

单目视觉的同时三维场景构建和定位算法

沈晔湖 刘济林 杜 歆

(浙江大学信息与电子工程学系, 浙江 杭州 310027)

摘要 同时场景构建和定位算法是机器人自主导航的重要组成部分。针对传统算法不能应用于室外环境和缺乏定量分析的缺点,提出了一种单摄像机恢复场景三维结构和摄像机位姿的新算法。提出了视频序列关键帧提取方法,降低了运算复杂度;利用特征点对和摄像机内参量计算场景三维结构和关键帧的位姿并提出一种估计关键帧位姿的简便方法;最后,提出一种兼顾优化效果和运算复杂度的自适应光束法平差算法优化场景结构和摄像机位姿,并生成适于机器人导航的数字高程图。室内和室外多种场景下的定量和定性实验结果表明,绕行误差低于 4%,该算法能够接近实时准确实现同时场景构建和摄像机定位。

关键词 光学测量;同时场景构建和定位;自适应光束法平差;单目视觉

中图分类号 TP391 **文献标识码**: A

Simultaneous Three-Dimensional Environment Reconstruction and Localization based on Monocular Vision

Shen Yehu Liu Jilin Du Xin

(Department of Information Science and Electronic Engineering, Zhejiang University,
Hangzhou, Zhejiang 310027, China)

Abstract Simultaneous localization and mapping (SLAM) is one of the most important components in robot navigation. A novel SLAM algorithm based on monocular vision is proposed to overcome the difficulties in outdoor applications and quantitative analysis with traditional methods. Firstly, a key frame selection method is proposed to reduce the computational cost. Then the three-dimensional (3-D) structure of the environment and the positions of the camcorder are estimated based on matched feature points and the intrinsic parameters of the camcorder. A simple method with reasonable optimizing effect and computing cost is applied to get position and orientation of the camcorder. Finally, an adaptive bundle adjustment is adopted to optimize the 3D structure of the environment and the positions of the camcorder simultaneously. Digital elevation map (DEM) which is more suitable for robot navigation is also obtained. Quantitative and qualitative experimental results show that the loop closure error is less than 4%. The algorithm can reconstruct the environment and localize the camcorder accurately in nearly real time.

Key words optical measurement; simultaneous localization and mapping; adaptive bundle adjustment; monocular vision

1 引 言

机器人自主导航是近几十年来机器人学界研究的一个重点问题。为了使自主机器人能够规划运行路径、躲避障碍,必须构建周围环境的地图并确定机器人在地图中的位置和姿态。因此,同时场景构建和定位算法成为机器人自主导航主要组成部分。该

算法最先由 J. J. Leonard 等^[1]提出,按距离传感器的不同大致可将现有算法分为两类。第一种采用主动距离传感器,如激光雷达^[2,3]或声纳^[4],这些算法一般采用扩展卡尔曼滤波^[5]或粒子滤波^[2]获得一致的地图构建和定位结果。激光雷达获得的数据精度较高,但具有重量大、电能消耗多和价格较高的缺

收稿日期: 2007-10-10; **收到修改稿日期**: 2007-11-07

基金项目: 国家自然科学基金重点项目(60534070)、国家自然科学基金(60502006)和浙江省科技厅项目(2005C14008)资助课题。

作者简介: 沈晔湖(1982-),男,江苏人,博士研究生,主要从事计算机视觉方面的研究。E-mail: paulsyh@zju.edu.cn

导师简介: 刘济林(1947-),男,湖北人,教授,博士生导师,主要从事计算机视觉、模式识别、智能交通等方面的研究。

E-mail: liujl@zju.edu.cn

点,在地外行星自主车载荷和能耗皆有限的情况下不适合。而声纳角分辨率低,且不能够在月球等无空气环境下工作,因此主要用于室内环境。另外一种是基于视觉传感器的同时场景构建和定位(SLAM)方法,它克服了激光雷达重量大和电能消耗多的缺陷,并且相机是全固态的,适合在地外行星自主车中使用。这些算法利用立体视觉系统^[6]。较之需要多台摄像机的立体视觉系统,单摄像机无需摄像机同步设备和保持摄像机相对位置的机械装置,因此使用更方便、价格更低廉、消耗电能更少,适合作为备份系统。A. J. Davison^[7]首次提出了基于单摄像机的同时场景构建和定位算法,但其仅跟踪 100 个左右的特征点,无法满足室外应用的需要。D. Nistér^[8]提出了利用单目视觉和立体视觉的视觉里程计算法,但仅提供了立体视觉的实验结果。J. Campbell 等^[9]利用一台消费级摄像机实现了视觉里程计,但算法着重于机器人定位。M.

