

采用逐层迭代方法的基础矩阵估计

杨磊^{1,2}, 李桂菊^{1*}

(1. 中国科学院 长春光学精密机械与物理研究所, 长春 130033; 2. 中国科学院大学, 北京 100049)

(* 通信作者电子邮箱 nemoleiyang@gmail.com)

摘要:为解决未知环境下运动序列中的基础矩阵估计问题,提出了一种逐层迭代优化的方法。该方法基于最优鲁棒估计方法,加入运动连续性以及多尺度对应的约束条件以减少虚假对应;然后,逐层将高层模型的数据内点添加到下层数据集,以更新数据集并同时估计单应性模型;最终,在底层全局优化并修正模型。实验表明,该方法的几何变换误差的均值不大于 2.891 821 pixel,误差波动范围的方差小于 0.295 172 pixel,相对于传统方法,当运动序列中场景表面的深度层次较多,深度变化连续时,误差均值及波动方差均有一定程度的降低。

关键词:基础矩阵;运动序列;逐层迭代;数据内点

中图分类号:TP394.1;TH691.9 **文献标志码:**A

Estimation of fundamental matrix by using layer-by-layer iterative approach

YANG Lei^{1,2}, LI Guiju^{1*}

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

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Concerning the estimation of fundamental matrix in motion sequences under the unknown environment, a layer-by-layer iterative approach was proposed. This approach was based on optimal robust estimation method, and added the constrained condition of motion continuity and multi-scale correspondence to reduce false correspondence. Then data inliers of model from high level were added into the data set of low level in order to update the data set and estimate the homographic model simultaneously. Finally, global optimization was done on bottom level to rectify the model. The experiments show that geometric transform error is no more than 2.891 821 pixel and variance of error fluctuated range is no more than 0.295 172 pixel, both the mean and the fluctuated variance of error are reduced in some degree when the scene surface has relatively plentiful depth levels and good continuity from motion sequences.

Key words: fundamental matrix; motion sequence; layer-by-layer iteration; data inlier

0 引言

透视投影使得真实世界中 3D 场景信息退化为 2D 图像信息,因此从 2D 图像中恢复 3D 场景成为视觉计算的重要任务;其中,场景目标和观察者之间的相对姿态估计是一个必要环节,其关键估计摄像机基础矩阵,从应用背景上可分为合作目标估计^[1]与非合作目标估计^[2]。非合作目标的几何特性未知,难以准确并稳定估计基础矩阵;对此,研究者们提出基于概率优化的鲁棒估计方法^[3],该方法是其他最优化基础矩阵估计方法的基础^[4]。

实际应用中,输入为未知场景运动序列时,连续运动导致视点不断变换,数据中将产生大量外点且噪声源的统计特性也难以预测,直接导致基本鲁棒估计结果出现严重畸变,造成图像严重扭曲。常采用最优化方法对基本鲁棒估计的结果进行全局优化,以克服由少数外点引发的误估。然而,直接采用全局优化容易导致模型的局部过拟合^[5],以致畸变加剧,采用合理的约束方法全局最优化基础矩阵成为改进的重点。

对此,可加入窄基线多尺度约束以及运动连续性约束,使用多尺度逐层迭代的方法,以高尺度层的整体视点变换约束低尺度层的局部细节,防止下局部过拟合,减小重投影误差。

1 基础矩阵估计的基本方法

初始模型计算多采用直接线性变换(Direct Linear Transform, DLT)^[3-4],主要由于 DLT 可直接推广到 n 点形式。令 $F = [f]_{3 \times 3}$ 为基础矩阵, $x = (x, y, 1)$ 和 $x' = (x', y', 1)$ 分别表示两帧中对应点的齐次坐标形式,则对应点应满足对极约束:

$$(x')^T F x = 0 \quad (1)$$

DLT 通过超定方程组求线性最小二乘解,故而无法应对数据外点,常以随机抽样一致(Random Sample Consensus, RANSAC)为代表的概率鲁棒估计方法去除外点,对 DLT 的解以 Sampson 距离^[6]为判别依据进行概率优化,最小化模型拟合误差。

