



平面运动约束下的相对位姿估计最小配置解

关棒磊^{1,2}, 赵季^{3*}, 尚洋^{1,2}, 于起峰^{1,2}

1. 国防科技大学空天科学学院, 长沙 410073;
2. 图像测量与视觉导航湖南省重点实验室, 长沙 410073;
3. 华中科技大学人工智能与自动化学院, 武汉 430074

* E-mail: zhaoji84@gmail.com

收稿日期: 2024-01-17; 接受日期: 2024-06-17; 网络版发表日期: 2024-11-12

湖南省优秀青年科学基金项目(编号: 2023JJ20045)资助

摘要 基于视觉的相对位姿估计是自动驾驶车辆等移动机器人进行自主定位与环境感知的核心关键技术. 为提高相对位姿估计算法的精度、效率和鲁棒性, 寻求相对位姿估计的最小配置解一直是重点关注的研究热点. 传统的相对位姿估计算法通常仅使用图像匹配点对中的图像坐标信息, 忽略特征描述子提供的特征旋转角和尺度等额外信息. 针对移动机器人应用中单目相机平面运动这一常见情景, 本文利用图像特征描述子提供的信息中的约束方程, 提出了一种新的平面运动估计最小配置解. 根据地面场景上图像特征满足的单应约束, 利用单个旋转不变特征中的图像坐标和旋转角信息, 闭式求解单目相机平面运动的相对位姿. 由于该方法所需要的特征数量最少, 因此可以高效地与RANSAC方法或直方图投票方法相结合, 用于两视图之间的初始运动估计和误匹配点对剔除. 仿真和公开数据集实验表明, 本文能够明显提高单目相机平面运动估计的精度和鲁棒性, 可运用于自动驾驶车辆等移动机器人在地面、道路等典型场景上的自主定位与视觉感知.

关键词 相对位姿估计, 单目相机, 最小配置解, 旋转不变特征

1 引言

数十年来, 基于视觉的同时定位与地图构建(simultaneous localization and mapping, SLAM)和运动恢复结构(structure from motion, SfM)是移动机器人进行自主定位与环境感知的重要方法, 已广泛应用于自动驾驶、视觉导航等场景^[1-3]. 两视图之间的相对位姿估计是SLAM和SfM系统的重要组成部分, 一直被视为移动机器人自主定位和环境感知的基础算法^[4,5], 如图1所示. 因此, 提高相对位姿估计算法的精度、效率

和鲁棒性一直是重点关注的研究热点^[6-8].

典型的单目SLAM和SfM系统由三个主要步骤组成: 首先, 通过图像特征匹配算法来建立不同视图之间的图像匹配点对; 然后, 采用随机抽样一致(random sample consensus, RANSAC)^[9]等算法剔除图像匹配点对中的误匹配点对; 最后, 利用图像匹配点对中的内点求解视图之间的相对位姿关系. 其中误匹配点对剔除步骤对于提高相对位姿估计算法的精度和鲁棒性至关重要, 同时误匹配点对剔除的效率直接影响SLAM和SfM系统的实时性能. 因为在相同的误匹配点对比例

引用格式: 关棒磊, 赵季, 尚洋, 等. 平面运动约束下的相对位姿估计最小配置解. 中国科学: 技术科学, 2024, 54: 2122-2130
Guan B L, Zhao J, Shang Y, et al. Minimal solver for relative pose estimation under planar motion (in Chinese). Sci Sin Tech, 2024, 54: 2122-2130, doi: 10.1360/SST-2024-0021

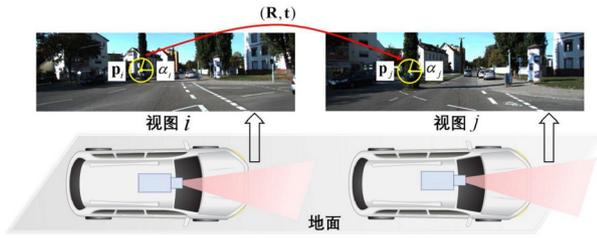


图 1 (网络版彩图)单目平面运动估计

Figure 1 (Color online) Monocular planar motion estimation.

条件下, RANSAC的随机采样次数随着最小配置解所需图像匹配点对数量的增加而成指数型增加. 减少最小配置解所需要的图像匹配点对数量能够有效提高相对位姿估计算法的效率. 以RANSAC置信度为99%、正确图像匹配点对所占比例50%的情况为例, 当最小配置解所需的图像匹配点对数量为1和2时, RANSAC采样次数分别为7和16. 因此, 研究获得相对位姿估计的最小配置解显得至关重要^[10-12].