Tomono^[10]提出了自动选择基线长度的方法,但缺乏定量分析。

针对以上各种算法的缺陷,本文利用图像中的特征点提高匹配的准确度,结合特征跟踪结果,提出了一种关键帧提取方法,所有的场景结构重建和摄像机位姿估计均在关键帧处进行,从而在提高精度的同时降低了算法复杂度。与许多运动求取结构(Structure from motion, SFM)方法^[11]不同,本文方法工作在串行模式下,这意味着该算法能够在拍摄视频序列的同时进行定位和地图重建。此外本文算法无需外加平面镜组^[12],因此简单可靠,适合自主车导航。

2 同时场景构建和定位算法

图 1 为整个算法的流程框图。该算法主要分为如下 6 个步骤。

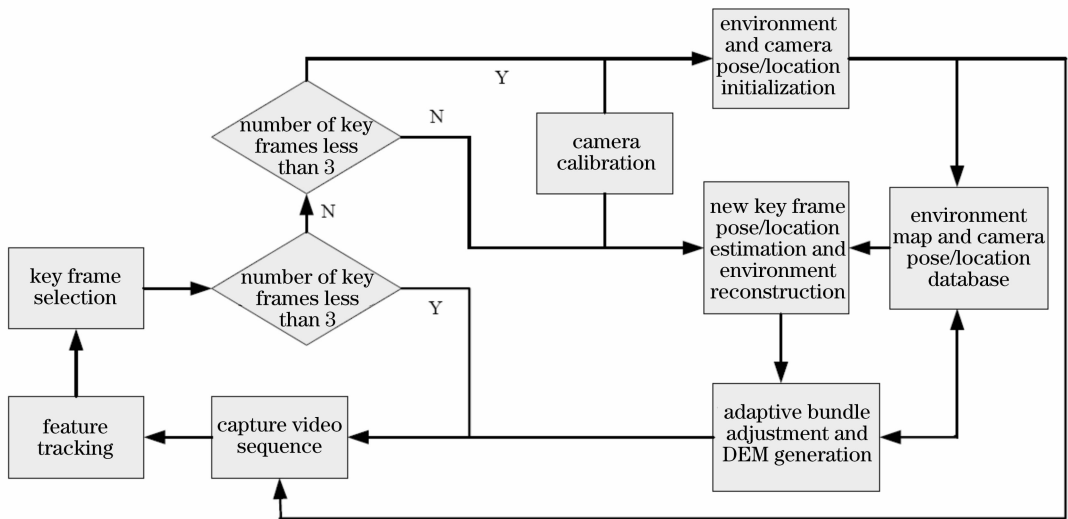


图 1 算法流程框图

Fig. 1 Algorithm flow chart

2.1 摄像机内参量标定

理想情况下,摄像机符合线性模型,摄像机内参量由矩阵 \mathbf{K} 表示:

$$\mathbf{K} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

式中 f_x 和 f_y 分别为以水平和垂直像素单位表示的焦距长度, (u_0, v_0) 为光轴与图像平面的交点, γ 为倾斜因子。一般摄像机均存在一定的成像畸变,可表示为:

$$\mathbf{x}_d = (1 + kc_1 r^2 + kc_2 r^4) \mathbf{x}_n, \quad (2)$$

式中 \mathbf{x}_d 为实际成像的归一化坐标, $\mathbf{x}_n = [x \ y]^T$ 为线性摄像机模型下的归一化坐标, $r^2 = x^2 + y^2$, kc_1 ,

kc_2 为非线性畸变系数。本文利用文献^[13]的方法计算 \mathbf{K} 和非线性畸变系数 kc_1 和 kc_2 。

2.2 特征点跟踪

为降低图像间的误匹配并且加快匹配速度,首先在图像中选取易于匹配的特征点,然后在下一帧图像中该特征点位置附近的区域内搜索与之最相似的位置,实现 Kanade-Lucas-Tomasi 特征跟踪算法^[14]。借助视频序列特征点的跟踪,可以高效实现大视角变化下图像对特征点的匹配。

2.3 关键帧选取

若对视频序列每帧均估计其位姿,则运算复杂度过高。由于前后两帧运动很小,估计场景结构是

一个病态问题,并且估计每帧的位姿对运行速度较慢的地外行星自主车意义不大。因此本文从视频序列中抽取关键帧,场景重建、摄像机位姿估计均在关键帧进行。算法如下:

1) 视频序列中的第一帧必为关键帧 I_1 。

2) 预设阈值 m_1 ,如果当前帧中与 I_1 公共特征点个数首次低于 m_1 时,该帧为第二关键帧 I_2 。

3) 假设当前找到 N ($N \geq 2$) 个关键帧。令 $P_1(k)$ 为第 k 帧和 I_N 中公共特征点的个数, $P_2(k)$ 为第 k 帧和 I_{N-1} 中公共特征点的个数,如果满足条件

$$\begin{aligned} & [P_1(k) < m_1 \wedge P_1(k-1) \geq m_1] \vee \\ & [P_2(k) < m_2 \wedge P_2(k-1) \geq m_2], \quad (3) \end{aligned}$$

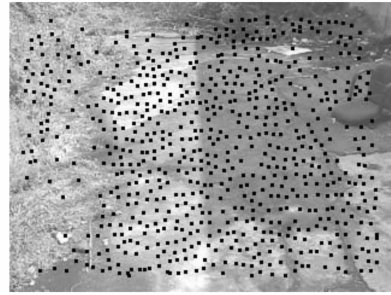
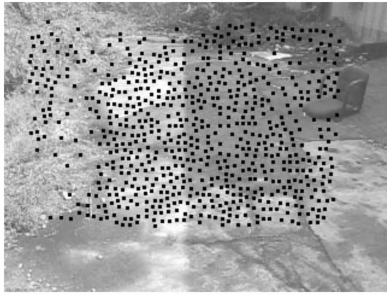