由于 RANSAC 必须从小样本上通过模型一致性扩充内点集,对数据中的高斯噪声缺乏必要的抵抗能力,得出病态解。可通过全局优化鲁棒估计,即:给定 3D 点的齐次坐标 X_j 以及在视角 i 上观测像素平面上的 2D 点齐次坐标 x_{ij} ,存在视角 i 上的投影矩阵 P_i :

$$(P_{opt}, X_{opt}) = \arg \min_{P_i, X_j} \left(\sum_{i=1}^m \sum_{j=1}^n \|x_{ij} - P_i X_j\|^2 \right) \quad (2)$$

光束平差法(Bundle Adjustment, BA)被证明解决该类优

收稿日期:2013-03-26;修回日期:2013-05-14。 基金项目:国家自然科学基金资助项目(61172111)。

作者简介:杨磊(1986-),男,湖北武汉人,博士研究生,CFF 会员,主要研究方向:数字图像处理、计算机图形学; 李桂菊(1964-),女,研究员,主要研究方向:数字图像处理、数字信号处理。

化问题的成功方法^[7],该方法通过优化式(2),可求得基本鲁棒估计的数据内点上所存在的最优模型。

2 逐层迭代下的优化估计

2.1 运动序列的特征对应

对于运动场景,输入为动态场景图像序列。当运动序列邻近帧之间不存在运动突变时,建立邻近帧对应关系的过程可看作窄基线匹配问题^[8],建立对应关系:

$$\begin{aligned} (r_{u_{\text{opt}}}, r_{v_{\text{opt}}}) = \\ \arg \min_{(r_u, r_v)} \| \text{des}^{F_{i,j}}(u, v) - \text{des}^{F_{i+n,j}}(u + r_u, v + r_v) \| \quad (3) \end{aligned}$$

其中: F_r 表示序列中各帧,下标 i 与 $i+n$ 分别表示第 i 帧与第 $i+n$ 帧,为保持窄基线特性,常令 $n < 3$;下标 j 表示特征点所处的尺度层; des 表示特征点描述器; (u, v) 表示该特征点的像素坐标; r_u 与 r_v 分别表示特征点在第 $i+n$ 帧上相对于第 i 帧上的水平和垂直运动范围。

式(3)中包含了两个额外约束:1)尺度约束 $F_{i,j} \rightarrow F_{i+n,j}$: 只有处于同尺度层的特征才能进行匹配,可减少误匹配。2)运动约束 $\text{des}^{F_{i,j}}(u, v) \rightarrow \text{des}^{F_{i+n,j}}(u + r_u, v + r_v)$: 超过运动孔径 (r_u, r_v) 的特征将作为外点剔除,以优化整体运动的估计^[6]。需指出:由于校正过程由顶层至底层,顶层特征数量不少于底层的 20%,使初始特征数可反映大尺度变化模型。

2.2 逐层迭代

基于 2.1 节提出的特征对应方法,本文以最小化单应性变换引起的几何误差为原则,通过指定距离函数逐层优化基础矩阵。设 \mathbf{x} 和 \mathbf{x}' 为两帧中对应点, \mathbf{F} 和 \mathbf{F}^+ 分别为基础矩阵及其伪逆,则图像单应性几何变换误差^[6]可由齐次形式坐标表示为:

$$\begin{cases} E(\mathbf{F}) = (\| \mathbf{x} - \mathbf{F}^+ \mathbf{x}' / s \| + \| \mathbf{x}' - \mathbf{F} \mathbf{x} / s' \|) \\ \mathbf{F} \mathbf{x} = (\mathbf{x}', y', s') \\ \mathbf{F}^+ \mathbf{x}' = (x, y, s) \end{cases} \quad (4)$$

其中 \mathbf{F} 应为一个 2 阶奇异矩阵,只存在广义逆,但实际求解 \mathbf{F} 常得到满秩矩阵,因此需对 \mathbf{F} 进行奇异值分解,即:

$$\begin{cases} \mathbf{F} = \mathbf{U} \mathbf{A} \mathbf{V}^T \\ \hat{\mathbf{F}} = \mathbf{U} \text{diag}(\sigma_1^2, \sigma_2^2, 0) \mathbf{V}^T \end{cases} \quad (5)$$

