术及其军事应用[J]. 国防科技,2015,36(1):31 - 34.

[2] Hartley R, Zisserman A. Multiple view geometry in computervision [M]. Cambridge: Cambridge University Press, 2004.

[3] Durrant-Whyte H, Bailey T. Simultaneous localization andmapping: Part I [J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99 - 110.

[4] Bailey T, Durrant-Whyte H. Simultaneous localization andmapping(SLAM): Part II[J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108 - 117.

[5] Klein G, Murray D. Parallel tracking and mapping for small ARworkspaces[C]. Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007: 225 - 234.

[6] Engel J, Sturm J, Cremers D. Semi-dense visual odometry for amonocular camera[C]. Proceedings of IEEE International Conference on Computer Vision, 2013: 1449 - 1456.

[7] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale directmonocular SLAM[C]. ECCV, 2014: 834 - 849.

[8] SUN Xinde, BO Shukui, LI Lingling. Study of infrared image clutter suppression based on background estimation [J]. Laser & Infrared, 2011, 41(5): 586 - 590. (in Chinese)
孙新德,薄树奎,李玲玲. 基于背景估计的红外图像杂波抑制方法研究[J]. 激光与红外, 2011, 41(5): 586 - 590.

[9] Engel J, Koltun V, Cremers D. Direct sparse odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, (99): 1 - 1.

[10] Z Zhang. AFlexible new techniquefor camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330 - 1334.

[11] J Civera, A Davison, J Montiel. Inverse depthparametrization for monocular SLAM[J]. Transactions on Robotics, 2008, 24(5): 932 - 945.

[12] G P Huang, A I Mourikis, S I Roumeliotis. Afirst-estimates Jacobian EKF for improving SLAM consistency [C]. Experimental Robotics, 2009: 373 - 382.

[13] S Leutenegger, S Lynen, M Bosse, et al. Keyframe-based visual-inertialodometry using nonlinearoptimization [J]. International Journal of Robotics Research, 2015, 34(3): 314 - 334.