

文章编号 1004-924X(2008)10-1982-06

估计多视点摄像机姿态的两步法

向长波^{1,2}, 谢 丹³, 刘太辉^{1,2}, 宋建中¹

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033 ;

2. 中国科学院 研究生院, 北京 100039 ; 3. 湖南师范大学 工学院, 湖南 长沙 410081)

摘要:提出了一种估计多个视点位置上摄像机姿态的新算法,即两步估计法。该方法采用两幅图像的特征匹配点重建出物体的局部模型,以局部模型作为世界坐标系的中心,利用模型的三维特征点与图像特征点的对应关系,求解出每个视点位置上的摄像机姿态。模拟数据与真实图像的实验结果表明,在特征点定位误差 <1 pixel的情况下,摄像机的平均方位误差 $<0.5^\circ$,平均位置误差 <0.1 cm,模型重投影误差 <1 pixel。

关键词:摄像机;姿态估计;两步法;三维重建;三维测量

中图分类号:TP391 文献标识码:A

A two-step algorithm for estimating postures of cameras located in different points of view

XIANG Chang-bo^{1,2}, XIE Dan³, LIU Tai-hui^{1,2}, SONG Jian-zhong¹

(1. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;*

2. *Graduate University of Chinese Academy of Sciences, Beijing 100039, China;*

3. *College of Engineering, Hunan Normal University, Changsha 410081, China*)

Abstract: A two-step algorithm to estimate the postures of cameras located in different points of view was proposed. The local 3D model of an object is reconstructed with matching points in two images. Then, by taking the local 3D models as the center of world coordinate system, the posture of every camera is estimated according to the correspondences between the 3D feature points and the image feature points of a object. The experimental results of synthetic data and real images show that the average orientation error, average location error and the re-projection model error of the camera are less than 0.5° , 0.1 cm and 1 pixel, respectively, under the condition of the position error for feature point in less than 1.0 pixel.

Key words: camera; pose estimation; two-step method; 3D reconstruction; 3D measuring

1 引 言

采用一台或多台摄像机从多个角度拍摄物体,然后利用获得的图像进行物体的三维建模与非接触测量是一个集计算机视觉、数字摄影测量、逆向工程技术为一体的重要研究方向,在国内外已经引起了广泛重视^[1-4]。

利用多幅图像进行物体的三维建模与测量,首先需要解决每个视点位置上摄像机的姿态估计问题。所谓摄像机的姿态,是指摄像机与被摄物体之间的相对旋转和平移关系,它的精确求解直接关系到最后三维模型重建与测量的精度。目前,经常采用序列分层合并的方法^[5-7]来估计每个视点位置上摄像机的姿态。这种方法以两幅或三幅图像为单位将多幅图像划分成相互重叠的多个子序列,通过求解子序列内部图像之间的基本矩阵或三焦点向量来计算每个摄像机的相对姿态,最后以第一个摄像机作为世界坐标系的原点进行合并操作。显然,序列分层合并方法必然存在误差的累积问题。

为了避免上述的累积误差,本文提出了用两步法来估计各个视点位置上摄像机的姿态。算法的第一步将第一个视点位置作为世界坐标系的原点,采用第一、二幅图像上的特征匹配点重建出物体的局部模型;算法的第二步将世界坐标系的原点移到物体局部模型上,以局部模型上的空间点与图像特征点的对应关系来求解每个视点位置上的摄像机姿态。这种两步法估计出的每个摄像机姿态都是相对于被重建物体而言的,彼此之间相对独立。因此,不存在误差累积问题,从而在一定程度上提高了姿态估计的精度。

2 重建物体的局部模型

利用两幅图像求解物体的局部模型,首先需要计算出拍摄这两幅图像时摄像机的投影矩阵 \mathbf{P} 。投影矩阵由下式定义

$$\mathbf{P}=\mathbf{K}[\mathbf{R}|\mathbf{t}], \quad (1)$$

其中, \mathbf{K} 是 3×3 的内部参数矩阵, \mathbf{R} 是 3×3 的旋转矩阵, \mathbf{t} 是平移矢量。 \mathbf{R} 和 \mathbf{t} 描述了摄像机在拍摄图像时的方位和姿态。通常可以假定第一个摄像机位于坐标系的原点,即第一个摄像机的投