单目相对位姿估计最小配置解问题的研究最早起源于Hartley等人^[4]的工作, 他们提出了利用7个图像匹配点对求解相机内参数未知条件下的基础矩阵. 在相机内参数部分已知、仅焦距未知的条件下, Kukulova等人^[13]和Stewénius等人^[14]通过6个图像匹配点对估计两视图之间的相对位姿. 在相机内参数已知的条件下, 针对三维场景, Nistér等人^[15]利用5个图像匹配点对求解两视图之间的本质矩阵, 由于单相机相对位姿估计无法确定平移尺度信息, 通过5个图像匹配点对可求解五自由度的相对位姿, 即3个自由度的旋转关系和2个自由度的平移向量; 针对平面场景, Hartley等人^[4]通过4个图像匹配点对求解两视图之间的单应矩阵. 为进一步减少相对位姿估计最小配置解所需的图像匹配点对数量, 国内外学者通过引入相机运动模型假设来降低相对位姿估计问题的自由度. 典型地, 平面运动是自动驾驶车辆等移动机器人应用中的常见情形. 当相机运动满足平面运动约束时, 两视图之间的相对位姿估计问题从5个自由度减少为2个自由度, 可通过1个旋转角和1个平移方向来表述. Choi和Kim^[16]、Ortín和Montiel^[17]和Chou等人^[18]采用2个图像匹配点对即可估计出平面运动约束下两视图之间的相对位姿. 此外, 对于满足阿克曼运动模型(Ackermann motion model)的轮式移动机器人, 两视图之间的相对位姿估计问题降低为1个自由度, 仅通过1个图像匹配点对即可恢复单

相机的相对运动^[19,20].

上述相对位姿估计算法通常采用SIFT(scale-invariant feature transform)^[21]、SURF(speeded up robust features)^[22]和ORB(oriented FAST and rotated BRIEF)^[23]等旋转尺度不变特征描述子获取图像匹配点对, 但仅使用图像匹配点对中的图像坐标, 直接忽略特征描述子提供的特征旋转角和尺度等额外信息. 最新的研究表明特征描述子中的额外信息能够提供新的约束方程, 可用于单目相机的相对位姿估计. 例如, Barath等人^[24]通过2个图像匹配点对及相应的旋转角和尺度信息估计两视图之间的单应矩阵; Ding等人^[25]采用地面场景上1个图像匹配点对及相应的旋转角和尺度信息求解已知垂直方向条件下的相机相对位姿. 通过充分挖掘特征描述子提供的图像匹配信息, 能够有效减少相对位姿估计最小配置解所需的图像匹配点对数量^[8,26].

本文针对移动机器人应用中单目相机平面运动这一常见情景, 通过SIFT和SURF等特征描述子提取两视图之间的旋转不变特征, 根据地面场景上图像特征满足的单应约束, 利用单个旋转不变特征中的图像坐标和旋转角信息, 闭式求解单目相机平面运动的相对位姿. 值得注意的是本文方法只利用旋转不变特征中旋转角信息, 不需要特征尺度信息, 因此提出的方法也适用于仅旋转不变的特征描述子, 例如ORB特征描述子. 仿真和公开数据集实验表明, 本文能够明显提高单目相机平面运动估计的精度和鲁棒性, 可运用于自动驾驶车辆等移动机器人在地面、道路等典型场景上的自主定位与视觉感知.

2 两视图几何约束

2.1 单应约束

假设单目相机内参数已知, 从两个不同视角对同一个三维平面场景拍摄图像, 分别获得视图*i*和*j*. 两视图之间的图像匹配点对满足平面单应约束, 可表示为^[4]

$$\lambda \mathbf{p}_j = \mathbf{H} \mathbf{p}_i, \quad (1)$$

其中, $\mathbf{p}_i = [u_i, v_i, 1]^T$ 和 $\mathbf{p}_j = [u_j, v_j, 1]^T$ 是三维平面上的场景点分别投影在视图*i*和*j*上的图像齐次坐标, λ 为比例因子. 单应矩阵 \mathbf{H} 可表示为

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}. \quad (2)$$

单应矩阵满足以下关系式:

$$\mathbf{H} = \mathbf{R} - \frac{1}{d} \mathbf{t} \mathbf{N}^T, \quad (3)$$

其中, \mathbf{R} 和 \mathbf{t} 分别表示视图 i 相机坐标系到视图 j 相机坐标系的旋转矩阵和平移向量, d 是相机坐标系 i 到三维平面的距离, $\mathbf{N} = [n_1, n_2, n_3]^T$ 是三维平面在相机坐标系 i 中的单位法向量.

2.2 特征旋转角和尺度约束

目前广泛采用的SIFT和SURF等旋转尺度不变特征描述子不仅提供了两视图之间的图像匹配点对坐标 $(\mathbf{p}_i, \mathbf{p}_j)$, 而且提供了相应的特征旋转角和尺度等额外信息. 这些额外信息本质上描述了图像匹配点之间的邻域关系, 可采用 2×2 的局部仿射矩阵来表示. 局部仿射矩阵 \mathbf{A}_{ij} 表示以图像坐标 \mathbf{p}_i 为中心的邻域与以图像坐标 \mathbf{p}_j 为中心的邻域之间的局部仿射变换关系, 可定义为^[27]

$$\begin{aligned} \mathbf{A}_{ij} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} q_u & w \\ 0 & q_v \end{bmatrix} \\ &= \begin{bmatrix} q_u \cos(\alpha) & w \cos(\alpha) - q_v \sin(\alpha) \\ q_u \sin(\alpha) & w \sin(\alpha) + q_v \cos(\alpha) \end{bmatrix}, \end{aligned} \quad (4)$$

