

文章编号: 1002-0446(2005)02-0137-05

一种基于点对的深度和运动估计方法*

钟志光, 易建强, 赵冬斌

(中国科学院自动化研究所复杂系统与智能科学实验室, 北京 100080)

摘要: 针对基于模型的姿态估计问题提出了一种新颖的深度和运动估计方法. 估计过程分为两步. 首先根据由 3 个模型点构成的 3 个点对估计旋转矩阵. 然后, 根据 3 个点对中的一个计算平移和深度. 本文还提出了一种简单而有效的优化算法用于提高估计结果的精度. 该估计方法的特点是独立地估计旋转矩阵, 而且只需少量的模型点就能获得好的估计结果. 模拟试验和实际试验结果表明了该方法的有效性.

关键词: 深度估计; 运动估计; 姿态估计; 外部相机参数标定

中图分类号: TP24 文献标识码: B

Depth and Motion Estimation from Point Pairs

ZHONG Zhi-guang, YI Jian-qiang, ZHAO Dong-bin

(Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: This paper presents a novel approach for depth and motion estimation, aiming at the model-based pose estimation problem. The estimation procedure consists of two steps. Firstly, it estimates the rotational matrix according to three point pairs constituted by three model points. Then, the translation vector and depths are computed from one of the three pairs. In order to enhance the precision of the estimated results, a simple and effective optimization algorithm is also proposed. The main advantages of this method are that it estimates the rotational matrix individually and can obtain good estimation results according to small number of model points. Experiments performed on synthetic data and real images demonstrate the effectiveness of this approach.

Keywords: depth estimation; motion estimation; pose estimation; extrinsic camera parameters calibration

1 引言 (Introduction)

深度和运动估计在移动机器人导航、工业检测等方面有着重要应用. 已有的方法可以大致分为三类: 基于光流的、基于梯度的和基于特征的. 其中, 基于光流的方法^[1]对光流计算误差是敏感的, 因为光流的估计是一个病态问题, 只有引入其它限制 (如平滑) 才能解决. 基于梯度的方法^[2], 也叫直接方法, 根据灰度图像的时空微分估计深度和运动参数. 这种方法一般基于亮度变化限制方程并以最小平方作为估计器, 因而对非高斯残差是敏感的^[2]. 基于特征的方法^[3,4]被广泛研究, 选取的特征可以是点、线、面等.

本文提出的方法是基于特征的, 目的是解决基于模型的姿态估计问题^[5-7]. 如 $O_f^{[6]}$ 所述, 这一问题

假设某个目标的三维结构已知且它的二维图像可以得到, 从而估计该目标的姿态. 许多研究者一直致力于该问题的研究. 例如, Oberkempf^[5]介绍了一种比例正交投影近似方法, 使用四个或更多的模型点. 该方法通常得到两个不同的姿态估计结果. $O_f^{[6]}$ 通过垂直投影计算模型点的近似深度, 并根据奇异值分解方法估计运动参数.

本文方法根据相机与模型点对之间的几何关系估计深度和运动参数. 首先, 根据由三个模型点构成的三个点对估计旋转矩阵. 然后, 根据这三个点对中的一个计算深度和平移向量. 最后, 通过本文提出的优化算法对这些估计结果进行优化. 值得一提的是, 大多数已有方法, 如 Oberkempf^[5]和 $O_f^{[6]}$ 的方法, 在

* 基金项目: 国家 973 计划资助项目 (2003CB517106).

收稿日期: 2004 - 07 - 17

进行非线性优化的同时估计旋转矩阵和平移向量,而且估计深度或运动时需要各种近似,但本文方法的特点是首先独立地估计旋转矩阵,然后直接计算出深度和平移向量.

一些研究者已经注意到使用特征点对的优势并利用它们解决各种问题.例如, Wang^[7]根据标定点对之间的相对位置信息标定外部相机参数. Newman^[8]利用特征点对重建相对地图用于移动机器人导航.本文方法主要受 Wang^[7]启发,但与 Wang的方法至少在以下方面是不同的.首先, Wang的方法是一种解析方法而本文方法是一种几何方法.我们认为本文的方法更加简洁和直观.其次, Wang的方法仅仅估计运动参数而不能明确地估计标定点(模型点)的深度.