影矩阵 $\mathbf{P}_1=\mathbf{K}[\mathbf{I}|0]$ 。第二个摄像机相对于第一个摄像机的旋转矩阵和平移矢量可以通过两幅图像的本质矩阵来求解。

设两幅图像上的匹配点为 $(\mathbf{x}, \mathbf{x}')$, 则本质矩阵的方程^[8]定义为:

$$\hat{\mathbf{x}}'^T \mathbf{E} \hat{\mathbf{x}}=0, \quad (2)$$

式中, $\hat{\mathbf{x}}=\mathbf{K}^{-1} \mathbf{x}$ 是图像点 \mathbf{x} 在归一化坐标下的表示。容易看出,采用 8 对图像匹配点可以线性求出本质矩阵 \mathbf{E} 。可以证明^[8], \mathbf{R} 、 \mathbf{t} 和 \mathbf{E} 之间存在 $\mathbf{E}=[\mathbf{t}]_x \mathbf{R}$ 的关系。因此,一旦本质矩阵已知,通过把 \mathbf{E} 分解成反称矩阵 $\mathbf{T}=[\mathbf{t}]_x$ 和旋转矩阵 \mathbf{R} 的乘积形式^[9-10], 就可以恢复出第二个摄像机的投影矩阵 $\mathbf{P}_2=\mathbf{K}[\mathbf{R}|\mathbf{t}]$ 。

知道摄像机的投影矩阵 \mathbf{P}_1 、 \mathbf{P}_2 和两幅图像之间的特征匹配点,采用线性三角形法就可以重建出物体上的三维特征点,获得物体的局部模型。

设 $\mathbf{X}=(X, Y, Z, 1)^T$ 为物体上的某个点,它在两幅图像上都可见,其对应的图像点坐标记为 $\mathbf{x}_1=(x_1, y_1, 1)^T$ 和 $\mathbf{x}_2=(x_2, y_2, 1)^T$ 。每个图像点 $\mathbf{x}_i(i=1, 2)$ 与 \mathbf{X} 之间的关系满足公式 $\mathbf{x}_i=\mathbf{P}_i \mathbf{X}$ 。由该公式可以看出,每个图像点给出了关于 \mathbf{X} 坐标值的两个线性独立方程。这样,一对图像点就有 4 个方程构成了关于 \mathbf{X} 的线性方程组,通过叉乘运算 $\mathbf{x}_i \times (\mathbf{P}_i \mathbf{X})=0$ 可以得到形如 $\mathbf{A} \mathbf{X}=0$ 的矩阵表达式,其中

$$\mathbf{A}=\begin{bmatrix} x_1 \mathbf{P}_1^{3T} & -\mathbf{P}_1^{1T} \\ y_1 \mathbf{P}_1^{3T} & -\mathbf{P}_1^{2T} \\ x_2 \mathbf{P}_2^{3T} & -\mathbf{P}_2^{1T} \\ y_2 \mathbf{P}_2^{3T} & -\mathbf{P}_2^{2T} \end{bmatrix}, \quad (3)$$

上式中, \mathbf{P}_i^{jT} 是第 i 个投影矩阵的第 j 行。对矩阵 \mathbf{A} 进行奇异值分解就可以获得方程 $\mathbf{A} \mathbf{X}=0$ 的最小二乘解 \mathbf{X} 。

3 估计每个摄像机的姿态

获得物体的局部模型后,利用物体上特征点与图像点之间的对应关系,采用 POSIT 算法^[11]估计出每个视点的摄像机姿态。POSIT 算法首先以弱透视投影模型获得像机姿态的近似解,然后采用循环迭代的方法寻找最优解,它的成像关系如图 1 所示。