图 2 前后两幅关键帧的匹配结果

Fig. 2 Matching results between two consecutive key frames

2.4 初始场景和摄像机位姿估计

得到 3 幅关键帧后进行初始化。令第 1 幅关键帧对应的摄像机坐标系为世界坐标系, G_i ($i=1 \dots 3$) 为特征点在 3 幅关键帧中的齐次坐标集合。对所有 $x_c \in G_i$ ($i=1 \dots 3$), 计算

$$x_n = d^{-1}(K^{-1}x_c), \quad (4)$$

式中 $d^{-1}(\cdot)$ 为(2)式的反函数。对于第 1 和第 3 幅关键帧中的对应点 x_n 和 x'_n , 有极线约束关系^[15]:

$$x_n^T E x'_n = 0, \quad (5)$$

式中 E 为本质矩阵。本文利用 8 点法^[16] 和随机一致采样算法(RANSAC)^[17] 估计 E 并进行奇异值分解:

$$E \propto U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T, \quad (6)$$

式中 U 和 V 均为矩阵, 它们的行列式均大于 0。令

$$D = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

可得 2 个不同的旋转矩阵 $R = UDV^T$, $R' = UD^T V^T$ 和平移矢量 $t = [u_{13} \ u_{23} \ u_{33}]^T$, 其中 u_{ij} 为矩阵 U 的第 i 行第 j 列元素。则第 3 幅关键帧位姿参量有 3 组可能的解, 分别为 R 和 t , R 和 $-t$, R' 和 t , R' 和 $-t$ 。取

则 $I_{N+1} = k$, 其中 m_2 为一预设阈值。

以上规则可在跟踪特征点的同时得到关键帧位置, 无需额外运算。若 m_1 和 m_2 过大则导致两帧间距较小, 无法克服估计场景结构中的病态问题; 若 m_1 和 m_2 过小, 则帧间公共点过少, 不但影响相机位姿求解精度, 而且导致场景地图过于稀疏, 实用性较差。本文在每个关键帧中选取 1500 个特征点进行跟踪, 因此可取 $m_1 = 700$, $m_2 = 350$ 。当新关键帧产生后, 补充新特征点使其在关键帧中保持 1500 个。图 2 显示了关键帧间的匹配结果, 黑点表示两帧匹配的特征点(未去除误匹配点)。

任意一对特征点, 结合位姿参量用三角化方法进行三维重建^[15]。若该点在两摄像机前方, 则即为所求。由于单目视觉只能在相似的意义下重建, 需要对 t 进行归一化使得 $\|t\| = 1$ 。

得到旋转矩阵和平移矢量后, 摄像机对应的投影矩阵可以表示为 $P = K[R \ t]$, 从而对第 1 和第 3 幅关键帧的对应特征点用三角化方法进行重建。获得特征点的三维坐标后, 可以求得第 2 幅关键帧的位姿。最后, 利用传统的光束法平差^[18] 对 3 帧的位姿参量和三维特征点坐标进行同时优化。

2.5 新关键帧位姿估计及场景三维重建

获得新关键帧后, 需估计其位姿。针对文献[10]的缺陷, 本文提出了一种无需非线性优化的位姿估计方法。

设新关键帧为 I_{N+1} , F_i 为第 i 关键帧中特征点齐次坐标的集合, S 为场景三维坐标集合, $L_i(\mathbf{X})$ ($\mathbf{X} \in S$) 为将三维点 \mathbf{X} 转化到 I_i 中的投影函数。其理论值可通过 I_i 对应的投影矩阵 $P^i = K[R^i \ t^i]$ 计算得到, 实际值可以通过特征跟踪获得:

1) 计算公共场景三维坐标集合

$$S' = \{\mathbf{X} | \mathbf{X} \in S \wedge L_{N-1}(\mathbf{X}) \in F_{N-1} \wedge L_{N+1}(\mathbf{X}) \in F_{N+1}\},$$

令 I_{N-1} 和 I_{N+1} 中的公共二维投影点集合分别为

$$G_{N-1} = \{x | x = L_{N-1}(X) \wedge X \in S'\},$$

$$G_{N+1} = \{x | x = L_{N+1}(X) \wedge X \in S'\},$$

对所有的 $x_c \in G_{N-1}$ 及其对应点 $x'_c \in G_{N+1}$, 计算 R 和 t , 且 $\|t\| = 1$. 则 $R_{N+1} = RR_{N-1}$.

2) 仅在相似的意义下得到了 I_{N-1} 到 I_{N+1} 的平移矢量, 因此不妨设实际平移矢量为 λt , 未知量 λ_i 为平移矢量的大小. 对于 $X \in S'$, 计算 $X_T = R_{N-1}X + t_{N-1}$, 则

$$RX_T + \lambda t \sim x = [u \ v \ 1]^T = L_{N+1}(X) \in G_{N+1}, \quad (7)$$

令 $RX_T = [m_x \ m_y \ m_z]^T$, $t = [t_x \ t_y \ t_z]^T$, 整理(7)式可得

$$\begin{cases} (ut_z - t_x)\lambda_i = m_x - um_z, \\ (vt_z - t_y)\lambda_i = m_y - vm_z, \end{cases} \quad (8)$$

通过一个三维点就可求出 λ_i . 为了增加可靠性, 对每一个 $X \in S'$ 均计算 λ_i , 最后结果取其均值. I_{N+1} 所对应的平移矢量为 $t_{N+1} = Rt_{N-1} + \lambda t$.

