

带约束的摄像机定标方法

高俊鑫¹⁾ 华 炜¹⁾ 章国锋¹⁾ 秦学英^{1), 2)}

¹⁾ (浙江大学 CAD&CG 国家重点实验室, 杭州 310027)

²⁾ (山东大学计算机学院, 济南 250100)

摘 要 由于积累误差和摄像机内部参数的校正误差, 基于视觉的摄像机定标结果尽管在图像上的重投影误差很小, 但是重建的空间结构往往存在一定程度的扭曲。提出了一种带约束的摄像机定标算法, 将场景中存在几何条件 (如摄像机路径在一条直线上) 作为约束条件, 对摄像机定标结果进行再优化, 使得重建的空间与真实世界的欧氏空间更为一致。在优化过程中, 将欧氏空间与图像空间的两种误差约束在同一范围, 自动获取最佳的约束系数。实验结果证明了该算法的有效性。

关键词 摄像机定标 路径约束 多视点

中图法分类号: TP391.41 文献标志码: A 文章编号: 1006-8961(2010)03-536-05

Camera Tracking with Constraint

GAO Junxin¹⁾, HUA Wei¹⁾, ZHANG Guofeng¹⁾, QIN Xueying^{1), 2)}

¹⁾ (State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310027)

²⁾ (College of Computer Science and Technology, Shandong University, Jinan 250100)

Abstract Structure reconstructed from motion is usually distorted due to accumulated error and calibration error from extrinsic parameters of video cameras although the re-projected error sometimes can be small. This paper proposes a novel approach to structure from motion with the geometry conditions existing in the scene, such as the trajectory of a video camera lying on a straight line. With this new method, distortion of the reconstructed results can be removed. We also propose a method to automatically select the optimal coefficients between the original cost and the constraint cost, through which the best reconstruction is obtained. The implemented examples demonstrate very precise structure and motion recovery, which prove the effectiveness and robustness of the proposed method.

Keywords structure from motion, path constraint, multiview

0 引 言

摄像机定标技术是计算机视觉领域的一个经典研究内容, 在增强视频和增强现实技术中有着重要的应用。基于结构和运动恢复 (SMR) 的摄像机定

标技术已日趋成熟^[1-4], 并且在实际应用中已体现出其优势。SMR 技术仅利用图像序列中的特征点对应^[5-6], 自动确定摄像机的方位路径, 同时重建了特征点的 3 维结构, 这些 3 维结构构成了场景的稀疏表达。

由于 SMR 技术的 3 维结构计算, 是以 3 维点在

基金项目: 国家高技术研究发展计划 (863) 项目 (2007AA01Z326); 国家自然科学基金项目 (60870003); 浙江大学 CAD&CG 国家重点实验室开放课题项目 (A0809)

收稿日期: 2008-10-15 改回日期: 2008-12-25

第一作者简介: 高俊鑫 (1984—), 男。2008 年于浙江大学获计算机应用专业硕士学位。主要研究方向为计算机图形学、计算机视觉。Email: gaojunxin@cad.zju.edu.cn

通讯作者: 秦学英。Email: xyqi@cad.zju.edu.cn

图像上的重投影误差最小化为约束条件的^[3-4], 因此, 其重建的 3 维空间与真实的欧氏空间的一致性, 容易受到误差等因素的干扰, 从而使得重建空间存在一定程度的扭曲。用自主研发的摄像机标定软件^[4-7], 对在摄像机沿平直公路直线运动情况下拍摄的视频进行标定, 发现标定结果的摄像机路径容易变为曲线, 并且重建的 3 维空间也存在一定程度的扭曲。即使对相机预先进行了精确的非线性畸变校正^[8], 标定结果的路径扭曲和空间扭曲依然明显存在。尤其是当视频达到上千帧时, 这种扭曲将会变得更加明显。这些扭曲主要来自标定中的特征点匹配、相机校正以及计算误差的积累, 很难完全避免。由于重建的稀疏结构空间将决定 3 维密集重建的几何框架^[9], 因此, 其空间与真实空间的一致性非常重要。

针对 SMR 技术的积累误差, Fitzgibbon 采用闭合环路序列产生的约束来消除^[10], 但该技术不能用于没有闭合环路的情况。由于真实世界的几何是可测量的, 因此, 重建空间的真实程度是可验证的。但是测量的方式非常繁琐, 需要简便快捷的方法。