2 估计旋转矩阵 (Estimation of rotation matrix)

假定某个目标上有 n 个特征点 $G_0, G_1, \dots, G_i, \dots, G_{n-1}$ 在相机的视域里.这些特征点被叫做模型点.目标坐标系由 G_0 表示, G_0 被叫做参考点.点 G_i 在目标坐标系中的坐标 $[G_{ix}, G_{iy}, G_{iz}]^T$ 是已知的. G_i 的成像被叫做图像点 g_i , 它们在相机坐标系 C 下的正规化图像坐标 $[g_{ix}, g_{iy}, g_{iz}]^T$ 是已知的. G_i 在相机坐标系下的坐标 $[X_i, Y_i, Z_i]^T$ 是未知的, 因为目标在相机坐标系下的姿态是未知的. 根据图 1, 我们将说明如何根据 3 个点对直接估计旋转矩阵而无需明确地解出 G_i 的坐标 $[X_i, Y_i, Z_i]^T$.

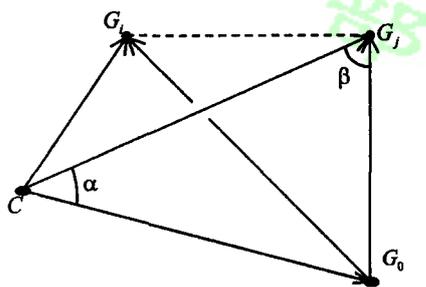


图 1 相机与点对的几何关系

Fig. 1 Geometric relationships between a camera and point pairs

首先, 选取 G_0 和 G_i 作为点对, 得到:

$$(\overline{CG_0} \times \overline{CG_i}) (R \overline{G_0 G_i}) = 0 \quad (1)$$

这里 R 是要估计的旋转矩阵. 表达式 $R \overline{G_0 G_i}$ 表示将向量 $\overline{G_0 G_i}$ 由 G_0 坐标系变换到 C 坐标系. R 一般由欧拉角表示. 假设 ω 是偏航角、 ϕ 是俯仰角、 θ 是翻滚角

以及 $c\phi = \cos\phi, s\phi = \sin\phi$ 等, 那么 R 可以写成:

$$R = R_z(\theta) R_y(\phi) R_x(\omega) = \begin{bmatrix} c\phi c\theta & s\omega s\phi c\theta - c\omega s\theta & c\omega s\phi c\theta + s\omega s\theta \\ c\phi s\theta & s\omega s\phi s\theta + c\omega c\theta & c\omega s\phi s\theta - s\omega c\theta \\ -s\phi & s\omega c\phi & c\omega c\phi \end{bmatrix} \quad (2)$$

如果旋转较小, 则 R 可以近似写成:

$$R = \begin{bmatrix} 1 & -w3 & w2 \\ w3 & 1 & -w1 \\ -w2 & w1 & 1 \end{bmatrix} \quad (3)$$

这里 $w = [w1, w2, w3]^T$ 是绕 C 的旋转向量而且它的三个旋转角较小. 为了简化后面的估计方程, 本文采用后一种表示方法. 由这种近似造成的估计误差将在第 5 节中加以分析. 在后面的讨论中, 由欧拉角表示的 R 被写成 R_i 以示区别.