由于每一关键帧会补充一些新的特征点, I_{N-1} 中新加入的部分特征点未进行三维结构重建. 因此, 得到 I_{N+1} 的位姿参量后, 对 I_{N-1} 中新加入的且在 I_{N+1} 仍可见的特征点进行三维结构重建.

2.6 自适应光束法平差及数字高程图生成

2.6.1 稳健的自适应光束法平差算法

在新加入关键帧并对其进行位姿估计和场景重建后采用光束法平差优化估计结果. 代价函数为

$$f(C, X) = \frac{1}{MN_{T_i=N-N_T+1}} \sum_{j=1}^M r[x_{ij} - L_i(X_j)]^2, \quad (9)$$

$(X_j \in X)$

式中 $M = |X|$, $C = \{R^i, t^i | N - N_0 < i \leq N\}$ 为待优化的关键帧位姿参量集合, N_0 为待优化位姿参量的关键帧数, 共考虑 N_T 幅关键帧中的特征点投影, X 为待优化的场景结构坐标集合, x_{ij} 为第 j 个三维点在第 i 关键帧中成像的坐标. $r(x)$ 为 L2 范数. 传统光束法平差^[18] 在三维结构和摄像机位姿参量较多的情况下算法复杂度仍较高, 它每次迭代需求解

$$\begin{bmatrix} U^* & W \\ W^T & V^* \end{bmatrix} \begin{bmatrix} \delta_a \\ \delta_b \end{bmatrix} = \begin{bmatrix} \varepsilon_a \\ \varepsilon_b \end{bmatrix}, \quad (10)$$

式中左侧为 Hessian 矩阵 $H = J^T J$ 对角线分量乘以 $(1 + \lambda)$ 后的矩阵, δ_a 和 δ_b 分别为每次迭代过程中摄像机位姿参量和场景结构坐标的增量, ε_a 和 ε_b 均为残差. 首先

$$(U^* - WV^{*-1}W^T)\delta_a = \varepsilon_a - WV^{*-1}\varepsilon_b, \quad (11)$$

求得 δ_a , 然后利用得到的 δ_a 求解 δ_b :

$$\delta_b = V^{*-1}(\varepsilon_b - W^T\delta_a),$$

该方法计算量较大的部分为求 Hessian 矩阵和解(11)式. 为简化分析, 设平均每个关键帧中特征点个数为 n . 求 Hessian 矩阵的运算复杂度为 $O(nN_T)$. WV^{*-1} 内非零块的个数与 W 相同, 所以运算复杂度为 $O(nN_0)$, 从而计算 $WV^{*-1}W^T$ 的复杂度为 $O(nN_0^2)$. 解(11)式的复杂度为 $O(N_0^3)$. 因此, 总的运算复杂度约为 $O(nN_T) + O(nN_0^2) + O(N_0^3)$. 传统光束法平差 $N_0 = N_T = N$, 因此随着关键帧个数 N 的增多, 算法复杂度快速增长. 本文提出如下算法:

1) 当 $N \leq 20$ 时, $N_0 = N_T = N$ 即进行传统光束法平差. 当 $N = 21$ 时, $N_0 = 3, N_T = N_0 + 5$; 当 $N > 21$ 时, 根据前 2 个关键帧的优化残差确定 N_0 和 N_T :

$$N_0 = \begin{cases} N_0 - 2, & (e_{N-1} < e_{N-2} \wedge N_0 > 3) \\ N_0 + 2, & (e_{N-1} > e_{N-2} \wedge N_0 < 9), \\ N_0, & \text{otherwise} \end{cases}$$

$$N_T = N_0 + 5, \quad (12)$$

2) 利用光束法平差进行优化:

$$\{C_{\text{opt}}, X_{\text{opt}}\} = \underset{C, X}{\operatorname{argmin}} f(C, X), \quad (13)$$

计算内点集合:

$$X_i = \{X_j | X_j \in X_{\text{opt}} \wedge r[x_{ij} - L_i(X_j)] < \delta \wedge L_i(X_j) \in F_i \wedge N - N_T < i \leq N\}, \quad (14)$$

在所有实验中都取阈值 $\delta = 2$. 如果 $X_i \neq X_{\text{opt}}$, 则再次利用光束法平差进行优化:

$$\{C_{\text{opt}}^n, X_{\text{opt}}^n\} = \underset{C_{\text{opt}}, X_i}{\operatorname{argmin}} f(C_{\text{opt}}, X_i), \quad (15)$$

否则, 令 $C_{\text{opt}}^n = C_{\text{opt}}, X_{\text{opt}}^n = X_{\text{opt}}$.