由于 SMR 技术的进步, 自动构建 3 维物体的精确表面几何正在成为现实^[9-11]。同时, 该技术也在其他领域有广泛的应用前景, 如多视点的全景图拼接^[12-13]。传统的基于图像的绘制技术^[14], 特别是全景图技术, 在图像注册时, 视点位置基本固定。SMR 技术使得基于图像的绘制技术可以处理更为复杂的情形。

本文提出了一种新的带约束的标定算法。摄像机与场景的 3 维结构, 往往存在明显的特征, 如相机位置在一条近似直线上、场景中的一些特征点在一条直线或者平面上等。对拍摄路径为直线的情况进行约束, 使得空间的扭曲得以消除。这样的方法也可推广到其他线性约束。

1 摄像机标定模型

摄像机运动状态, 一般用 7 个参数来表示, 即旋转参数 (用欧拉角表示) $\Theta_i = (\theta_x, \theta_y, \theta_z)$, 位移向量 $t_i = (t_x, t_y, t_z)$ 以及焦距参数 f_i 。可将内部参数矩阵表示如下:

$$K_i = \begin{bmatrix} f_i & 0 & c_x \\ 0 & f_i & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

式中, (c_x, c_y) 是图像中心。这里的内部参数矩阵略去了纵横比, 并假设横轴和纵轴正交。

图像序列中的每个跟踪点代表了一个空间 3 维点 $X_j = (X_1, X_2, X_3, 1)^T$, 其在每帧画面上的投影表达为

$$u_{ij} = P_i X_j = K_i (R_i | t_i) X_j \quad (2)$$

$$i = 1, \dots, n; \quad j = 1, \dots, m$$

式中, u_{ij} 表示 X_j 在第 i 帧上投影点位置的齐次坐标; P_i 为第 i 帧的透视投影矩阵, R_i 为具有 3 个自由度的旋转矩阵, 由 3 个欧拉角决定, t_i 为平移向量, n 为图像序列的帧数, m 为 3 维点的个数。 u_{ij} 是通过特征点匹配来获得其图像坐标的, 即凡是特征匹配上的点, 就认为是同一个 3 维点的投影。因此, 结构和运动恢复 (SMR) 技术, 是通过 u_{ij} 的关联来获取所有的 $R_i, t_i, K_i, i = 1, \dots, n$, 以及 $X_j, j = 1, \dots, m$ 。

2 摄像机标定中的约束

在进行摄像机标定之前, 预先校正了待标定图像的径向畸变, 因为径向畸变也会导致 3 维空间的扭曲。采取定标盘或定标盒的标定方法^[8], 可以获得较好的效果, 但是难以完全消除相机的畸变。

结构和运动恢复技术, 根据 3 维点及其帧画面上对应的 2 维图像位置, 通过最小化如下目标函数求解这些帧的摄像机参数:

$$S_p = \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^m d(u_{ij}, \hat{u}_{ij})^2 \quad (3)$$

式中, S_p 为视频序列上所有参与计算的特征点的重投影误差, 函数 d 为图像点间的欧氏距离, u_{ij} 为特征点的图像位置, \hat{u}_{ij} 为重建后的 3 维点 \hat{X}_j 在第 i 帧图像上的重投影坐标。因此, 根据式 (2), 有:

$$\hat{u}_{ij} = K_i (R_i \hat{X}_j + t_i)$$

经典的 SMR 技术, 是通过最小化函数 (3), 获得摄像机参数 R_i, t_i, K_i 以及特征点的 3 维位置, 并且一般已经采用集束调整^[14-15] 以保证获得最优解。但是, 这样求得的解, 仍然在空间上有一定程度的扭曲。因此, 对 3 维空间进行进一步约束, 其约束主要来自场景的 3 维结构, 如相机的 3 维空间位置。实例中, 是在一条平直的公路上拍摄的视频序列, 因此, 相机的位置应该近似地在一条直线上。采用主元分析 (PCA) 的方法, 对这条直线进行拟合, PCA 的第一主元的方向即为直线的方向, 然后取所有摄像机位置点的平均点为直线上一点。设这条直线为 l , 相机的中心位置记为 C_i , 且有 $C_i = -R_i^T t_i$ 。所有

相机位置偏离该直线的误差为 $S_c = \frac{1}{n} \sum_{i=1}^n D^2(l, C_i)$,

其中, S_c 表示约束误差, 函数 D 表示点 C_i 到直线 l 的距离。附加约束条件, 即同时使得 S_p 和 S_c 最小来优化 SMR 的各项参数。根据拉格朗日法, 即要使得下式最小化:

$$S = S_p + \lambda S_c = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m d(u_{ij}, u'_{ij})^2 + \frac{\lambda}{n} \sum_{i=1}^n D^2(l, C_i) \quad (4)$$

