

文章编号: 1000-5862(2012)05-0466-06

基于李群代数表达帧间位姿变化矩阵的 3D 视觉跟踪研究

张 磊, 张兴国

(南通大学机械工程学院, 江苏 南通 226019)

摘要: 基于 RAPiD 边缘跟踪方法的思想, 采用以帧间位姿变化更新目标物绝对位姿值的办法, 引入李群代数表达帧间位姿变化矩阵; 然后将帧间位姿变化矩阵参数的求解问题转换为线性最小二乘的最优求解问题. 在李群代数的表达方式下, 最小二乘的最优问题可较容易获得线性化表达, 方便了求解. 实验显示了该算法的 3D 跟踪效果与目标物位姿值的估计结果.

关键词: 视觉跟踪; 3 维; 最小二乘问题; 位姿估计

中图分类号: TP 391.41 **文献标志码:** A

0 引言

3D 视觉跟踪是指计算机视觉中通过图像获得目标物的实际 3 维位姿(位置与空间姿态), 比起仅仅在图像上获得跟踪性能的 2D 跟踪要求高得多, 因此目标对象常限制为低速或静止. 3D 跟踪的需求主要集中在视觉机器人操作 3 维目标物、扩展的虚拟现实, 以及有较高要求的智能交通监控等方面. 欲获得目标物的 3 维空间位姿, 较自然的想法是通过双目确认包括深度信息在内的 3 维信息. 事实上在双目或多目视觉的研究中, 由于要面临不同视角下的视差匹配、多摄像机标定等难题, 且对匹配、标定误差敏感, 实施并不方便. 而在单目视觉中, 一个点的 3 维信息虽然不易获得, 但对一个有大小的物体, 可以利用多个点的约束关系, 间接获得 3 维信息. 人们在研究中发现, 单目 3D 跟踪如果有足够的先验知识, 实现将变得简单易行.

自 20 世纪 80 年代以来, 有关研究人员^[1-3]提出基于目标物 3 维 CAD 模型的 3D 跟踪研究方式更能应对实际的视频图像, 性能更鲁棒. David G Lowe 和 C. Harris 在这方面较早地开展了有关研究工作^[2,4-5], 提出了经典的 RAPiD 方法, 如图 1. 该方法基于模型投影与图像边缘特征之间的差异最小为目标函数,

进行位姿矩阵或投影矩阵估计. 因其可避免特征匹配等困难问题且易满足实时性能, 至今仍然是这类研究的基本方法之一. 本文的研究对 RAPiD 方法进行了有关改进.

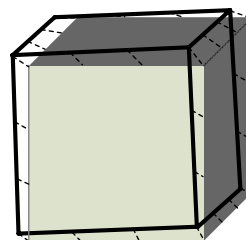


图 1 RAPiD 方法示意图

1 3 维位姿矩阵的参数化表达问题分析

3D 视觉跟踪的研究中需要面临一个基本问题: 目标物 3 维位姿的参数化表达. 在经典摄像机针孔模型下, 空间某点与其在图像上投影的像素坐标的关系记为: $m = KEC_W / s$, 其中 K 为摄像机内参数矩阵, E 为需要求解的位姿矩阵(pose), 也是摄像机针孔模型的外参数矩阵, 常写成 $\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$ 的形式, 其中 R 为 3×3 旋转矩阵, 代表目标物相对相机的姿态; t 为位移向量, 代表目标物相对于相机的位置; C_W

收稿日期: 2012-04-08

基金项目: 江苏省高校自然科学基金(11KJB460007)和南通市应用研究计划(K2010001)资助项目

作者简介: 张 磊(1973-), 男, 江苏南通人, 讲师, 博士, 主要从事机器人与机器视觉的研究.

为可事先指定的空间点世界坐标; s 表示比例因子, 与目标物在摄像机坐标系中的深度有关, 非独立的参数, 可以并入位姿矩阵的未知参数中一并求解, 故以下表达将其隐去, $P=KE$ 为投影矩阵.

摄像机针孔线性模型中旋转矩阵 R 的参数化表达是影响位姿矩阵 E 参数求解的关键点, 其经典的欧拉角表达为

$$R = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma & -\sin \gamma \\ 0 & \sin \gamma & \cos \gamma \end{bmatrix}, \quad (1)$$

该形式虽然直观, 方便根据目标物空间某点的姿态(即欧拉角)正向求解其投影坐标, 然而却很不便于由投影点坐标求逆解 R , 何况还存在万向节死锁的可能而引入病态解.