O 为摄像机中心,物体坐标系 UVW 的原点与空间点 X_0 重合, X_0 在图像平面上的像点为

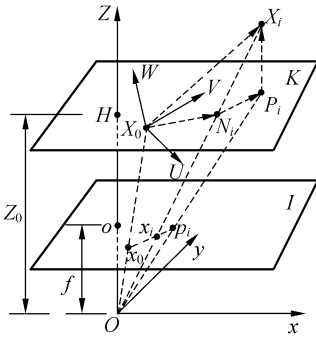


图 1 弱透视投影

Fig. 1 Weak perspective projection

x_0 。因此,摄像机的平移矢量

$$t = OX_0 = \frac{Z_0}{f} \cdot OX_0$$

设 i, j 和 k 分别代表摄像机坐标系 X, Y, Z 轴的单位矢量,根据坐标系之间的变换关系,摄像机的旋转矩阵可以表示为:

$$R = \begin{bmatrix} i_U & i_V & i_W \\ j_U & j_V & j_W \\ k_U & k_V & k_W \end{bmatrix}$$

i_U, i_V, i_W 是单位矢量 i 在模型坐标系中的坐标。由于 $k = i \times j$,如果能够求解出矢量 i, j 在模型坐标系中的坐标,就可以完全确定旋转矩阵 R 。

为了求解 i, j 和 Z_0 ,以弱透视投影模型来近似透视投影。弱透视摄像机将特征点 X_i 正交投影到过 X_0 且与图像平面 I 平行的平面 K 上,得到点 P_i 。紧接着将 P_i 通过摄像机的中心 O 透视投影到图像平面 I 上,得到像点 p_i ,该像点称为点 X_i 的弱透视像点。连接摄像机中心 O 和 X_i 的射线交平面 K 于 N_i 点,交图像平面 I 于 x_i ,可以看出 x_i 是 X_i 的透视像点。可以证明,在上述成像关系中,有下面两个等式成立。

$$\begin{aligned} X_0 X_i \cdot I &= x_i(1 + \epsilon_i) - x_0 \\ X_0 X_i \cdot J &= y_i(1 + \epsilon_i) - y_0 \end{aligned} \quad (4)$$

其中, $\epsilon_i = \frac{1}{Z_0} X_0 X_i \cdot k, I = \frac{f}{Z_0} i, J = \frac{f}{Z_0} j$ 。由式(4)可知,每对空间点 X_i 和图像点 x_i 提供了关于矢量 I, J 的一个约束条件。因此,在已知的情况,只要三个不共面的空间点就可以线性求出矢量 I, J ,从而进一步解出 $i = \frac{I}{\|I\|_2}, j = \frac{J}{\|J\|_2}$ 和 $Z_0 = \frac{f}{\|I\|_2}$ 。为了求解 ϵ_i ,可以在最初令所有的 $\epsilon_i =$

0。一旦得到了近似的 i, j, Z_0 ,就可以采用循环迭代的方法逐渐地获得 ϵ_i 的更精确值。当 ϵ_i 的值趋于稳定时,迭代过程结束,最后的 R, t 就是所求摄像机在物体坐标系中的姿态。

4 算法流程

算法的具体实现流程如下:

第一步:以第一个视点位置作为世界坐标系的中心,求解物体的局部模型。

(1)采用平面模板的标定方法^[12-13]计算摄像机的内部参数 K ;

(2)采用鲁棒贪心估计算法^[14]计算两幅图像的基本矩阵 F 和本质矩阵 E ;

(3)将本质矩阵分解成 $E = [t]_x R$ 的乘积形式^[9],获得摄像机在第二个视点位置上的投影矩阵 $P_2 = K[R|t]$ 。

(4)利用投影矩阵 P_1, P_2 和两幅图像上的特征匹配点 (x_i, x_i') 构建式(3)的矩阵 A ,通过奇异值分解得到空间点 X_i ,多个空间点 X_i 构成了物体的局部模型。