式中, λ 为拉格朗日系数。 λ 越大, 直线越直。

这是一个高度复杂的非线性优化问题, 其变量个数与 SMR 问题的变量相同。由于 SMR 法已经获得较为精确的参数, 因此成为求解的一个很好的初值。假设优化的摄像机位置数为 n , 优化的 3 维点个数为 m , 则优化的变量个数为 $6n + 3m$, 由于计算中涉及到雅可比矩阵的求逆计算, 和矩阵乘法运算, 所以优化计算量与未知数个数成平方关系, 即 $O(6n + 3m)^2$ 。因此, 采用关键帧及剔除多余的 3 维点, 可以使得优化函数尽可能地简化, 但其计算量仍然庞大。

不带约束的集束调整利用了优化函数式 (3) 的性质, 以及雅可比矩阵的稀疏性, 可以利用矩阵性质推导一个简便的求解方法, 进行快速计算。但是, 新的约束条件的加入, 破坏了这些性质, 使得传统的集束调整不能够顺利进行。

本文算法采取的标定策略是基于关键帧的标定方法^[3], 在作整体优化时, 采用非线性的优化算法 Levenberg-Marquardt 方法^[16], 直接利用稀疏矩阵的性质^[7], 而非集束调整的矩阵性质, 来获取快速精确的解。由于不同的拉格朗日系数 λ 会获得不同的解, 那么其取值会影响计算结果, 因此需接着讨论如何选取 λ 。

3 参数 λ 的选取

一般真实的摄像机拍摄路径并非是严格位于一条直线上的, 由于无法避免的抖动, 摄像机往往偏离一条直线。因此, λ 的选取是这个问题的关键, 因为 λ 选得过小, 起不到约束的作用, 但 λ 取得过大, 会造成过矫正, 导致路径被过分拉直, 使得重投影误差过大。

在代价函数 (式 (4)) 中, 由于 S_p 是结构点的重投影误差, 以像素平方为单位, 而 S_c 是摄像机空间位置到拟合直线的距离, 二者的绝对数值大小不同,

量纲也不同。由于 S_p 和 S_c 在同一个事件中, 误差的来源是一样的, 而 λ 作为一个加权参数, 如果要使 S_p 和 S_c 放在一起优化, 就应使 S_p 和 λS_c 的误差在同一水平。因此, 首先单独优化 $\min(S_p)$ 和 $\min(S_c)$, 从而分别得到其残差 $S_p^{min} = \min(S_p)$ 和 $S_c^{min} = \min(S_c)$ 。由于 S_p 和 S_c 的理想优化值是 0, 那么 S_p 和 S_c 的绝对误差分别是 S_p^{min} 和 S_c^{min} , 这样取 $\lambda_0 = S_p^{min} / S_c^{min}$, 就可以使 S_p 和 λS_c 的误差处于同一水平。

利用此 λ_0 , 再次优化函数 $S = S_p + \lambda_0 S_c$, 可以获得一组解, 这组解将最大程度地满足了要求。

4 实验结果与讨论

为了验证算法的有效性, 在校园的平直道路上拍摄了一段视频, 首先把视频解析成图片, 然后采用 OpenCV 对图片进行径向畸变校正, 最后用本算法进行标定。实验用的计算机配置为 P4 2.4 GHz CPU, 2 048 M 内存, 显卡为 Nvidia 6 800。

为了保证相机位置的平直, 将相机用三脚架固定, 并放在小轿车内, 将方向盘放平后, 沿着一条笔直道路前行。拍摄场景的前景是路边的低矮灌木丛和沿线种植的树木, 均大致在一条直线上。这段视频一共有 450 帧, 在图 1 中, 选取了第 1 200 400 帧展示空间的 3 维场景, 给出了其视频截图, 以便了解大致的场景架构。采用传统的方法, 匹配点个数一共 2 238 个, 关键帧个数 41 个, 未进行路径约束前重投影误差为 0.498 3 个像素。图 2 显示了传统方法重建的场景 3 维结构, 在结果图中画出了摄像机轨迹 (实线) 和跟踪点 (离散点) 在欧氏空间的 3 维位置。可以看出, 在未加约束的情况下, 相机路径和 3 维空间出现了明显的弯曲。