其中, α 表示图像匹配点对的特征旋转角, 可由视图 i 和 j 上同名点对的旋转角之差 $\alpha_j - \alpha_i$ 计算得到; q_u 和 q_v 分别表示图像特征在图像 x 轴和 y 轴方向的尺度大小; w 表示错切参数. 从几何关系上看, 局部仿射矩阵 \mathbf{A}_{ij} 是图像匹配点对邻域满足的单应矩阵在图像匹配点对的一阶近似矩阵^[27]. 矩阵 \mathbf{A}_{ij} 中的各元素可表示为

$$\begin{cases} a_{11} = \frac{\partial u_j}{\partial u_i} = \frac{h_{11} - h_{31} u_j}{m}, & a_{12} = \frac{\partial u_j}{\partial v_i} = \frac{h_{12} - h_{32} u_j}{m}, \\ a_{21} = \frac{\partial v_j}{\partial u_i} = \frac{h_{21} - h_{31} v_j}{m}, & a_{22} = \frac{\partial v_j}{\partial v_i} = \frac{h_{22} - h_{32} v_j}{m}, \end{cases} \quad (5)$$

其中, $a_{ij} (i = 1, 2; j = 1, 2)$ 表示局部仿射矩阵 \mathbf{A}_{ij} 中第 i 行第 j 列的元素. $m = u_i h_{31} + v_i h_{32} + h_{33}$ 表示投影深度. 根

据式(4)和(5), 可以建立图像特征旋转角和尺度与单应矩阵之间的约束公式.

3 平面运动估计算法

3.1 平面运动参数化

针对移动机器人应用中典型的平面运动, 单目相机的图像平面垂直于地面, 如图2所示. 相机坐标系的 X 轴和 Y 轴分别平行于图像平面的横轴和纵轴, Z 轴与相机光轴重合, 相机坐标系的 Y 轴垂直于地面. 因此, 两个视图之间仅存在绕相机 Y 轴的旋转和二维平面位移, 因此从视图 i 到 j 的旋转矩阵 $\mathbf{R} = \mathbf{R}_y$ 和平移向量 \mathbf{t} 可以分别表示为

$$\mathbf{R}_y = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \quad (6)$$

$$\mathbf{t} = -\mathbf{R}_y \begin{bmatrix} \rho \sin(\phi) \\ 0 \\ \rho \cos(\phi) \end{bmatrix}, \quad (7)$$

其中, θ 表示相机坐标系绕 Y 轴的偏航角, ϕ 表示相机在平面上移动的方向角, ρ 表示两视图之间相机运动的距离.

单目相机在地面运动的过程中, 地面在相机坐标系的法向量可表示为 $\mathbf{N} = [0, 1, 0]^T$. 将式(6)和(7)代入式(3), 两视图之间的单应矩阵可简化为

$$\mathbf{H} = \begin{bmatrix} \cos(\theta) & -s \sin(\theta - \phi) & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & s \cos(\theta - \phi) & \cos(\theta) \end{bmatrix}, \quad (8)$$

其中, 尺度因子 $s = \rho / d$. 由式(8)可知, 在相机平面运动

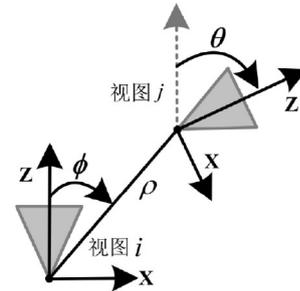


图2 (网络版彩图)两视图之间的平面运动参数化
Figure 2 (Color online) Parametrization of planar motion between two views.

条件下, 两视图之间的地面单应矩阵由相机的偏航角 θ 、平移方向角 ϕ 和尺度因子 s 等三个未知参数组成. 该单应矩阵描述了单目相机平面运动条件下同一地面点在两视图之间的同名点对映射关系. 已知地面单应矩阵和视图 i 中地面点对应的图像坐标, 可直接求出视图 j 中同名点的图像坐标.

3.2 平面运动几何约束

两视图之间的平面运动几何约束由地面上图像匹配点对的单应约束、特征旋转角约束两部分组成. 对于图像匹配点对的单应约束, 为了消除式(1)中的比例因子 λ , 在方程的两边乘以反对称矩阵 $[\mathbf{p}_j]_{\times}$:

$$[\mathbf{p}_j]_{\times} \mathbf{H} \mathbf{p}_i = 0. \quad (9)$$

通过把式(8)代入式(9), 并扩展可以得到以下三个方程:

$$\begin{cases} v_i v_j s \cos(\theta - \phi) + u_i v_j \sin(\theta) + v_j \cos(\theta) - v_i = 0, \\ -u_j v_i s \cos(\theta - \phi) - v_i s \sin(\theta - \phi) - (u_i u_j + 1) \sin(\theta) \\ \quad + (u_i - u_j) \cos(\theta) = 0, \\ v_i v_j s \sin(\theta - \phi) - u_i v_j \cos(\theta) + v_j \sin(\theta) + u_j v_i = 0. \end{cases} \quad (10)$$

由于反对称矩阵 $[\mathbf{p}_j]_{\times}$ 的秩是2, 式(10)三个方程中任取两个方程都是独立的, 故对未知参数 θ , ϕ 和 s 仅存在两个独立的约束.