其中 $\mathbf{A} = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ 。

非线性最小二乘优化的实现多基于列文伯格-马夸尔特 (Levenberg-Marquardt, LM) 法,对初值并不十分敏感^[9],且采用了稀疏存储的方法,极大地削减了内存占用^[10]。仍需指出:当初值较差时 LM 也难以在少量的迭代次数下收敛到全局最优解,并且算法稳定性下降,而较好的初值可使 LM 在较少的迭代次数下迅速收敛,提高算法稳定性。

运动序列中邻近帧之间具有稳定运动模式时,目标点可被连续观察;可作出假设:在第 i 帧中观察到的场景点 \mathbf{X} 在邻近的第 $i+n$ 帧中也能被观察到,即特征点估计 \mathbf{F} 只在预测范围 $\{r_u, r_v\}$ 内进行;通过鲁棒方法初始化估计顶层的 $\mathbf{F}_{i+n,j-1}$ 及顶层的内点,对顶层内点按尺度插值到下层,合并下层特征样本以估计下层的 $\mathbf{F}_{i+n,j}$,并在 $\mathbf{F}_{i+n,j}$ 上使用 BA 得到最佳估计 $\mathbf{F}_{i+n,\text{opt}}$ 。最佳估计基于多尺度估计,对于高层整体视点变换和低层局部相似变换有较好体现。令第 i 帧点为 \mathbf{x}' ,第 $i+n$ 帧点为 \mathbf{x} ,则有:

$$\begin{cases} (\mathbf{x}'_{j-1})^T \mathbf{F}_{j-1} \mathbf{x}_{j-1} = 0 \\ (\mathbf{x}'_j + \sigma_{j,j-1} \mathbf{x}'_{j-1})^T \mathbf{F}_j (\mathbf{x}_j + \sigma_{j,j-1} \mathbf{x}_{j-1}) = 0 \\ \mathbf{F}_{\text{opt}} = \arg \min_{\mathbf{F}_j} \| E(\mathbf{F}_j) \| \end{cases} \quad (6)$$

其中: j 表示所处的尺度层; \mathbf{F}_j 表示在 j 层估计的基础矩阵; \mathbf{x}'_{j-1} 表示 $j-1$ 层中满足 $(\mathbf{x}'_{j-1})^T \mathbf{F}_{j-1} \mathbf{x}_{j-1} = 0$ 的内点; $\sigma_{j,j-1}$ 表示第 j 层与其上一层尺度所相差的倍数; $\mathbf{F}_{\text{opt}} = \arg \min_{\mathbf{F}_j} \| E(\mathbf{F}_j) \|$ 则表示 BA 对式(6)优化的结果。

3 实验及其结果分析

3.1 实验设计

实验中测试数据为 Technical University of Denmark (DTU) Robot data set, 该数据库由运动序列图像构成。本文选取数据库中序列 1, 4, 31, 32, 50 共计 5 个大型目标,此类场景具有视角变化明显、纹理及空间结构复杂且深度层次丰富、更符合实际应用场景的特点;而其余场景中多为小型目标,其纹理单一或结构简单,因此各方法的基础矩阵估计结果相近。每个测试序列均由连续的 30 帧组成,各场景表面的深度层次及深度变化幅度均不同。

实验对比的算法为基本鲁棒估计(基于 RANSAC)、最优鲁棒估计(基于 RANSAC + LM)和逐层迭代估计。评价标准采用单应性几何变换的重投影误差^[6]。实验中,各算法均在各场景序列的连续 30 帧中的相邻两帧上连续估计基础矩阵,并对每种算法在各序列上均给出 29 次单应性变换误差评价,算法的准确性以及稳定性均可直观地通过误差折线评价。