这样, 方程 (1) 就可以写成:

$$\begin{bmatrix} g_{0x} \\ g_{0y} \\ 1 \end{bmatrix} \times \begin{bmatrix} g_{ix} \\ g_{iy} \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -w3 & w2 \\ w3 & 1 & -w1 \\ -w2 & w1 & 1 \end{bmatrix} \cdot \begin{bmatrix} G_{ix} - G_{0x} \\ G_{iy} - G_{0y} \\ G_{iz} - G_{0z} \end{bmatrix} = 0 \quad (4)$$

为了表达方便, 假设:

$$A_{0i} = \begin{bmatrix} g_{0x} \\ g_{0y} \\ 1 \end{bmatrix} \times \begin{bmatrix} g_{ix} \\ g_{iy} \\ 1 \end{bmatrix}$$

以及:

$$B_{0i} = \begin{bmatrix} G_{ix} - G_{0x} \\ G_{iy} - G_{0y} \\ G_{iz} - G_{0z} \end{bmatrix}$$

由方程 (4) 可得:

$$P_{0i} \begin{bmatrix} w1 \\ w2 \\ w3 \end{bmatrix} = Q_{0i} \quad (5)$$

这里:

$$P_{0i} = [A_3 B_2 - B_3 A_2, A_1 B_3 - A_3 B_1, A_2 B_1 - A_1 B_2]$$

$$Q_{0i} = - (A_1 B_1 + A_2 B_2 + A_3 B_3)$$

A_i 和 B_i ($i=1, 2, 3$) 分别是向量 A_{0i} 和 B_{0i} 的第 i 个元素. 根据前面的假设, 它们都是已知的.

类似地, 如果 G_0 和 G_j 或者 G_i 和 G_j 被选为点对, 可以得到与式 (5) 类似的方程. 因此, 旋转向量 w 可以被解出为:

$$w = \begin{bmatrix} P_{0i} \\ P_{0j} \\ P_{ij} \end{bmatrix}^{-1} \begin{bmatrix} Q_{0i} \\ Q_{0j} \\ Q_{ij} \end{bmatrix} \quad (6)$$

方程 (6) 意味着 3 个点足以确定旋转向量. 值得指出, 多数已有方法都是根据大量的已知模型点, 通过非线性优化同时估计旋转和平移向量. 尽管这些方法可以获得一些好的估计结果, 但它们是复杂的. 而且, 如果只有 3 个已知模型点或者只需估计旋转向量, 那该怎么办? 因此, 我们认为根据尽可能少的模型点独立地估计旋转矩阵是非常重要的.

3 计算平移和深度 (Computation of translation and depths)

一旦旋转矩阵被估计出来, 就可以根据相机和一个点对之间的几何关系直接计算出平移向量和深度. 假设 G_0 和 G_j 是这个点对, 如图 1 所示, 可以得到:

$$\cos\alpha = \frac{\overline{(CG_0 \cdot CG_j)}}{\|CG_0\| \cdot \|CG_j\|} \quad (7)$$

$$\cos\beta = \frac{\overline{(R G_0 G_j) \cdot CG_j}}{\|R G_0 G_j\| \cdot \|CG_j\|} \quad (8)$$

即:

$$\cos\alpha = \frac{([g_{0x}, g_{0y}, 1]^T \cdot [g_{jx}, g_{jy}, 1]^T)}{\sqrt{g_{0x}^2 + g_{0y}^2 + 1} \cdot \sqrt{g_{jx}^2 + g_{jy}^2 + 1}} \quad (9)$$

$$\cos\beta = \frac{\overline{(R G_0 G_j) \cdot [g_{jx}, g_{jy}, 1]^T}}{\|R G_0 G_j\| \cdot \sqrt{g_{jx}^2 + g_{jy}^2 + 1}} \quad (10)$$

根据这两个方程, 可以容易地解出 α 和 β . 假设 $d_0 = \|CG_0\|$ 、 $d_j = \|CG_j\|$ 以及 $d_{0j} = \|R G_0 G_j\|$, 则可以得到下面的方程:

$$\frac{d_0}{\sin\beta} = \frac{d_j}{\sin(\alpha + \beta)} = \frac{d_{0j}}{\sin\alpha} \quad (11)$$

由于 d_{0j} 已知, d_0 就可以由上面的方程解出. 因此, 可以得到平移向量为:

$$T = \frac{d_0}{\sqrt{g_{0x}^2 + g_{0y}^2 + 1}} \begin{pmatrix} g_{0x} \\ g_{0y} \\ 1 \end{pmatrix} \quad (12)$$

类似地, 可以得到 G_j 的深度 Z_j 是:

$$Z_j = \frac{d_j}{\sqrt{g_{jx}^2 + g_{jy}^2 + 1}} \quad (13)$$

显然, 平移向量和深度的计算是非常简洁和直观的, 这进一步表明了使用点对以及单独估计旋转矩阵的优势.