对于图像特征旋转角约束, 通过分析式(4)和(5)可知, 采用仿射矩阵中的 a_{11} 和 a_{21} 相除能够消除图像特征的尺度因子, 进而建立特征旋转角与单应矩阵之间的关系:

$$\frac{a_{11}}{a_{21}} = \frac{\cos(\alpha)}{\sin(\alpha)} = \frac{h_{11} - h_{31} u_j}{h_{21} - h_{31} v_j}. \quad (11)$$

通过把式(8)代入式(11), 扩展得到以下公式:

$$(v_j \cos(\alpha) - u_j \sin(\alpha)) \sin(\theta) + \sin(\alpha) \cos(\theta) = 0. \quad (12)$$

上式提供了仅关于未知参数 θ 的约束方程.

3.3 平面运动估计

根据式(12)可知, 可以直接计算获得两视图之间的相机偏航角, 偏航角 θ 的计算公式表示为

$$\theta = \arctan\left(\frac{\sin(\alpha)}{u_j \sin(\alpha) - v_j \cos(\alpha)}\right). \quad (13)$$

考虑到移动机器人等典型应用场景, 相机的采集帧率一般在每秒20帧以上, 故相邻两帧图像之间的偏航角大小通常处于 $[-\pi/2, \pi/2]$. 因此根据式(13)可以求解得到唯一的两视图偏航角.

进而, 通过把两视图之间的偏航角解算结果代入式(10)中的第一行和第三行, 可以计算得到 $s \sin(\theta - \phi)$ 和 $s \cos(\theta - \phi)$:

$$\begin{cases} s \sin(\theta - \phi) = \frac{u_i v_j \cos(\theta) - v_j \sin(\theta) - u_j v_i}{v_i v_j}, \\ s \cos(\theta - \phi) = \frac{v_i - u_i v_j \sin(\theta) - v_j \cos(\theta)}{v_i v_j}. \end{cases} \quad (14)$$

为了消除尺度因子 s , 式(14)的第一行与第二行相除, 计算到两视图之间的平移方向角 ϕ :

$$\phi = \theta - \arctan\left(\frac{u_i v_j \cos(\theta) - v_j \sin(\theta) - u_j v_i}{v_i - u_i v_j \sin(\theta) - v_j \cos(\theta)}\right). \quad (15)$$

基于单个旋转不变特征的单目平面运动估计方法的算法流程如表1所示.

综上所述, 本方法提出的方法只采用单个旋转不变特征中的图像坐标和旋转角信息, 即可闭式求解单目相机平面运动条件下的两视图相对位姿, 同时获得两视图图像匹配点对中的内点. 常用的SIFT, SURF和ORB等特征描述子可以直接获得图像匹配点对及相应的特征旋转角. 对于 $n(n \geq 2)$ 个旋转不变特征的条件,

表1 算法流程图

Table 1 Algorithm flow chart

算法	基于单个旋转不变特征的单目平面运动估计
应用场景:	平面运动
输入:	单个图像匹配点对坐标 $(\mathbf{p}_i, \mathbf{p}_j)$ 和相应的图像特征旋转角 α
输出:	单目相机的偏航角 θ 和平移方向角 ϕ
1:	采用SIFT, SURF和ORB等特征描述子提取两视图之间的旋转不变特征 $(\mathbf{p}_i, \mathbf{p}_j, \alpha)$
2:	根据图像匹配点对坐标 $(\mathbf{p}_i, \mathbf{p}_j)$ 的单应约束, 可构建两个独立约束方程(式(10))
3:	根据图像特征旋转角 α 的特征旋转角约束, 可构建一个约束方程(式(12))
4:	通过式(12)的约束方程求解相机的偏航角 θ , 计算公式如式(13)所示
5:	将偏航角 θ 代入式(15)中求解得到相机的平移方向角 ϕ
6:	根据式(6)和(7)恢复出视图 i 到 j 的旋转矩阵 \mathbf{R} 和平移向量 \mathbf{t}

根据式(10)第一和第三行的两个独立约束和式(12), 可得到 $3n$ 个关于 $\sin(\theta)$, $\cos(\theta)$, $s\sin(\theta-\phi)$ 和 $s\cos(\theta-\phi)$ 的线性方程组, 通过最小二乘方法求解得到相机偏航角 θ 和平移方向角 ϕ . 此外, 在内点判断过程中采用以下策略: 通过 \mathbf{R} 和 \mathbf{t} 构造视图 i 和 j 之间的本质矩阵 $\mathbf{E} = [\mathbf{t}]_{\times}\mathbf{R}$, 用于判断图像匹配点对是否为内点. 相比于通过单应矩阵判断内点的策略, 只有相机视场内来自地面的图像匹配点对才满足平面单应约束, 而非地面的图像匹配点对则直接被当作外点进行了剔除, 这将导致获得的内点数量减少. 本文采用本质矩阵判断内点的策略, 可适用于相机视场内的全部场景, 包括来自地面、建筑等的图像匹配点对. 这样整个场景中的正确图像匹配点对都能被选为内点, 而不仅限于地面上的图像匹配点对, 能够极大提高图像匹配内点的数量. 由于本文方法所需要的特征数量最少, 因此可以高效地与RANSAC方法或直方图投票方法相结合, 提高平面运动估计的精度和鲁棒性, 同时快速获得两视图图像匹配点对中的内点, 可作为输入用于后续的非线性优化等方法, 实现更精确的相机运动估计.