关于旋转矩阵的另一种常见表达为指数映射的形式: 设 $\omega = [\omega_x, \omega_y, \omega_z]^T$, 绕 ω 旋转 θ 可以表达为

$$R(\Omega) = \exp(\Omega) = I + \Omega + \frac{1}{2!}\Omega^2 + \frac{1}{3!}\Omega^3 + \dots \text{ 其中}$$

$$\Omega = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}.$$

这种指数映射表达避免了万向节死锁, 然而无穷级数的形式仍然不利于求逆解 R . 但如果 θ 旋转角为微小变化, 则前述指数映射可以简化为 $I + \Omega$.

文献[6-7]采取单应矩阵表达空间平面目标的空间坐标与图像坐标之间的关系, 并以帧间单应矩阵的形式求解: $H_w^t = H_{t-1}^t H_{t-2}^{t-1} \dots H_0^1 H_w^0$, 其中 H_w^0 为初始单应矩阵估计值, H_{t-1}^t 为帧间单应矩阵. 因帧间单应矩阵为微小旋转变化, 故而指数映射可以用 $I + \Omega$ 的形式近似, 方便了解. 然而单应矩阵将摄像机内参数一并包括在内, 且要求跟踪的目标为平面或显著的含有平面(单应矩阵丢掉了未知参数量), 并不利于求解一般的目标物相对于摄像机的位姿.

2 基于李群代数表达帧间位姿变化矩阵

本文借鉴文献[8-10]的做法引入李群代数的表

达方法, 提出以帧间位姿变化更新位姿值, 同时根据基于边缘进行跟踪的特点, 结合帧间边缘距离的表达, 转换为线性最小二乘问题, 得到目标物位姿值.

首先, 将投影矩阵以动态的形式表达为 $P = KEM$, 其中 M 为帧间位姿变化矩阵, M 也为 $\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$ 的表达形式, E 为初值位姿或前一帧位姿.

以帧间更新的形式将当前帧与前一帧的位姿关系表达为 $E_{t+1} = E_t M_{t+1}^t$, 此处 t 代表图像帧顺序号, 投影矩阵 $P = KE_0 M_1^0 \dots M_{t+1}^t$. 如果帧与帧之间未发生变化, 理应 $M_{t+1}^t = I$, I 表示单位矩阵. 实际的情况, 即便目标物相对摄像机未发生运动变化, 然而如果前一帧估计的位姿不够准确, 也可以通过求解 M_{t+1}^t 来修正 E_{t+1} , 此时的 M_{t+1}^t 不等于 I .

其次, 引入李群代数表达形式, 将 M_{t+1}^t 基于群 $SE(3)$ 的 4 阶方阵形式, 以 6 维的李群来表达. $SE(3)$ 确切地表达了系统输出的姿态空间, 李代数则是同一点处的切空间(指某曲面上的一点的切平面内, 所有切向量组成的空间), 从而成为表达微分量(速度、小位移)的一个自然空间. 构建代表 x, y, z 轴移动与旋转的一组基如下:

$$\begin{aligned} G_1 &= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, G_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \\ G_3 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, G_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \\ G_5 &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, G_6 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (2)$$

M_{t+1}^t 视为群中的元素, 可由这些基通过指数映射来获得, 即 $M_{t+1}^t = \exp(\sum_i \alpha_i G_i)$, 此处 i 表示每个基及其因子的编号. 考虑到 M_{t+1}^t 表示的是相邻帧下的摄像机相对目标物的姿态变化, 这种变化很微小, 可以用一阶线性表达的方式, 即

$$M_{t+1}^i \approx I + \sum_i \alpha_i G_i. \quad (3)$$

求解帧间位姿变化参数的问题转化求解未知参数向量 $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)^T$. 这种表达简洁高效, 方便优化算法的求解, 基本上可以避免万向节死锁现象.