4 优化算法 (Optimization algorithm)

上述估计过程仅仅根据 3 个模型点估计深度和运动参数. 如果已知模型点不只 3 个, 那么估计结果

可以被优化. 假设给定 n 个模型点, 任取其中的 3 个都可以根据方程 (6) 计算出一个旋转向量 w_i . 这样计算出的 w_i 共有 C_n^3 个. 求取所有这些可能的 w_i 的平均值. 将每个 w_i 与该平均值比较, 找出差别最大的那个 w_i . 排除这个 w_i 并重新计算剩下的所有的 w_i 的平均值. 重复这一过程直到只剩下两个 w_i . 将它们的平均值作为最终估计的旋转向量 w . 深度和平移向量的优化是类似的. 不同的是, 平移向量共有 $n-1$ 个估计值, 但每个模型点都有 $n-1$ 个估计的深度值.

相对大多数已有的估计方法而言, 这一优化算法是非常简单的. 然而, 它是有效的. 由于只需 3 个模型点就可以估计出旋转向量 w . 如果给定的模型点很多, 那么估计出的 w_i 就非常多, 例如, 给定 10 个模型点可以得到 120 个估计的 w_i . 另一方面, 造成 w_i 估计不准确的主要原因是图像点检测的不准确性, 但在一般情况下, 完全检测错误的图像点的数目只占所有图像点的很少部分甚至没有, 因而在所有估计的 w_i 中, 完全估计错误的 w_i 是很少的甚至没有. 因此, 上述优化算法可以排除完全估计错误的 w_i 并最终得到比较准确的旋转向量估计. 后面的试验结果将表明, 如果错误检测的图像点的数目小于总图像点数目的一半, 那么通过该算法总能得到好的估计结果.

5 试验和分析 (Experiments and analysis)

5.1 模拟试验

为验证该方法的准确性, 在所有这些试验中, 长度的单位是厘米而角度的单位是度. 在边长为 150cm 的立方体内随机产生 3 个模型点, 它们由 (R, T) 变换后再根据针孔相机模型投影到焦距为 1 的图像平面. 根据前面的深度和运动估计方程, 对不同的 (R, T) 估计的运动参数以及计算的这 3 个点的深度值与它们的实际值完全一致. 这就表明, 如果 3 个模型点及其图像点精确给定, 则该方法可以准确地估计深度和运动参数.

然而在实际应用中, 由于各种影响因素 (如图像噪声), 很难精确得到所有模型点的图像点. 因此, 分析该方法的鲁棒性是必要的. 在后面的试验中, 如果没有明确说明, 则随机产生的 10 个模型点在经过变换和投影后被用于深度和运动估计, 如同前面的试验一样. 平移向量都是 $[120, 100, 120]^T$.

现在, 我们在有错误图像点的情况下估计深度和运动参数. 为了模拟, 这里的错误图像点指的是该图像点的实际图像坐标被加上了 0 到 10 个像素之间

的随机噪声.当实际的旋转向量 w 是 $[1, 1, 1]^T$ 、 $[3, 3, 3]^T$ 或 $[5, 5, 5]^T$, 以及对每个 w 错误的图像点数目由 0 变到 10 (即所有的图像点都可能是错误的) 时, w_1 的估计结果如图 2 所示. 为了清晰, w_2 和 w_3 的估计结果没有图示, 事实上, 它们与 w_1 的类似. 图 2 表明, 当 10 个图像点中有 6 个错误时, 该方法依然可以准确估计出 w_1 . 另一方面, 如果错误的图像点超过 6 个, 估计的 w_1 只稍微偏离其实际值. 但必须指出, 这两个结论不是总成立. 另外的试验结果表明, 只有当错误图像点的数目小于总图像点的一半时, 该方法总可以得到准确的估计结果, 而且与错误图像点的噪声大小无关, 也就是说, 即使所加随机噪声大于 10 像素, 同样可以得到图 2 所示的估计结果. 所有这些试验结果表明: 该方法可以自动排除错误图像点并根据剩下的图像点准确估计旋转向量, 或者如果所有的图像点都在不同程度上是错误的, 那么它也能根据相对正确的图像点比较准确地估计旋转向量. 事实上, 假设错误图像点只占有所有图像点的很少部分是合理的, 因此该方法是有效的.