3) 计算关键帧 I_N 的优化残差

$$e_N = f(C_{\text{opt}}^n, X_{\text{opt}}^n), \quad (16)$$

当 $N \leq 20$ 时, 传统光束法平差复杂度不高, 且能获得较准确的初始估计. 当关键帧个数增长后需限制 N_0 和 N_T , 但是当前一次优化残差增大时, 需要加入更多优化关键帧; 反之, 可降低优化关键帧的数量. 把 N_0 限制在 3~9 之间是考虑到性能和速度的折中. 为了保证优化前后结果的一致性, N_T 必须满足: $N_T \geq N_0 + 2$, 所以取 $N_T = N_0 + 5$ 也是性能和速度的折中. 对于待优化的三维场景点, 取在 N_0 个关键帧中至少有一幅可见的场景点.

为了消除特征匹配过程中误匹配点的影响, 本文提出了一种两步光束法平差法. 由于仅部分关键帧需要进行两次光束法平差, 且第二次光束法平差初始值较好, 所需迭代次数少, 因此对算法复杂度的

增加不明显。

2.6.2 数字高程图生成

为将环境三维点数据转换为更适合自主车导航的数字高程图,本文采用均匀采样网格。对于三维点 $\mathbf{X} = (x, y, z)^T \in S$, 将其转换到世界坐标系下。该坐标系下 XY 平面与地平面平行, Y 轴水平向前, Z 轴垂直向上, 结合该点的 X 和 Y 坐标将其分配到对应的数字高程图格点中。对于高程图每个格点, 选取所有三维点中 Z 坐标最大值作为该格点的高度值。

3 实验结果

3.1 算法有效性验证

为了验证算法主要部分的有效性, 同时展示算法在室内场景下的结果, 选取了 Dinosaur 序列。该序列共有 36 帧图像, 分辨率为 720×576 , 摄像机固

定, 将模型置于前方转盘中心, 转盘每旋转 10° 拍摄 1 帧, 恰好环绕模型一周。图 3(a) 和图 3(b) 分别为序列的第 1 帧和 18 帧。由于已提供各帧二维特征点的坐标数据, 直接利用本文算法, 每帧均为关键帧。图 3(c) 和图 3(d) 分别为重建结果的顶视图和侧视图, 其中白色部分为摄像机在各帧下的位姿。由图可见摄像机基本等间隔位于同一个圆上, 模型重建结果也较可信。

为进行定量分析, 用最小二乘法对恢复的摄像机位置坐标拟合一个圆, 则真实的摄像机位置坐标均匀分布在该圆上, 分别对摄像机位置和姿态进行了误差分析, 如图 4 所示。由于位置绝对误差没有意义, 分析了相对误差(即将位置绝对误差除以摄像机绕行一圈的总长度)。

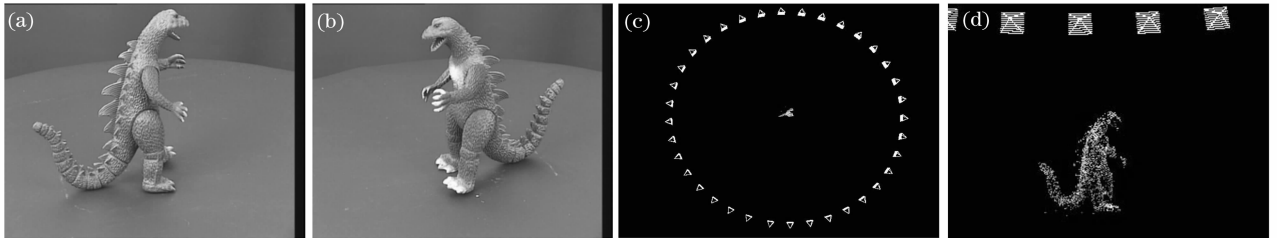


图 3 Dinosaur 序列。第 1 帧图像(a), 第 18 帧图像(b), 重建结果顶视图(c), 侧视图(d)

Fig. 3 Dinosaur sequence—first frame (a), 18th frame (b) and top view (c), side view (d) of reconstruction results

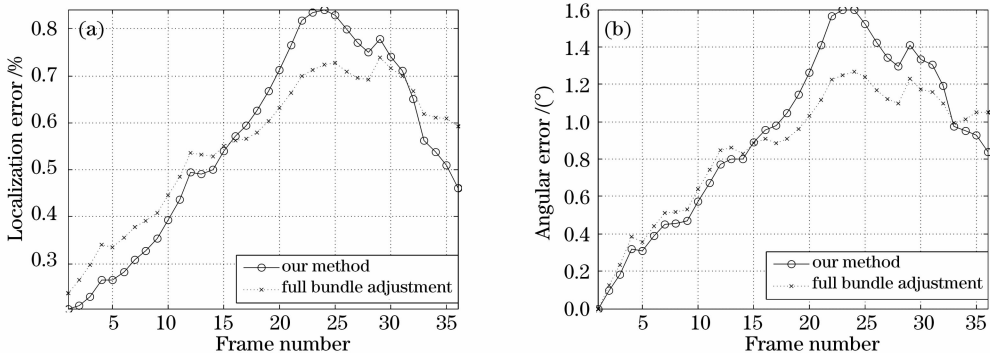


图 4 摄像机位置相对误差(a)和摄像机姿态绝对误差(b)

Fig. 4 (a) Relative error of the camera positions and (b) absolute error of the camera poses