采用本文提出的方法, 首先求得 $S_p^{min} = 342.016$, $S_c^{min} = 202.952$, 算得 $\lambda = 1.685$ 。再对式 (4) 进行优化, 计算得重投影误差为 0.536 4 获得的重建结果如图 3 所示。从结果中可以看出, 虽然重投影误差稍微变大, 但是在路径和空间整体上的结构得到明显改善, 摄像机路径和 3 维空间恢复了接近正常的状态, 基本上达到平直。又注意到, 在 λ 确定后的优化过程中, 对空间中的所有 3 维位置进行了一次集束调整, 也即对相机参数和特征点的空间位置都同时进行了修正, 而 λ 的选择策略保证了获得的结果是最优的。

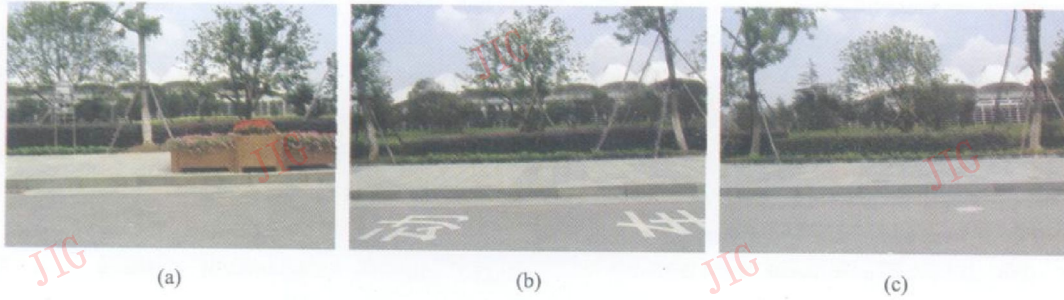


图 1 视频序列中的帧画面
Fig. 1 Three frame images from the video sequence

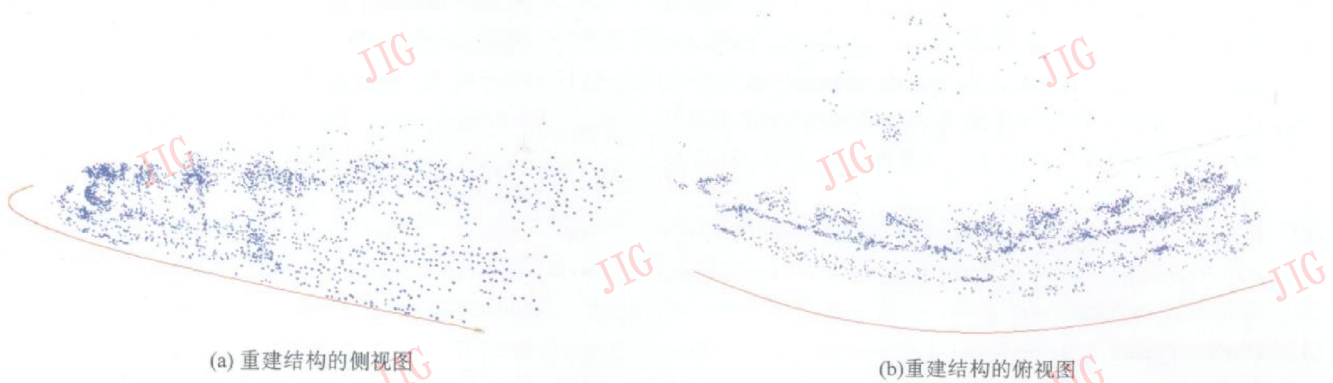


图 2 未进行路径约束的摄像机标定结果
Fig. 2 Result without path constraint

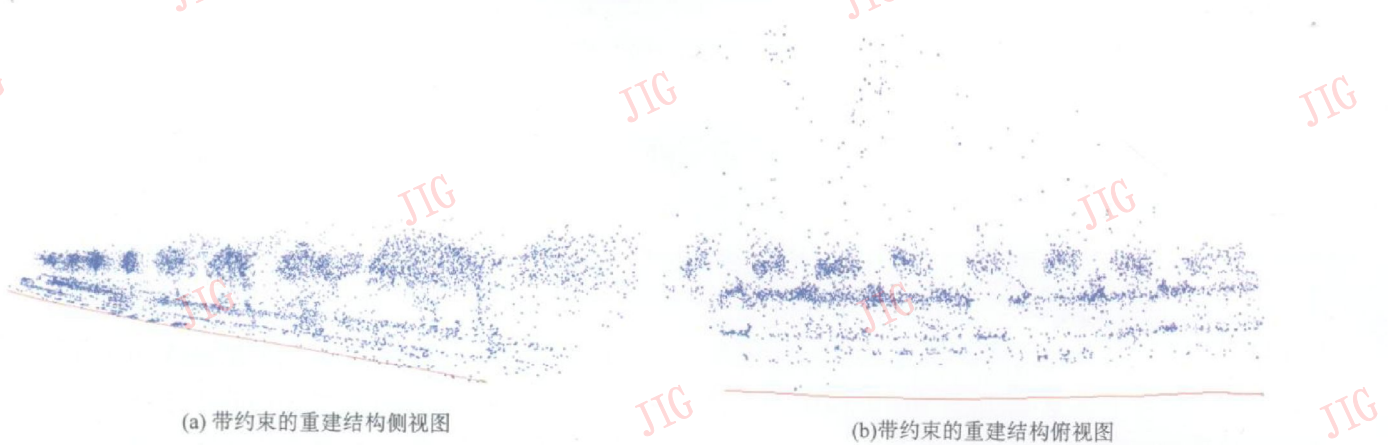


图 3 $\lambda = 1.685$ 时路径约束后的摄像机标定结果
Fig. 3 Result of path constraint with $\lambda = 1.685$

通过理论分析和实践验证了关于 λ 选取策略的有效性。式 (4) 给出的优化函数中并未对路径约束的具体形式给予限定, 虽然例子中是对直线约束进行了讨论与求解, 但是优化式可以推广到其他形式的路径约束中。

5 结 论

提出了一种带约束的摄像机标定的优化算法, 并对加权参数的选取进行了详细的讨论。最后实验

结果显示了算法的有效性, 使得空间扭曲的摄像机标定结果得到较好的矫正。

参考文献 (References)

- [1] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision[M]. Cambridge, UK: Cambridge University Press, 2000.
- [2] Fitzgibbon A, Zisserman A. Automatic Camera Tracking[M]. In Shah M and Kumar R, ed. Video Registration, Chapter 2. UK: Kluwer Academic Publishers, 2003: 18-35.
- [3] Pollefeys M. Visual 3D Modeling from Images[EB/OL]. [2008-

- 10-10]. <http://www.cs.unc.edu/~marc/tutorial.pdf>
- [4] Zhang G F, Qin X Y, Hua W, et al. Robust metric reconstruction from challenging video sequences [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis Minnesota USA: IEEE Press, 2007: 1-8.
- [5] Shi J B, Tomasi C. Good features to track [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE Computer Society, 1994: 593-600.