3.2 实验结果分析

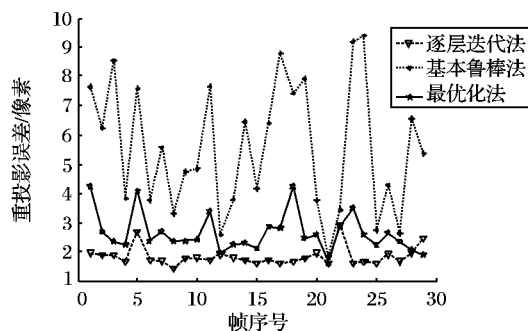
表 1 统计了三种参测算法在 5 个测试场景序列的最大、平均单应性误差以及误差波动方差。图 1~3 为 5 个测试序列的重投影误差折线图。

表 1 五个不同场景的实验结果

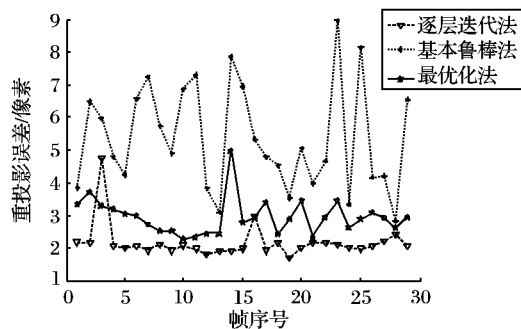
算法	比较项	场景 1	场景 4	场景 31	场景 32	场景 50
基本鲁棒法	最大误差	9.396 707	9.870 246	5.854 354	8.950 960	8.478 334
	平均误差	5.537 033	6.893 329	3.944 010	5.364 465	5.304 125
	波动方差	4.915 527	2.458 211	1.152 768	2.681 856	3.269 923
最优化法	最大误差	4.232 251	5.214 238	2.786 858	4.969 806	5.724 934
	平均误差	2.645 257	3.841 698	2.172 924	2.945 580	3.066 538
	波动方差	0.433 499	0.332 552	0.088 204	0.298 430	1.158 660
逐层迭代法	最大误差	2.916 497	4.205 488	4.077 125	4.740 515	2.344 591
	平均误差	1.854 720	2.891 821	2.229 976	2.160 012	1.806 971
	波动方差	0.104 511	0.146 483	0.154 200	0.295 172	0.031 760

场景 1 与场景 32 较为相似,深度层次较少,场景表面连续性极好。场景 1 中逐层估计的平均单应性误差是最优化估计的 70.11%,误差震荡范围是最优化估计的 24.11%。场景 32 中逐层估计的平均误差是最优化估计的 73.33%,误差震荡范围是最优化估计的 98.91%。两场景的误差震荡范围相差较大是由于场景 1 变化幅度小而场景 32 变化幅度较大所致。

场景 31 较为特殊,深度层次较少,变化幅度较大,场景表面连续性较差;逐层估计的平均误差是最优化估计的 102.63%,误差震荡范围是最优化估计的 174.82%;该场景中未剔除的上层外点误差传入下层,导致下层得出病态解,造成迭代估计得出非最优模型。



(a) 场景1



(b) 场景32

图1 场景1和场景32中三种算法的误差曲线

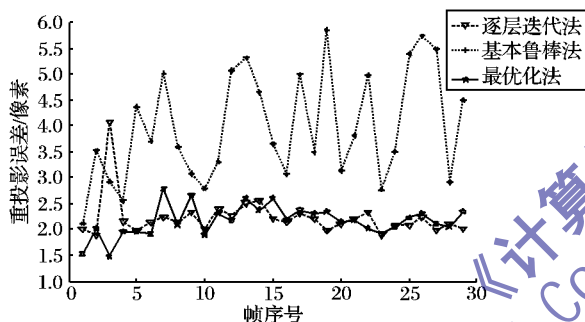
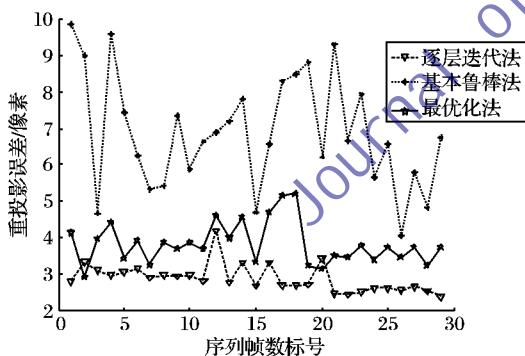
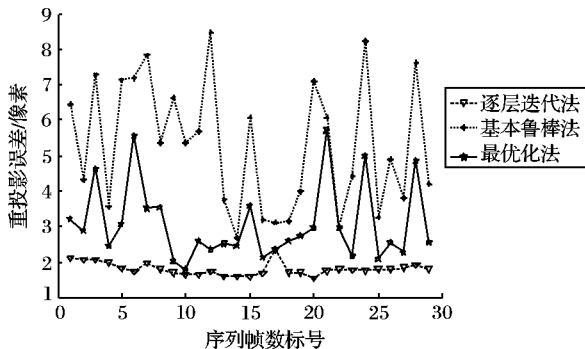