由图 4 可见, 本文方法误差较小。在第 15 帧之前, 两种方法误差接近, 在第 15 帧后, 本文方法误差略高, 但最后几帧本文方法误差更小。本文方法位置相对误差低于 1%, 姿态误差低于 1.6° 。本文算法耗时 30.3 s, 利用传统光束法平差方法耗时 76.4 s, 为本文算法的 2.5 倍以上。可见本文算法在性能和速度上找到了一个平衡点。

3.2 室外场景实验结果

将 Canon A630 安装到 Pioneer 3 自主机器人

顶部, 拍摄了 2 段视频序列, 摄像机视角约 53° , 视频分辨率为 640×480 , 帧率为 15 frame/s。所有结果利用本文完整算法全自动获得。

图 5(a) 显示了第 1 段视频序列的场景概貌, 沿图中黑色箭头方向绕行一周回到初始点。相机指向和运动方向接近垂直。图 5(b) 和图 5(c) 为视频序列中的 2 帧示意图, 图 5(d)~图 5(f) 为 3 个不同视角下的重建结果。图 5(g) 为 Google Earth 的截图, 图中白色轨迹为运动路径。利用 Google Earth 测

量得该路径长度约为 108.45 m。图 5(h)为场景的数字高程图。格点大小为 $0.5 \text{ m} \times 0.5 \text{ m}$ 。该序列

距离较长,算法的成功表明了其稳健性。

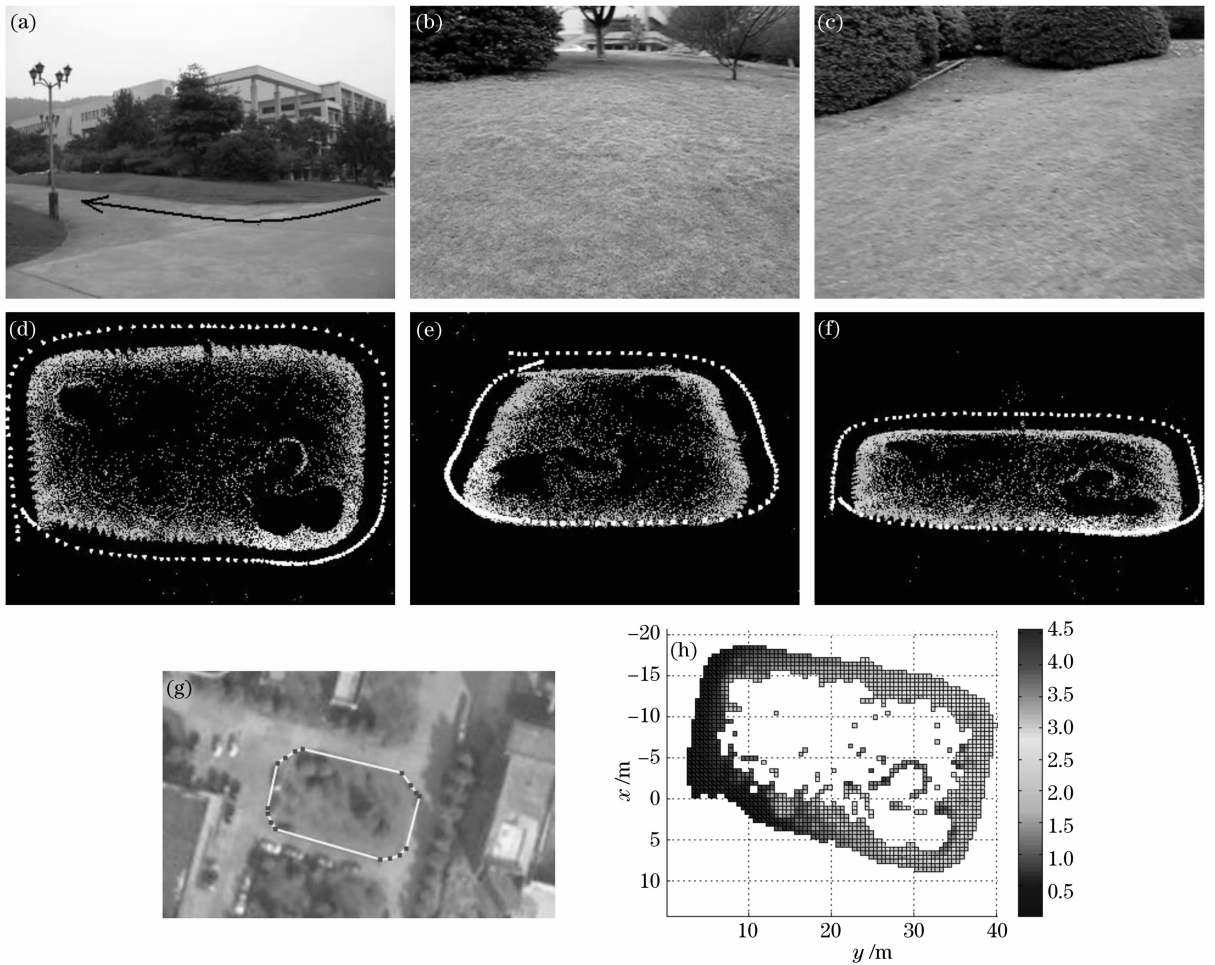


图 5 第 1 段视频序列(a)场景概貌, (b), (c)序列中的 2 帧图像, (d)~(f) 3 个不同视角下重建结果, (g) Google Earth 的截图, (h) 数字高程图