第二步:以物体局部模型为坐标系的中心,分别求解每个视点位置上的摄像机姿态。

(1)以物体局部模型上的某一点(假设为点 X_0)作为世界坐标系的中心,计算模型上其它点在该世界坐标系下的坐标 $\bar{X}_i = X_i - X_0$;

(2)假设在物体局部模型上有 m 个点在当前图像中可见,其中 $m \geq 3$ 。建立这 m 个空间点 \bar{X}_i 与图像点 x_i 之间的对应关系 (\bar{X}_i, x_i) ;

(3)将 m 对 (\bar{X}_i, x_i) 和当前的 ϵ_i 值代入式(4)中,通过奇异值分解获得 I, J 的最小二乘解;

(4)由式 $i = \frac{I}{\|I\|_2}, j = \frac{J}{\|J\|_2}, k = i \times j$ 和 $Z_0 = \frac{f}{\|I\|_2}$ 得到矢量 i, j, k 和标量 Z_0 的解;

(5)更新所有的 $\epsilon_i = \frac{1}{Z_0} X_0 X_i \cdot k = \frac{1}{Z_0} \bar{X}_i \cdot k$ 。设更新前的 ϵ_i 为 ϵ_{i-old} ,更新后的 ϵ_i 为 ϵ_{i-new} 。如果 $|\epsilon_{i-new} - \epsilon_{i-old}|$ 小于某个阈值,程序执行下一步,否则转到第(3)步;

(6)当前的 i, j, k 就是所求旋转矩阵 R 的 3 个行矢量,需要求解的平移矢量 $t = \frac{Z_0}{f} \cdot x_0$ 。

5 实验结果

首先以一组真实图像数据来直观地检测算法的有效性;然后,通过一组模拟数据来定量地检测算法在不同噪声等级下的鲁棒性和所求摄像机姿态的精确性。

5.1 图像数据的实验结果

图 2 是采用 CANON 520 数码相机从 4 个角

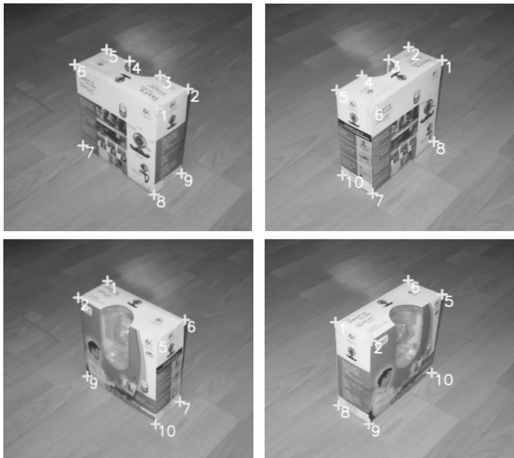


图 2 4 个角度拍摄图像

Fig. 2 Images taken by a camera located in four viewpoints

度拍摄的盒子图像,图像上的白色十字所示位置是选择的特征点。摄像机的内部参数采用平面模板的标定方法求解,其中图像中心坐标 $(u_0, v_0) = (349.10, 207.21)$,等效焦距 $f = 880.895$ pixel。

算法第一步采用图 2 前两幅图像的 8 对特征点重建物体的局部模型,结果如表 1 所示。局部模型与两个摄像机的方位关系,如图 3 所示,其中,世界坐标系 XYZ 与第一个摄像机 YC1 重合。

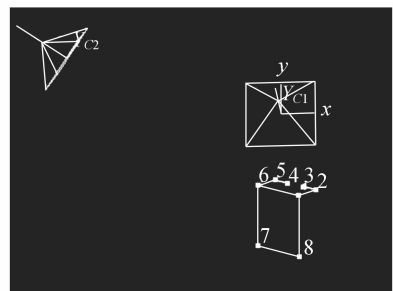


图 3 两个摄像机与局部模型的方位关系

Fig. 3 Position relations between two cameras and local model

表 1 物体局部模型的重建结果

Tab. 1 Reconstructed results of object's local model

特征点的标号	特征点三维坐标	特征点的标号	特征点三维坐标
1	(0.014 6, -0.024 5, -0.664 1)	5	(-0.077 8, 0.092 8, -0.842 8)
2	(0.070 0, 0.011 3, -0.716 7)	6	(-0.133 9, 0.055 1, -0.788 0)
3	(0.024 3, 0.035 8, -0.755 3)	7	(-0.143 2, -0.118 3, -0.924 8)
4	(-0.030 9, 0.066 0, -0.802 4)	8	(0.008 2, -0.197 5, -0.800 5)