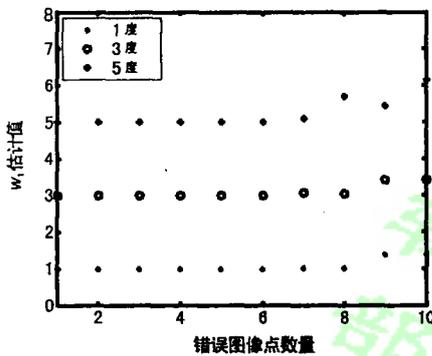


图 2 当 w 不同且错误图像点增加时 w_1 的估计结果
Fig. 2 Estimated w_1 when w varies and the number of false image point increases

估计出旋转矩阵后, 所有模型点的深度就能被计算出来, 结果图 3 所示. 在该试验中, 前面 4 个图像点是错误的, 但其余的是正确的. 加到这 4 个错误图像点上的随机噪声大小是 0 到 5 个像素. 结果表明, 对正确的图像点计算的深度值也是完全准确的, 这是因为错误图像点的数目小于总图像点的一半因而估计的旋转矩阵是准确的. 然而, 对错误图像点而言, 由于随机噪声的影响, 计算的深度值不同程度地偏离其实际值. 由于平移向量的估计与深度的类似, 限于篇幅没有列出其估计结果.

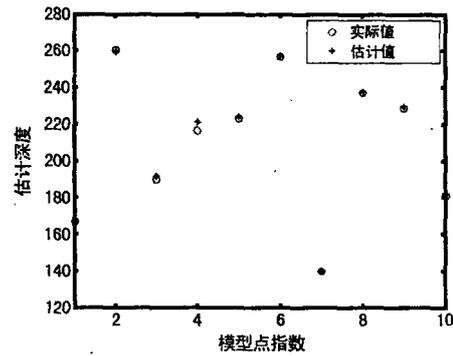


图 3 深度估计结果
Fig. 3 Estimated depths

上面的试验中, 我们假设旋转很小因而旋转矩阵 R_1 可以由 R 近似. 显然, 随着旋转角度的增大, 这种近似会导致估计误差. 图 4 显示了实际的偏航角 ω 与估计的 w_1 之间的差值, 即估计误差. 我们用 5 个或 10 个模型点估计这两种情形.

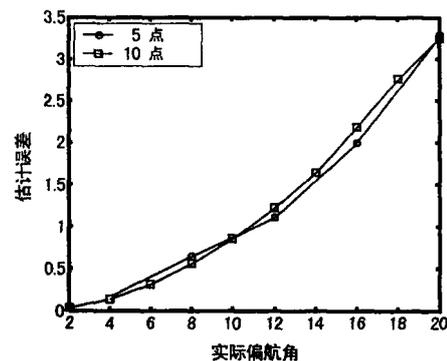


图 4 ω 增加时的估计误差
Fig. 4 Estimation errors when ω increases

为了方便, 如果模型点是 5 个, ω 每次增加 4° , 否则 ω 每次增 2° . 所有的模型点由 (R_1, T) 而不是 (R, T) 变换后再根据针孔相机模型投影, 即所有的图像点精确已知. 由图 4 可以明显看出估计误差随 ω 增加而增大. 当 ω 为 10 时, 误差大致是 1. 当 ω 为 20 时, 误差大致是 3.3. 对不同数目的模型点, 估计误差非常类似. 这些结果进一步表明: 只要模型点不是太少, 该方法就能根据这少量的模型点得到比较准确的估计结果. 这是该方法的一个吸引人的特征. 如果不是所有的图像点都正确但错误图像点的数目小于总图像点的一半, 则估计误差是类似的. 限于篇幅, 不再给出试验结果.