Fig. 5 The first video sequence (a) environment overview, (b), (c) two sample frames in the video sequence, (d)~(f) reconstruction results in three different views, (g) snapshot from Google Earth, (h) digital elevation map

由于无法测得场景的真实结构,也无法得到摄像机在各关键帧处的位姿,为进行定量分析,引入绕行误差,设序列共有 N 个关键帧, $\mathbf{p}_i (i=1, \dots, N)$ 为第 i 个关键帧的摄像机位置坐标,则绕行误差为

$$E = \frac{\|\mathbf{p}_N - \mathbf{p}_1\|}{\left(\sum_{i=2}^N \|\mathbf{p}_i - \mathbf{p}_{i-1}\| \times 100\%\right)}, \quad (17)$$

由(17)式可见,绕行误差越小说明起点与终点重合程度越好。结果见表 1。

图 6 显示了第 2 段视频序列的结果。该段序列相机指向和运动方向基本一致。这种情况对运动估计和场景重建带来了挑战。数字高程图的格点大小为 $0.3 \text{ m} \times 0.3 \text{ m}$ 。该段序列的成功表明算法适用

于地外行星自主车。定量分析结果见表 1。

表 1 两段视频序列的定量分析结果

Table 1 Quantitative analysis of two video sequences

Sequence index	1	2
Loop closure error /%	2.70	3.35
Traveled distance /m	108.45	32
Total frame number	4139	3209
Key frame number	168	203
Number of three-dimensional points in scene structure	16738	22284

由表 1 可见,本文算法的绕行误差较小,与文献[8]中利用立体视觉方法的结果接近,因此具有较好的实用性。

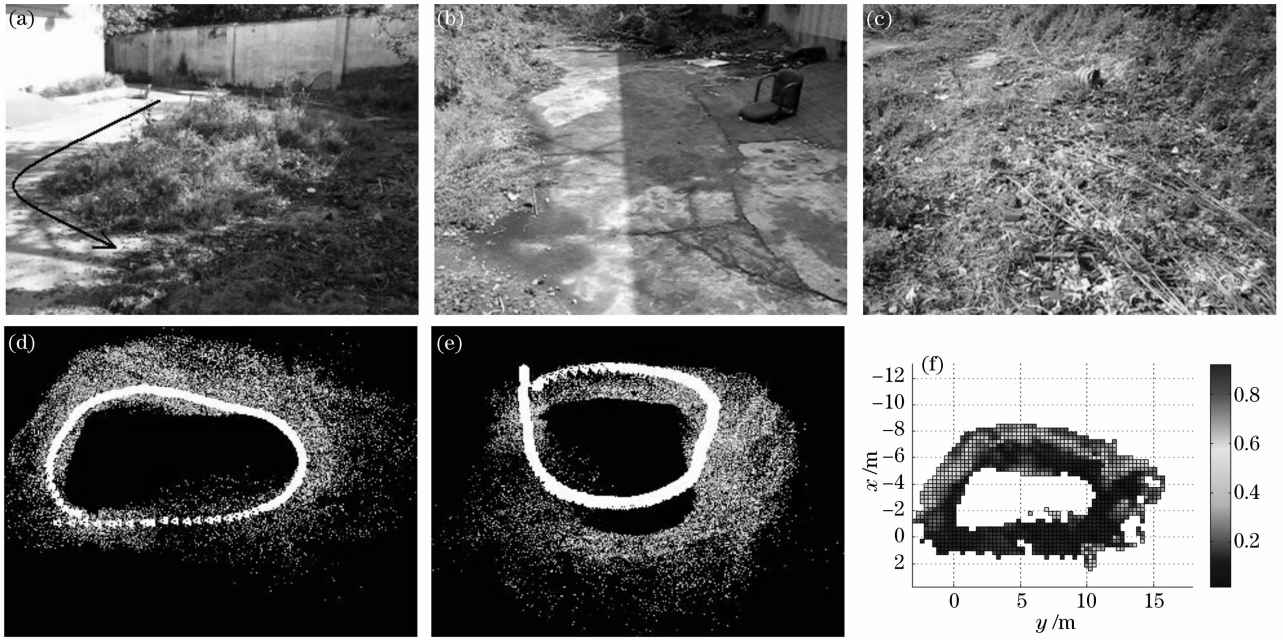


图 6 第 2 段视频序列 (a) 场景概貌, (b), (c) 序列中的 2 帧图像, (d), (e) 2 个不同视角下重建结果, (f) 数字高程图
 Fig. 6 The second video sequence (a) environment overview, (b), (c) two sample frames in the video sequence, (d), (e) reconstruction results in two different views, (f) digital elevation map

3.3 算法耗时分析

算法测试平台为 P4 2.8 GHz 和 1 GB 内存, 先通过摄像机采集数据, 然后传至计算机离线运行。运行时视频序列按串行方式顺序读入以模拟实际运用中视频采集过程。特征跟踪和关键帧提取未进行优化, 跟踪一帧约需 290 ms。对第 1 段序列除特征跟踪和关键帧提取部分进行耗时分析, 结果如图 7 所示。