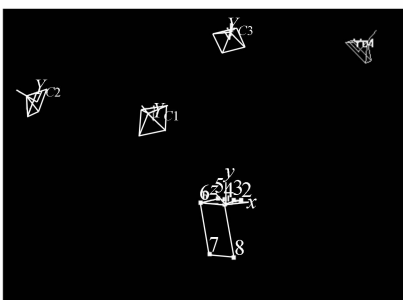


图 4 4 个摄像机与局部模型的方位关系

Fig. 4 Position relations between four cameras and local model

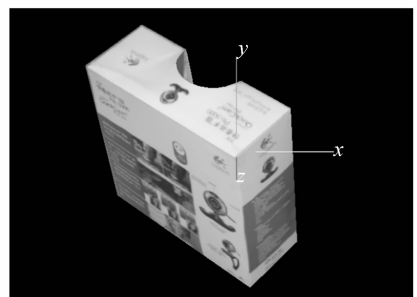


图 5 重建的三维纹理模型

Fig. 5 Reconstructed 3D texture-mapped model

算法第二步以物体局部模型作为世界坐标系的中心,计算得到的 4 个摄像机与局部模型的方位关系如图 4 所示,利用所求 4 个摄像机姿态重建的盒子三维纹理模型如图 5 所示,模型上特征点的重投影误差为 0.356 3 pixel。

5.2 模拟数据的实验结果

以三维结构已知的景物作为场景,投影到内部参数相同、外部参数不同的 6 个模拟摄像机中,即可以获得无噪声模拟图像。再将这些图像分别

加上均值为 0,标准差为 0.5,1.0,1.5 的高斯噪声(单位为像素)来测评算法的鲁棒性。

采用两步法和序列分层合并法估计的 6 个摄像机的平均方位误差、平均位置误差如表 2 所示。在频率 1.10 GHz 的 CPU 上,两步法的运行时间为 0.031 s,序列分层合并法的运行时间为 0.075 s。从实验数据可以看出,本文所提出的两步法在姿态参数的估计精度、运行时间、抗噪声能力等方面均优于序列分层合并算法。

表 2 摄像机姿态的测量误差

Tab.2 Measuring errors of camera posture

标准方差	两步法		序列分层合并法	
	平均方位误差(°)	平均位置误差(cm)	方位绝对误差(°)	平均位置误差(cm)
0.5	0.43	0.05	0.79	0.37
1.0	0.96	0.17	1.32	0.52
1.5	1.71	0.35	2.38	0.81

6 结 论

本文提出了两步法来估计多个视点位置上摄像机姿态。该算法首先用两幅图像重建出物体的局部模型,然后以物体的局部模型为坐标系的中心,计算每个摄像机的姿态。由于每个摄像机的姿态都是相对于被重建物体而言的,彼此之间

相对独立。因此,所提出的算法克服了序列分层合并方法中存在的误差累积问题。大量的模拟数据与真实图像的实验结果表明,所提算法可以提高姿态参数的估计精度,减少运行时间。在特征点定位误差 <1 pixel 的情况下,计算出来的摄像机平均方位误差 $<0.5^\circ$,平均位置误差 <0.1 cm,模型的重投影误差 <1 pixel。

参考文献:

- [1] WU H, YU Y ZH. Photogrammetric reconstruction of free-form objects with curvilinear structures[J]. *The Visual Computer*, 2005, 21(4): 203-216.
- [2] DEBEVEC P E, TAYLOR C J, MALIK J. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach[C]. *Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, New Orleans, Louisiana*, 1996: 11-20.
- [3] QUAN L, TAN P, ZENG G. Image-based plant modeling[J]. *ACM Transactions on Graphics*, 2006, 25(3): 646-653.
- [4] 叶海加, 陈昱, 邢渊. 双目 CCD 结构光三维测量系统中的立体匹配[J]. *光学 精密工程*, 2004, 12(1): 71-75.
YE H J, CHEN G, XING Y. Stereo matching in 3D measurement system using double CCD structured light[J]. *Opt. Precision Eng.*, 2004, 12(1): 71-75. (in Chinese)
- [5] POLLEFEYS M, GOOL L V, VERGAUWEN M, et al. Visual modeling with a hand-held camera[J]. *International Journal of Computer Vision*, 2004, 59(1): 207-232.
- [6] RAMALINGAM S, LODHA K S, STURM P. A generic structure-from-motion framework[J]. *Computer Vision and Image Understanding*, 2006, 103(3): 218-228.
- [7] 陈泽志. 由非定标图像序列重建和测量三维物体[D]. 西安: 西安电子科技大学, 2002.

- CHEN Z ZH. *Reconstruction and metrology of 3D objects from uncalibrated image sequences*[D]. Xian: Xidian University, 2002. (in Chinese)
- [8] 韦穗,杨尚骏,章权兵,等. 计算机视觉中的多视图几何[M]. 合肥:安徽大学出版社, 2002.
WEI S, YANG SH J, ZHANG Q B, *et al.*. *Multiple View Geometry in Computer Vision*[M]. Hefei: Anhui University Press, 2002. (in Chinese)
- [9] HARTLEY R I. Estimation of relative camera positions for uncalibrated cameras[C]. *Proceedings of the European Conference on Computer Vision, Santa Margherita*, 1992:579-587.
- [10] HORN B K P. Recovering baseline and orientation from essential matrix[OL]. <http://people.csail.mit.edu/bk-ph/articles/Essential.pdf>
- [11] DANIEL F, DEMENTHON, DAVIS L S. Model-based object pose in 25 lines of code[J]. *International Journal of Computer Vision*, 1995, 15(2):123-141.
- [12] ZHANG ZH Y. A flexible new technique for camera calibration[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- [13] STURM P, MAYBANK S J. On plane-based camera calibration: a general algorithm, singularities, applications [C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Colorado, USA*, 1999: 432-437.
- [14] 向长波, 刘太辉, 宋建中. 基本矩阵的鲁棒贪心估计算法[J]. 计算机辅助设计与图形学学报, 2007, 19(5): 651-655.
XIANG CH B, LIU T H, SONG J ZH. A robust greedy algorithm for estimating the fundamental matrix[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2007, 19(5): 651-655. (in Chinese)

作者简介:向长波(1978—),男,湖南怀化人,博士研究生,主要从事计算机视觉、视频图像处理技术、嵌入式系统设计等方面的研究。E-mail:cbxiang@126.com
宋建中(1942—),男,研究员,博士生导师,主要从事模式识别与电视跟踪技术方面的研究。

●下期预告

大速比摩擦传动中的柔轮精度分析

张景和,张顺国,龙剑,何远新

(哈尔滨工业大学 机电工程学院 精密工程研究所,黑龙江 哈尔滨 150001)

为了消除柔轮加载时的扭曲变形误差,提高传动系统精度,对实柔轮变形量的模拟方法进行了研究。所讨论的大速比摩擦传动负载力矩为 2 Nm,传动比为 $I=200$,分别用弹性力学理论和有限元分析方法对柔轮建模,并用 ANSYS 软件分析柔轮在不同载荷作用下的变形量,对柔轮变形量进行实际测量。在微小变形情况下,用 0.2" 电子水平仪测量了柔轮在不同载荷作用下的角度变形量,给出了一组柔轮理论变形量与实际变形量的对比数据,当加载载荷为 0.88 Nm 时,理论计算值和实测平均值的差值为 1.85"。结果表明:试验结果与弹性力学建模和有限元建模的两种分析结果基本一致。