5.2 实际试验

我们还做了一些实际试验以验证该方法的实用

性.图 5 是 $T=[-4.5, 5.1, 88.6]^T$ 以及 $[\theta, \phi, \omega] = [-0.1, 10.0, 10.0]$ 时拍摄的一幅图像.为了准确地获取模型点的三维坐标,我们用一个标定参照物作为目标.总共选取 6 个模型点,它们的三维坐标分别是 $[0, -12.0, 0]$ 、 $[16.0, -12.0, 0]$ 、 $[0, -4.0, 0]$ 、 $[16.0, -4.0, 0]$ 、 $[12.0, -8.0, 0]$ 和 $[4.0, -8.0, 0]$.相应的图像点分别是 $[256, 212]$ 、 $[436, 210]$ 、 $[257, 301]$ 、 $[437, 301]$ 、 $[391, 256]$ 和 $[301, 256]$.这些图像坐标是由手工检测得到的.这 6 个模型点的实际深度值分别是 86.4、83.5、87.8、85.0、85.0 和 86.4.

图 5 实际试验图像

Fig. 5 An image for real experiments

根据前面的估计方程,估计的旋转向量 w 是 $[-0.09, 9.7, 9.8]$,计算的深度分别是 88.4、86.2、89.8、86.8、87.1 和 88.3.显然,估计结果与它们的实际值是很接近的.估计误差的来源主要是图像坐标检测的不准确性.我们认为,能根据 6 个模型点获得这样的估计结果是非常令人鼓舞的.

6 结论 (Conclusion)

本文针对基于模型的姿态估计问题提出了一种新颖而简单的深度和运动估计方法.该方法基于相

机和点对之间的几何关系估计深度和运动参数.为估计旋转向量,需要 3 个模型点但只需一个点对计算模型点的深度.该方法的主要优点是独立地估计旋转矩阵而且根据少量的模型点就可以得到好的估计结果.试验结果表明了该方法的鲁棒性和实用性.

参考文献 (References)

- [1] Tagawa N, Inagaki A, Minagawa A. Parametric estimation of optical flow from two perspective views [J]. IEICE Transactions on Information and Systems, 2001, E84-D(4): 485-494.
- [2] Park S K, Kweon I S. Robust and direct estimation of 3-D motion and scene depth from stereo image sequences [J]. Pattern Recognition, 2001, 34(9): 1713-1728.
- [3] Ho A, Pong T C. Cooperative fusion of stereo and motion [J]. Pattern Recognition, 1996, 29(1): 121-130.
- [4] Nuno G, Helder A. Estimation of 3D motion from stereo images - uncertainty analysis and experimental results [A]. Proceedings of the IEEE International Conference on Intelligent Robots and Systems [C]. 2002. vol.1. 7-12.
- [5] Oberkempf D, Dementhon D F, Davis L S. Iterative pose estimation using coplanar feature points [J]. Computer Vision and Image Understanding, 1998, 63(3): 495-511.
- [6] Or S H, Luk W S, Wong K H, et al. An effective iterative pose estimation algorithm [J]. Image and Vision Computing, 1998, 16(5): 353-362.
- [7] Wang F. A simple and analytical procedure for calibrating extrinsic camera parameters [J]. IEEE Transactions on Robotics and Automation, 2004, 20(1): 121-124.
- [8] Newman P M. On the structure and solution of the simultaneous localisation and map building problem [D]. Australian: The University of Sydney, 1999.

作者简介:

钟志光 (1974-),男,博士研究生.研究领域:移动机器人,计算机视觉.

易建强 (1963-),男,博士生导师.研究领域:移动机器人,智能控制.

赵冬斌 (1972-),男,硕士生导师.研究领域:移动机器人,智能控制.