4 实验验证

本节通过仿真和实际实验来验证所提出的单目平面运动估计方法. 仿真实验主要是对本文方法在图像噪声和非平面噪声影响下的精度进行验证, 实际实验中通过公开数据集来验证本方法在真实数据条件下的精度. 为方便描述, 根据所需图像匹配点对的个数对方法的名称进行简写. 本文提出的单目平面运动估计方法称为1PC-Our方法, 其中1PC特指两视图之间的旋转不变特征. 实验中将1PC-Our方法与现有经典的两视图平面运动估计方法2PC-Choi^[16]进行了详细对比. 为克服图像误匹配点对对运动估计精度的影响, 所有的方法都加入到RANSAC等鲁棒估计框架或者直方图投票中. 对于RANSAC框架, 选取生成最多内点数量的相对位姿结果作为最终的估计结果; 对于直方图投票框架, 选取直方图峰值对应的相对位姿结果作为最终的估计结果. 实验中2PC-Choi方法加入到RANSAC框架中. 1PC-Our方法仅需要单个旋转不变特征点进行平面运动估计, 该方法则与直方图投票框架或RANSAC框架结合.

实验评估中分别对旋转估计误差和平移估计误差进行衡量, 其中旋转估计误差是比较旋转估计结果和真值之间的角度差. 由于两视图之间的平移向量只能恢复到尺度大小, 所以同样比较平移向量估计结果和真值之间的角度差. 旋转和平移的估计误差计算公式分别为

$$\zeta_{\mathbf{R}} = \arccos\left(\frac{(\text{trace}(\mathbf{R}_{gt}\mathbf{R}^T) - 1)}{2}\right),$$

$$\zeta_{\mathbf{t}} = \arccos\left(\frac{(\mathbf{t}_{gt}^T \mathbf{t})}{(\|\mathbf{t}_{gt}\| \|\mathbf{t}\|)}\right),$$

其中, \mathbf{R}_{gt} 和 \mathbf{t}_{gt} 分别表示视图 i 和 j 之间的旋转矩阵和平移向量真值. \mathbf{R} 和 \mathbf{t} 分别表示旋转矩阵和平移向量的估计结果.

4.1 仿真实验

仿真场景由1个地平面和50个随机平面组成. 这些三维平面随机分布在 X 坐标 $[-5, 5]$ m, Y 坐标 $[-5, 5]$ m, Z 坐标 $[10, 20]$ m的范围内. 从地平面随机选取50个三维场景点, 并从50个平面上随机选取1个三维场景点, 因此场景中共有100个三维场景点. 单目相机的图像分辨率是 640×480 像素, 焦距为400像素, 主点定义在图像中心 320×240 . 单目相机平面运动由绕 Y 轴的偏航角 θ 和平移方向角 ϕ 表示, 两个角度在 $[-10^\circ, 10^\circ]$ 随机生成.

根据单目相机的内参数和外参数, 将三维场景点投影到两视图中获得图像匹配点对. 此外, 图像匹配点对相应的特征旋转角通过图像匹配点对邻域中的局部仿射矩阵分解得到, 具体生成步骤如下: 首先在三维平面内随机选取4个额外场景点, 并投影到两视图中获得图像同名点对; 然后根据4个图像同名点对, 计算三维平面在两视图中满足的单应矩阵; 再根据式(5)通过单应矩阵在图像匹配点对处进行一阶近似求解, 获得图像匹配点对邻域之间的局部仿射矩阵; 最后根据式(11)求解得到图像匹配点对对应的特征旋转角. 仿真过程中图像高斯噪声不仅被加入图像匹配点对坐标中, 也被加入估计局部仿射矩阵的额外图像同名点对坐标中, 从而模拟图像匹配点对坐标误差和相应的特征旋转角误差^[24]. 在仿真实验中, 每级噪声下进行1000次独立实验, 统计五分位中前2个区间误差的均方差来评估旋转和平移的估计误差. 通过仿真实验来

验证提出方法的精度受图像噪声和非平面噪声的影响。

4.1.1 效率对比

通过评估平面运动估计方法的运行时间来进行方法效率对比。实验中CPU处理器采用Intel Core i7-8550U 1.8GHz。所有算法均在采用MATLAB软件上运行计时。实验中采用10000次随机运行时间的平均值来衡量, 表2总结了平面运动估计方法的运行时间。2PC-Choi方法将平面运动估计问题建模为寻找直线与单位圆交点的问题, 并提出了相应的闭式求解方法。1PC-Our方法的计算量主要包括式(13)和(15)的线性求解计算量, 计算复杂低。因此相比2PC-Choi方法, 本文提出的1PC-Our方法计算效率提高约16倍。