- [6] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [7] Zhang Guofeng, Qin Xueying, Dong Zilong, et al. Camera tracking based on structure and motion recovery for augmented video [J]. *Chinese Journal of Computers*, 2006, 29(12): 2104-2111. [章国锋, 秦学英, 董子龙, 等. 面向增强视频的基于结构和运动恢复的摄像机定标 [J]. *计算机学报*, 2006, 29(12): 2104-2111.]
- [8] Zhang Z Y. A flexible new technique for camera calibration [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- [9] Zhang G F, Jia J Y, Wong T T, et al. Recovering consistent video depth maps via bundle optimization [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska USA: IEEE Computer Society, 2008: 1-8.
- [10] Fitzgibbon A, Zisserman A. Automatic camera recovery for closed or open image sequences [C] // Proceedings of the Computer Vision-ECCV'98. Lecture Notes in Computer Science. Berlin: Springer-Verlag, 1998: 1406-311-326.
- [11] Pollefeys M, Gool L V, Vergauwen M. Visual modeling with a hand-held camera [J]. *International Journal of Computer Vision*, 2004, 59(3): 207-232.
- [12] Agarwala A, Agrawala M, Cohen M, et al. Photographing long scenes with multi-viewpoint panoramas [J]. *ACM Transactions on Graphics*, 2006, 25(3): 853-861.
- [13] Rav-Acha A, Engel G, Peleg S. Minimal aspect distortion (MAD) mosaicing of long scenes [J]. *International Journal of Computer Vision*, 2008, 78(7): 187-206.
- [14] Shum H Y, Kang S B. A review of image-based rendering techniques [C] // Proceedings of the IEEE/SPIE Visual Communications and Image Processing. Perth, Australia: SPIE, 2000: 20-23.
- [15] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle adjustment—a modern synthesis [C] // Proceedings of Workshop on Vision Algorithms. Freiburg, Germany: Springer-Verlag, 1999: 298-372.
- [16] Heraklion, Crete. Levenberg-Marquardt nonlinear least squares algorithms in C/C++ [EB/OL]. [2008-03-23]. <http://www.ics.forth.gr/~bourakis/levmar/>.