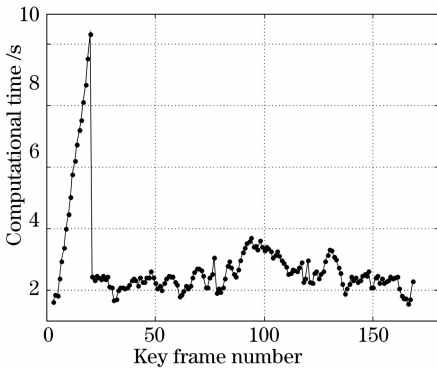


图 7 算法耗时分析

Fig. 7 Analysis of computational time cost

当关键帧个数小于 20 时, 计算时间快速增长, 这说明了对于较长的序列, 利用传统光束法平差运算复杂度极高, 而本文算法在关键帧个数大于 20 时, 处理每一关键帧的时间在 2.5 s 左右, 兼顾了效果和运算复杂度。对于第 1 段序列本文算法总处理时间为 1665.7 s, 折合 $4139 \div 1665.7 \approx 2.5$ (frame/s), 可见本

文算法接近实时。

4 结 论

提出了一种利用单目视觉进行同时场景构建和定位的算法, 该算法利用关键帧提取技术, 大大降低了运算复杂度。提出了一种简便的摄像机位姿估计算法, 避免了传统方法需要进行非线性优化的缺陷, 采用自适应光束法平差在保证优化效果的前提下提高了运算速度。实验结果表明, 该算法在室内和室外多种场景下的绕行误差低于 4%, 且运算速度接近实时, 因此适合于运行速度较低且不能采用激光雷达的场合, 例如: 月球车、火星车等地外行星自主车。

参 考 文 献

- 1 J. J. Leonard, H. F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons[J]. *IEEE Trans. on Robotics and Automation*, 1991, 7(3): 376~382
- 2 D. Hähnel, W. Burgard, D. Fox et al.. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements[C]. *Proc. of IROS*, 2003, 1: 206~211
- 3 Ma Yong, Lin Hong, Ji Hang et al.. Lidar system design based on brim detecting technology[J]. *Chin. J. Lasers*, 2007, 34(2): 170~175
马 泳, 林 宏, 冀 航 等. 基于边缘探测技术的激光雷达系统设计[J]. *中国激光*, 2007, 34(2): 170~175
- 4 Choi J, Ahn S, Chung, W. Robust sonar feature detection for the SLAM of mobile robot[C]. *Proc. of IROS*, 2005. 3415~3420
- 5 M. W. M. Gamini, G. Dissanayake, H. F. Durrant-Whyte. A

- solution to the simultaneous localization and map building (SLAM) problem [J]. *IEEE Trans. on Robotics and Automation*, 2001, **17**(3): 229~241
- 6 B. Kwolek. Visual odometry based on gabor filters and sparse bundle adjustment[C]. *Proc. of ICRA*, 2007. 3573~3578
- 7 A. J. Davison. Real-time simultaneous localization and mapping with a single camera[C]. *Proc. of ICCV*, 2003, **2**: 1403~1410
- 8 D. Nistér, O. Naroditsky, J. Bergen. Visual odometry[C]. *Proc. of CVPR*, 2004, **1**: 652~659
- 9 J. Campbell, R. Sukthankar, I. Nourbakhsh *et al.*. A robust visual odometry and precipice detection system using consumer-grade monocular vision[C]. *Proc. of ICRA*, 2005. 3421~3427
- 10 M. Tomono. 3-D localization and mapping using a single camera based on structure-from-motion with automatic baseline selection [C]. *Proc. of ICRA*, 2005. 3342~3347
- 11 Qiu Zhiqiang, Liang Yonghui, Yu Qifeng. Three-dimensional structure reconstruction from image sequence based on affine approximation[J]. *Acta Optica Sinica*, 2007, **27**(6): 1004~1010
邱志强, 梁永辉, 于起峰. 基于仿射近似从序列图像重建目标三维结构[J]. *光学学报*, 2007, **27**(6): 1004~1010
- 12 Zhao Chuangxin, Xu Jinliang, Zhang Yongli *et al.*. Three-dimensional reconstruction of free-flying insect based on single camera[J]. *Acta Optica Sinica*, 2006, **26**(1): 61~66
赵创新, 徐进良, 张永立 等. 基于单摄像机的昆虫自由飞行参量三维重构[J]. *光学学报*, 2006, **26**(1): 61~66
- 13 Zhengyou Zhang. A flexible new technique for camera calibration [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000, **22**(11): 1330~1334
- 14 Jianbo Shi, C. Tomasi. Good features to track[C]. *Proc. of CVPR*, 1994. 593~600
- 15 R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision* [M]. Cambridge: Cambridge University Press, 2003
- 16 R. I. Hartley. In defense of the eight-point algorithm[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997, **19**(6): 580~593
- 17 M. A. Fischler, R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *ACM Commun.*, 1981, **24**(6): 381~395
- 18 B. Triggs, P. F. McLauchlan, R. I. Hartley *et al.*. Bundle adjustment—A modern synthesis[C]. *In Vision Algorithms: Theory and Practice*, LNCS, 2000. 298~372