4.1.2 图像噪声对平面运动估计的影响

为评估所提出方法在不同图像噪声水平下的表现, 在图像坐标中加入均值为0、方差为 σ 的高斯噪声。图像高斯噪声的方差 σ 从0增加到2像素, 间隔0.1像素。图3展示了平面运动估计方法的精度随图像噪声增加的变化曲线, 其中图3(a)为旋转估计误差, 图3(b)为平移估计误差。由图可知, 两个平面运动估计方法的精度随图像噪声的增加而逐渐降低, 其中1PC-Our方法的旋转和平移估计精度均明显优于2PC-Choi方法。实验表明本文提出的1PC-Our方法对图像噪声具有较好的鲁棒性, 可以用于单目平面运动的快速估计和两视图之间内点的高效筛选。

4.1.3 非平面噪声对平面运动估计的影响

现实场景中的运动不一定符合完美的平面运动假设, 因此有必要验证本文方法在非平面运动条件下的估计精度。为评估所提出方法受非平面噪声的影响, 在这个仿真实验中增加非平面运动, 即绕相机X轴和Z轴的旋转运动、YZ平面上的平移运动。这些非平面噪声的均值为0, 方差从0增加到1°。图像高斯噪声的大小设定为1像素。图4展示了平面运动估计方法的精度随非平面噪声增加的变化曲线, 其中图4(a)为旋转估计误差, 图4(b)为平移估计误差。由图可知, 1PC-Our方法的旋转估计精度一直优于2PC-Choi方法。此外, 在非平面噪声小于0.4°时, 1PC-Our方法的平移估计精度要优于2PC-Choi方法。实验表明本文提出的1PC-Our方法能够在一定程度上克服非平面噪声的影响, 适用于于

表2 平面运动估计方法的运行时间对比

Table 2 Runtime comparison of plane motion estimation methods

方法	运行时间 (ms)
2PC-Choi	0.098
1PC-Our	0.006

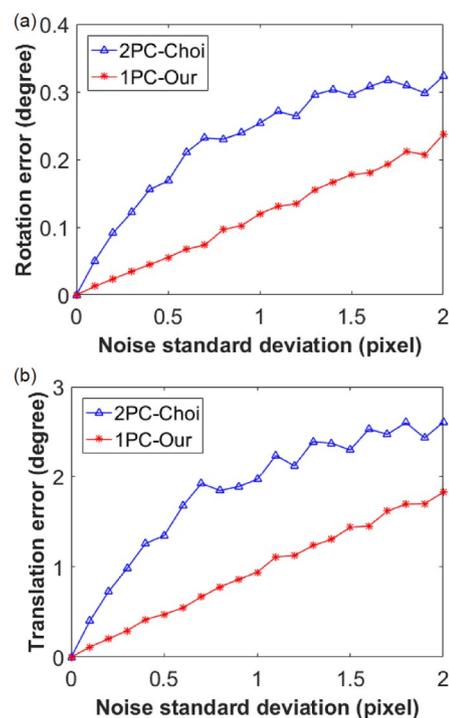


图3 (网络版彩图)平面运动估计误差随图像噪声的变化曲线图。(a) 旋转估计误差; (b) 平移估计误差

Figure 3 (Color online) Curve diagrams of the plane motion estimation error varying with image noise. (a) Rotation error; (b) translation error.

现实场景中的平面运动估计。

4.2 实际实验

本文采用自动驾驶领域知名的KITTI数据集^[28]来评估提出的平面运动估计方法。KITTI数据集提供了11个图像序列的位姿真值, 这些位姿真值通过GPS和惯性测量单元高精度融合定位获得。这11个图像序列中的所有相邻视图都用于验证平面运动估计方法的精度, 共计23000个图像对用于实际实验验证。

实验中采用SIFT算法提取两相邻视图之间的旋转不变特征点, 即获取图像匹配点对坐标和相应的特征旋转角。平面运动估计方法的相机偏航角和平移方

向角分别对相应的角度真值进行比较来计算误差, 并采用各序列中全部误差的中值来衡量方法的精度. 本文提出方法结合直方图投票框架来剔除误匹配点对: 首先根据每个旋转不变特征点计算单目相机的偏航角, 构成相机偏航角直方图; 再取直方图峰值为最终的偏航角; 最后在已知相机偏航角的条件下估计相机的平移方向角. 图5为实验中某两相邻视图的偏航角估