图2 场景31中三种算法的误差曲线



(a) 场景4



(b) 场景50

图3 场景4和场景50中三种算法的单应性误差曲线

场景4与场景50较为相似,深度层次较多,变化幅度较

大,场景表面连续性较好。场景4中逐层估计的平均误差是最优化估计的75.27%,误差震荡范围是最优化估计的44.04%;个别帧上,最优化估计可获得比逐层迭代估计更高的精度。场景50中逐层估计的平均误差是最优化估计的58.93%,误差波动范围是最优化估计的27.41%;该场景下最优化估计性能较差,主要原因是该场景表面深度层次较多且存在明显局部深度变化,导致局部过拟合。

4 结语

当输入为未知场景的运动序列时,逐层迭代估计方法可通过高层点集估计整体视点变化,将高层模型的点带入下层进行逐层优化估计。实验表明,与主流的方法相比,当上层数据外点被有效剔除时,可有效防止下层模型的局部过拟合;当场景深度层次性较多且表面连续性较好时,逐层迭代与常见的最优化估计的平均单应性几何变换误差比值为58.93%到75.27%,误差的波动范围比值为27.41%到98.91%,提高了准确性和稳定性。

参考文献:

- [1] 彭凯. 基于光靶标的双相机三坐标测量系统关键技术的研究[D]. 天津: 天津大学, 2007.
- [2] SNAVELY N, STEVEN M, SZELISKI S. Modeling the world from Internet photo collections[J]. International Journal of Computer Vision, 2008, 80(2): 189-210.
- [3] FAUGERAS O, LUONG Q T, MAYBANK S. Camera self-calibration: theory and experiments[C]// ECCV 1992: Proceedings of the Second European Conference on Computer Vision, LNCS 588. Berlin: Springer, 1992: 321-334.
- [4] HARTELY R, ZISSERMAN A. Multiple view geometry in computer vision, second edition[M]. 2nd ed. Cambridge: University of Cambridge Press, 2004: 87-150.
- [5] NI K, STEEDLY D, DELLAERT F. Out-of-core bundle adjustment for large-scale 3D reconstruction[C]// ICCV 2007: Proceedings of the 11th IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2007: 1-8.
- [6] CHOI S, KIM T, YU W. Performance evaluation of ransac family[J]. Journal of Computer Vision, 1997, 24(3): 271-300.
- [7] AGARWAL S, SNAVELY N, SIMON I, et al. Building rome in a day[C]// Proceedings of the 12th IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2009: 72-79.
- [8] TOMASI C, KANADE T. Detection and tracking of point features[EB/OL]. [2012-10-10]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.45.5770>.
- [9] FAN J. A modified Levenberg-Marquardt algorithm for singular system of nonlinear equations[J]. Journal of Computational Mathematics, 2003, 21(5): 625-636.
- [10] LOURAKIS M, ARGYROS A. SBA: a software package for generic sparse bundle adjustment[J]. ACM Transactions on Mathematical Software, 2009, 36(1): 1-30.
- [11] AANAS H, DAHL A L, PEDERSEN K S. Interesting interest points[J]. International Journal of Computer Vision, 2012, 97(1): 18-35.