3 帧间位姿变化位姿参数量的最小二乘优化求解

设第 i 控制点的前一帧图像像素坐标为 m_t , 根据针孔线性投影模型有: $m_t = KE_t C_{Wi}$, 该点在当前帧的 $m_{t+1} = KE_t M_{t+1} C_{Wi}$, 如图 2 所示. 在经典的 RAPID 方法中, 距离 $l_i = n_i^T (m_{t+1} - m_t)$, 结合公式(3)可得 $l_i = n_i^T KE_t (\sum_{r=1}^6 \alpha_r G_r) C_{Wi}$. 对于 N 个控制点, 该线性变换简记为 $A\alpha = l$, 其中距离向量 l 是由 N 个 l_i 构成的列向量. 于是问题转化为 $A\alpha - l = 0$ 的线性最小二乘求解问题, 有关目标函数 $\min F(\alpha) = \min \sum_{i=1}^N f_i^2(\alpha) = \min (A\alpha - l)^T (A\alpha - l)$, 其中 $A\alpha - l = (f_1(\alpha), f_2(\alpha), \dots, f_N(\alpha))^T$, 可通过广义逆求解得

$$\alpha = (A^T A)^{-1} A^T l. \quad (4)$$

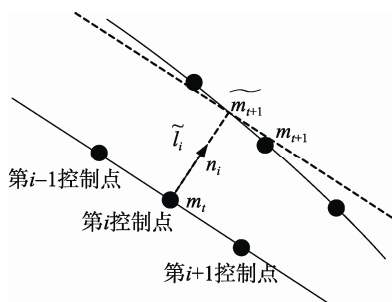


图 2 寻找新的控制点位置

然而在基于边缘的跟踪方式中, 控制点在下一帧的准确对应点并不容易确定, 一般采用沿前一帧(第 t 帧)图像的边缘垂直方向搜索当前帧(第 $t+1$ 帧)的边缘点, 以如图 2 的 \widetilde{m}_{t+1} 替代 m_{t+1} 点, 造成距离向量 l 并不准确. 本文拟采用迭代方式, 更新 l , 在优化搜索的方式下, 使求解的 α 满足 m_t 点向 m_{t+1} 点

更准确的逼近. 根据公式(4) 迭代公式为

$$\alpha^{(k+1)} = \alpha^{(k)} + (A_k^T A_k)^{-1} A_k^T l^{(k)}, \quad (5)$$

其中 $\alpha^{(k)}$ 为第 k 次的迭代值, 初始值 $\alpha^{(0)}$ 为零向量, A_k 为第 k 次的 $\alpha^{(k)}$ 值下所更新的系数矩阵, $l^{(k)}$ 为第 k 次更新的距离向量.

为了加快收敛速度, 引入了经典 Levenberg-Marquadt 动步长的方法^[11-12], 公式(5)化为

$$\alpha^{(k+1)} = \alpha^{(k)} + (A_k^T A_k + \lambda I_k)^{-1} A_k^T l^{(k)}. \quad (6)$$

L-M 算法中步长 λ 的计算, 可参见文献[11-12], 本文不展开. 对每一帧图像, 迭代过程在 $\|\alpha^{(k+1)} - \alpha^{(k)}\| \leq \varepsilon$ 时终止, 但图像在不断的变化中, 对于下帧, 迭代过程将重新开始, 前一帧的估计值作为下一帧的初始值.

4 3D 视觉跟踪的鲁棒性措施

对实际图像的目标进行位姿估计, 存在大量的干扰点或者伪数据, 必须采取鲁棒性的对策. 本文引入 M 估计器, 目标函数 $F(x)$ 不采用 $\sum_i f_i^2(\alpha)$ 的平方和形式. 因为噪声的 $f_i(\alpha)$ 一般较大, 单纯的平方和使得噪声对整个值影响较大, 从而使模型易收敛于错误的地方. 设 $f_i(\alpha) = r_i$, 根据 M 估计器的思想, 目标函数定义为 $\sum_i \rho(r_i)$, 通过函数 ρ 改变干扰点的影响力, 其中 ρ 函数取 Tukey 函数的形式为

$$\rho_{Tuk}(x) = \begin{cases} c^2 [1 - (1 - (r_i/c)^2)^3] / 6, & |r_i| \leq c, \\ c^2 / 6, & \text{其它}. \end{cases} \quad (7)$$

该估计器对 r_i 明显增大后不敏感, 从而增加了对噪声的鲁棒性. 使用 M 估计器后, 迭代公式(6)改进为

$$\alpha^{(k+1)} = \alpha^{(k)} + (A_k^T W A_k + \lambda I_k)^{-1} A_k^T W l^{(k)}, \quad (8)$$

其中权值矩阵 W 可简单取为 $\text{diag}(\dots, w_i, \dots)$, $\rho(r_i) = w_i^2 r_i^2$, $w_i = \rho(r_i)^{1/2} / r_i$.