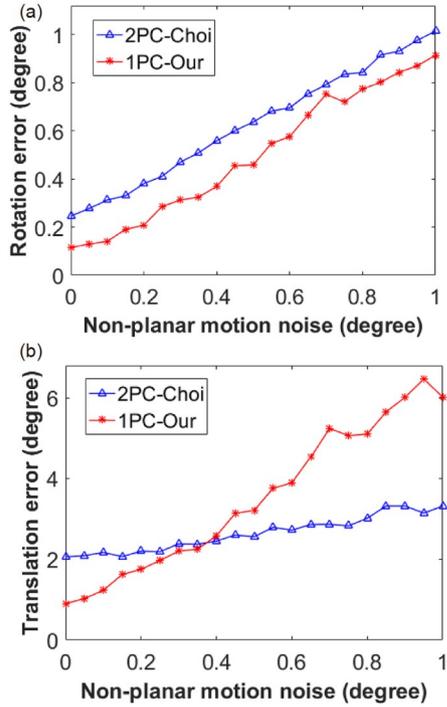


图4 (网络版彩图)平面运动估计误差随非平面噪声的变化曲线图. (a) 旋转估计误差; (b) 平移估计误差

Figure 4 (Color online) Curve diagram of plane motion estimation error varying with non-planar noise. (a) Rotation error; (b) translation error.

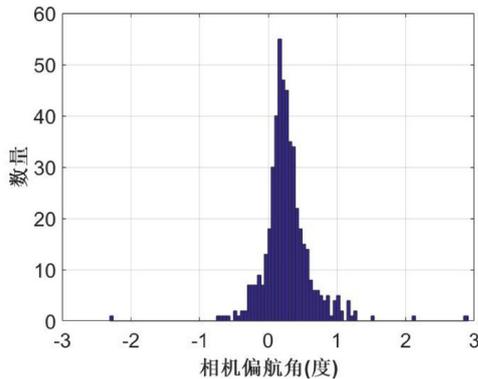


图5 (网络版彩图)直方图投票方法示例

Figure 5 (Color online) Example of histogram voting method.

计值直方图, 可见直方图的峰值相当明显.

表3总结了平面运动估计的旋转和平移估计误差. 由表3所知, 本文提出方法的相对旋转估计精度优于 0.180° , 相对平移估计精度优于 2.241° . 1PC-Our方法估计精度明显优于经典的2PC-Choi方法. 表4统计了KITTI数据集中平面运动估计方法在RANSAC框架下的相对位姿估计耗时平均值, 结果表明1PC-Our方法估计效率明显优于2PC-Choi方法. 实验结果表明提出的1PC-Our方法适用于自动驾驶场景下单目平面运动估计.

5 结论

针对自动驾驶车辆等移动机器人应用中常见的单目相机平面运动, 本文提出了一种基于单个旋转不变特征的单目平面运动估计方法, 可以高效地与直方图

表3 平面运动估计的旋转和平移估计误差

Table 3 Rotation and translation estimation errors of plane motion estimation

方法	2PC-Choi		1PC-Our	
	ζ_R	ζ_t	ζ_R	ζ_t
序列00	0.283°	5.362°	0.163°	1.956°
序列01	0.288°	5.502°	0.135°	2.097°
序列02	0.183°	3.741°	0.180°	1.709°
序列03	0.264°	7.091°	0.162°	1.705°
序列04	0.104°	3.019°	0.051°	1.559°
序列05	0.158°	4.641°	0.115°	1.784°
序列06	0.153°	3.793°	0.084°	1.471°
序列07	0.129°	4.836°	0.128°	2.241°
序列08	0.164°	4.954°	0.135°	1.949°
序列09	0.157°	3.856°	0.123°	1.465°
序列10	0.172°	4.064°	0.171°	1.708°

表4 平面运动估计方法在RANSAC框架下的运行时间对比

Table 4 Runtime comparison for plane motion estimation methods under the RANSAC framework

方法	运行时间 (s)
2PC-Choi	0.283
1PC-Our	0.109

投票方法或RANSAC方法相结合. 该方法采用SIFT, SURF和ORB等特征描述子提取两视图之间的旋转不变特征, 根据地面场景上图像特征满足的单应约束, 通过单个旋转不变特征中的图像坐标和旋转角信息, 实

现了两视图相对位姿的闭式求解. 仿真实验和KITTI数据集实验表明, 本文方法在精度和效率上均优于经典的平面运动估计方法, 可用于两视图之间的初始运动快速估计和误匹配点对高效剔除.

参考文献

- 1 Scaramuzza D, Fraundorfer F. Visual odometry: The first 30 years and fundamentals. *IEEE Robot Automat Mag*, 2011, 18: 80–92
- 2 Schönberger J L, Frahm J. Structure-from-motion revisited. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016. 4104–4113
- 3 Ma Y Q, Liu S C, Jia Y, et al. Experimental research of navigation and localization algorithm based on stereo images for the lunar rover (in Chinese). *Sci Sin Tech*, 2014, 44: 1097–1104 [马友青, 刘少创, 贾阳, 等. 基于立体图像的月球车导航定位试验研究. *中国科学: 技术科学*, 2014, 44: 1097–1104]
- 4 Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press, 2003
- 5 Liu C K, Li D S, Wang J K, et al. Precise pose measurement of lunar sampling manipulator based on binocular ellipse feature (in Chinese). *Sci Sin Tech*, 2021, 51: 1453–1464 [刘传凯, 李东升, 王俊魁, 等. 基于双目椭圆特征的月面采样机械臂末端位姿精确测量方法. *中国科学: 技术科学*, 2021, 51: 1453–1464]
- 6 Barath D. Five-point fundamental matrix estimation for uncalibrated cameras. In: *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018. 235–243
- 7 Zhao J. An efficient solution to non-minimal case essential matrix estimation. *IEEE Trans Pattern Anal Mach Intell*, 2022, 44: 1777–1792
- 8 Guan B, Zhao J, Barath D, et al. Minimal solvers for relative pose estimation of multi-camera systems using affine correspondences. *Int J Comput Vis*, 2023, 131: 324–345
- 9 Fischler M A, Bolles R C. Random sample consensus. *Commun ACM*, 1981, 24: 381–395
- 10 Barath D, Toth T, Hajder L. A minimal solution for two-view focal-length estimation using two affine correspondences. In: *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017. 2557–2565
- 11 Zhao J, Kneip L, He Y, et al. Minimal case relative pose computation using ray-point-ray features. *IEEE Trans Pattern Anal Mach Intell*, 2020, 42: 1176–1190
- 12 Duff T, Kohn K, Leykin A, et al. PLMP—Point-line minimal problems in complete multi-view visibility. In: *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, 2019. 1675–1684
- 13 Kukulova Z, Kileel J, Sturmfels B, et al. A clever elimination strategy for efficient minimal solvers. In: *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017. 4912–4921
- 14 Stewenius H, Nistér D, Kahl F, et al. A minimal solution for relative pose with unknown focal length. In: *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. San Diego, 2005. 789–794
- 15 Nistér D. An efficient solution to the five-point relative pose problem. *IEEE Trans Pattern Anal Mach Intell*, 2004, 26: 756–770
- 16 Choi S, Kim J H. Fast and reliable minimal relative pose estimation under planar motion. *Image Vis Comput*, 2018, 69: 103–112
- 17 Ortín D, Montiel J M M. Indoor robot motion based on monocular images. *Robotica*, 2001, 19: 331–342
- 18 Chou C C, Seo Y W, Wang C C. A two-stage sampling for robust feature matching. *J Field Robot*, 2018, 35: 779–801
- 19 Scaramuzza D, Fraundorfer F, Siegwart R. Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC. In: *Proceedings of 2009 IEEE International Conference on Robotics and Automation*. Kobe, 2009. 4293–4299
- 20 Huang K, Wang Y, Kneip L. Motion estimation of non-holonomic ground vehicles from a single feature correspondence measured over N views. In: *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, 2019. 4912–4921
- 21 Lowe D G. Distinctive image features from scale-invariant keypoints. *Int J Comput Vision*, 2004, 60: 91–110
- 22 Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF). *Comput Vis Image Understand*, 2008, 110: 346–359
- 23 Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF. In: *Proceedings of 2011 International Conference on Computer Vision*. Barcelona, 2011. 2564–2571

- 24 Baráth D, Kukulova Z. Homography from two orientation- and scale-covariant features. In: Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019. 1091–1099
- 25 Ding Y, Barath D, Kukulova Z. Homography-based egomotion estimation using gravity and SIFT features. In: Ishikawa H, Liu C L, Pajdla T, et al., eds. Computer Vision—ACCV 2020. ACCV 2020. Lecture Notes in Computer Science, vol 12622. Cham: Springer, 2020
- 26 Guan B, Zhao J, Li Z, et al. Relative pose estimation with a single affine correspondence. *IEEE Trans Cybern*, 2021, 52: 10111–10122
- 27 Barath D, Hajder L. Efficient recovery of essential matrix from two affine correspondences. *IEEE Trans Image Process*, 2018, 27: 5328–5337
- 28 Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset. *Int J Robot Res*, 2013, 32: 1231–1237

Minimal solver for relative pose estimation under planar motion

GUAN BangLei^{1,2}, ZHAO Ji³, SHANG Yang^{1,2} & YU QiFeng^{1,2}

¹ College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China;

² Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, China;

³ School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

Visual-based relative pose estimation is a core technology for autonomous localization and environment perception of mobile robots such as autonomous driving vehicles. In order to improve the accuracy, efficiency, and robustness of the algorithm, the minimal solvers for relative pose estimation are an important research topic. Traditional relative pose estimation algorithms typically utilize only the image coordinate information of the matched feature points, which ignores the additional information provided by feature descriptors, such as feature rotation angles and scales. In this paper, focusing on the common scenario in mobile robot applications, we propose a new minimal solver for planar motion estimation using the constraints provided by the feature descriptors. By exploiting the homography constraints, we obtain a closed-form solution for the relative pose of the monocular camera utilizing a single rotation-invariant feature. Due to the minimal number of features required by the proposed solver, it can be efficiently combined with RANSAC or histogram voting methods for initial motion estimation and removal of outlier matches. Experiments on synthetic data and public datasets demonstrate that our method significantly improves the accuracy and robustness of monocular camera planar motion estimation, which makes it applicable to autonomous localization and visual perception of mobile robots.

relative pose estimation, monocular camera, minimal solver, rotational invariant feature

doi: [10.1360/SST-2024-0021](https://doi.org/10.1360/SST-2024-0021)