5 实验结果

选择一个并口接头作为目标物, 在相机镜头相对

目标物缓慢运动的情况下,对目标物相对摄像机的位姿进行了估计.目标物相对摄像机的位姿是基于图3所示的摄像机坐标系 xyz 描述的,世界坐标系 $x_wy_wz_w$ 建立在目标物上,世界坐标系的原点代表目标物的位置.

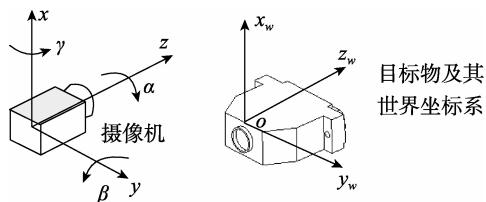


图3 目标物在摄像机坐标系中的位姿描述

图4~图6显示了目标物CAD框架模型在本文算法估计的目标物相对摄像机的位姿下的投影效果.为了节省版面,这些图是在原 640×480 分辨率图中截取了 420×340 的大小,并且以一半显示.图4显示了在某初始值下,摄像机相对目标物静止的瞬间,本文算法(基于公式(8))的迭代过程.图5显示了摄像机相对目标物变换到另一姿态后,在静止的瞬间,本文算法在前一位姿值下的迭代过程.图6是在摄像机相对目标物作缓慢运动的图像序列中,本文算法的连续

跟踪效果(图6选择了几百幅序列图中的3幅).

图4~图6虽然直观,但不能定量表达本文算法的跟踪性能.图7显示了类似图4~图6状态下图像序列中典型帧下本算法的位姿估计值(因在线的实验并不方便得到每帧图像的目标物位姿参考准确值,这里在线选择了有代表的典型帧计算结果,同时将这些典型帧另存后,以离线方式手工标定出相应的参考准确值),包括 x, y, z 的3个方向的位移;而姿态矩阵 R 的表达,由帧间位姿变化,经过 $E_{t+1} = E_t M_{t+1}^t$ 计算,最终以欧拉角的回转角 α ,俯仰角 β 和偏转角 γ 表达,如图3所示,其更为直观.从结果可知,位姿估计与参考值总体上误差不大,偶尔有估计误差较大的情况,主要原因是在背景干扰或噪声下,边缘检测将产生伪边缘,算法偶尔会收敛于这些伪边缘而影响了准确性.

实验发现,在本文所示图像的背景下,算法很少出现估计误差大的情况,但当目标物处于相对混乱的背景或噪声较多的环境时,因大量伪边缘的存在,算法估计的准确性与稳定性有所下降,这将是算法今后需要改善的地方.

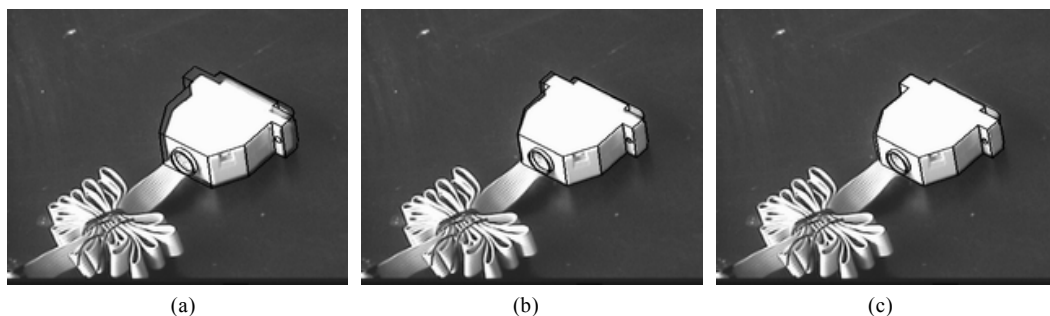


图4 摄像机相对目标物静止的瞬间,基于某初始值的跟踪算法迭代过程图像效果显示

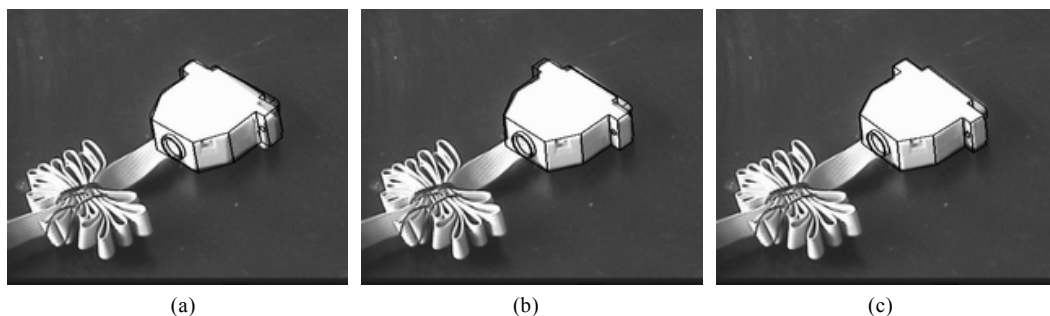


图5 摄像机相对目标物静止的另一瞬间,基于前一位姿值的跟踪算法迭代过程图像效果显示

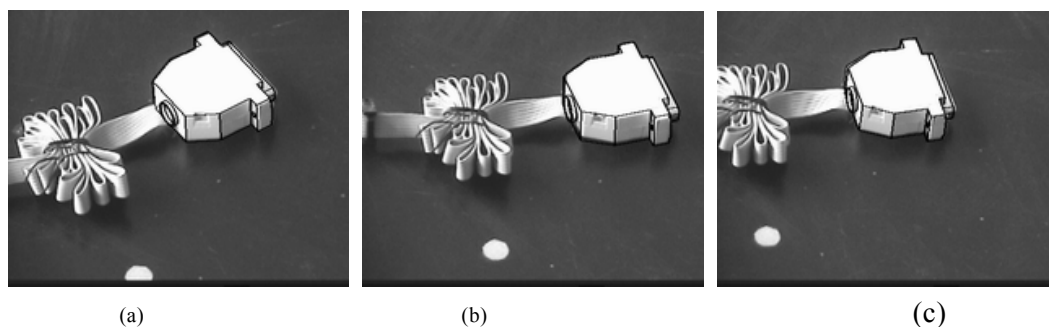


图 6 摄像机相对目标物运动时, 跟踪的图像效果显示

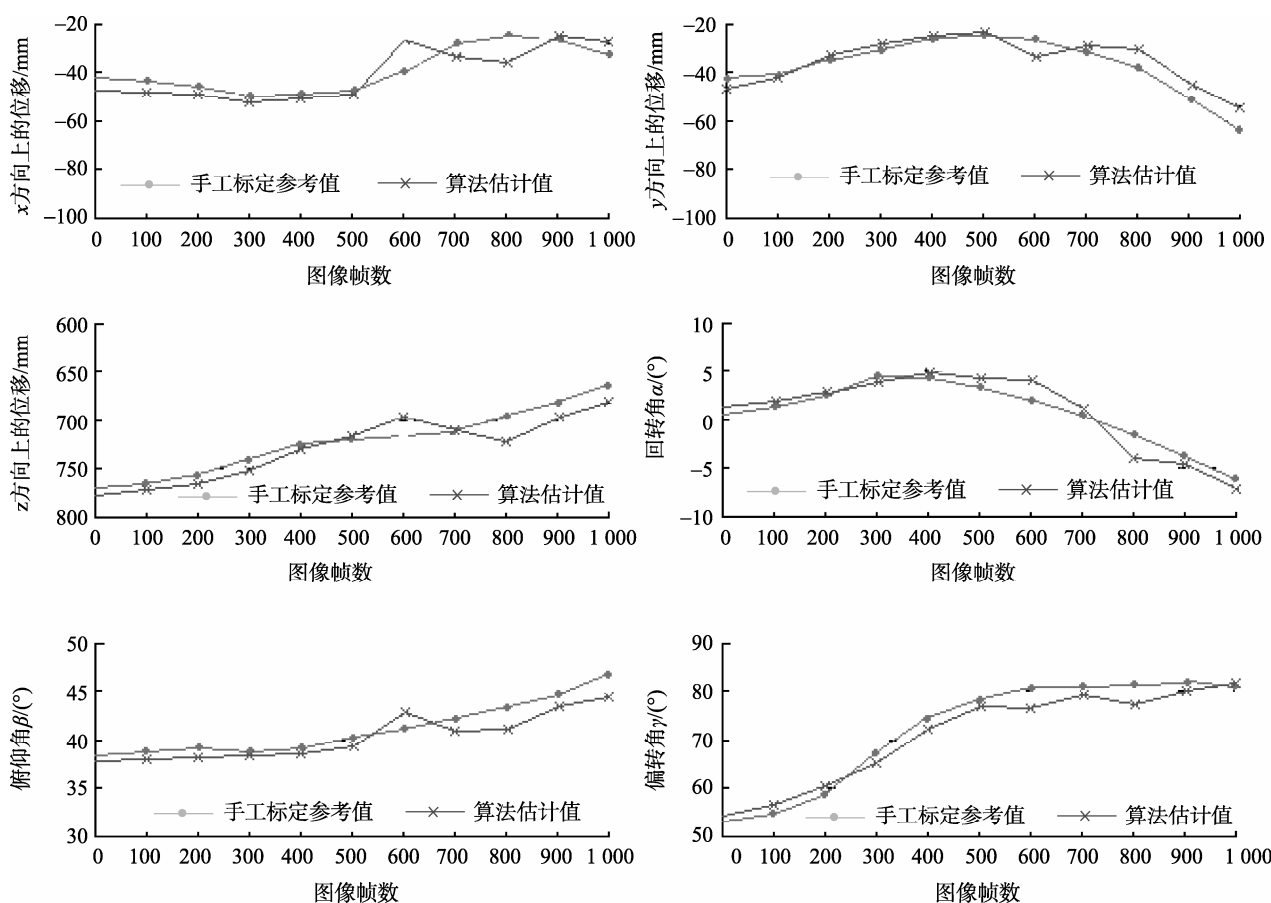


图 7 本文算法目标物跟踪的估计值与参考准确值

6 参考文献

- [1] Gennery D B. Tracking known three-dimensional objects [EB/OL]. [2011-11-18]. <http://www.aaai.org/Papers/AAAI/1982/AAAI82-004.pdf>.
- [2] Lowe D G. Fitting parameterized three-dimensional models to images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 3(5): 441-450.
- [3] Lepetit V, Fua P. Monocular model-based 3d tracking of rigid objects: A survey [EB/OL]. [2012-01-09]. http://cvlab.epfl.ch/~lepetit/papers/lepetit_ftcg05.pdf.
- [4] Lowe D G. Robust model-based motion tracking through the integration of search and estimation [J]. International Journal of Computer Vision, 1992, 8(2): 113-122.
- [5] Harris C. Tracking with rigid objects [M]. Massachusetts: MIT Press, 1992.
- [6] Simon R, Fitzgibbon A, Zisserman A. Markerless tracking using planar structures in the scene [EB/OL]. [2012-02-16].

- <http://www.robots.ox.ac.uk/~vgg/publications/papers/simon00.pdf>.
- [7] Simon G, Berger M O. Pose estimation from planar structures [J]. Computer Graphics and Applications, 2002, 22: 46 - 53.
- [8] Changhyun Choi, Henrik I. Christensen robust 3D visual tracking using particle filtering on the SE(3) group [EB/OL]. [2012-03- 12]. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05980245>.
- [9] Drummond T, Cipolla R. Real-time visual tracking of complex structures [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 27: 932-946.
- [10] Changhyun C , Henrik I C. Real-time 3D model-based tracking using edge and keypoint features for robotic manipulation [EB/OL]. [2012-02-11]. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05980245>.
- [11] Levenberg K. A method for the solution of certain problems in least squares [J]. Quarterly of Applied Mathematics, 1944, 2: 164-168.
- [12] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters [J]. SIAM Journal on Applied Mathematics, 1963, 11: 431-441.

3D Visual Tracking Based on the SE(3) Group Representation of Interframe Pose Transform Matrix

ZHANG Lei, ZHANG Xing-guo

(School of the Mechanical Engineering, Nantong University, Nantong Jiangsu 226019, China)

Abstract: The SE(3) group is introduced for representing the interframe pose transform which can be used for updating the absolute pose of the object. Then, the 3D tracking problem can be equivalent to the least square optimization problem on parameter solving of the interframe pose transform. Owe to the SE(3) group representation, simpler linear representation can be used in optimization process, which pave the way for parameter solving. Experiments show the 3D tracking results of an object and the related pose estimation.

Key words: visual tracking; 3D; least square problem; pose estimation

(责任编辑: 冉小晓)