

文章编号: 1001-4888(2024)02-0183-12

# 基于单个 SIFT 特征的相对位姿估计方法\*

谭泽, 关棒磊, 孙祥一

(国防科技大学 空天科学学院, 图像测量与视觉导航湖南省重点实验室, 湖南长沙 410073)

**摘要:** 位姿估计是精密光测和自动驾驶的基本问题之一。针对自动驾驶等实际应用中, 相机在平面上运动, 相机位姿的自由度为 3 的情况, 本文提出了基于单个 SIFT 特征的相机相对位姿估计方法。由于单目相机无法恢复平移尺度, 因此相机运动的自由度减少为仅有旋转角和平移角的两自由度。通过观测地面, 可以得到包含相机运动和平面法向量的地面单应信息, 因此可以通过提取地面同名点估计单应矩阵来恢复相机运动。为了减少 RANSAC 迭代次数、提高算法效率, 引入 SIFT 特征进行位姿估计。SIFT 特征包括 2 幅图像中同名点图像坐标以及其特征旋转和特征尺度, 可以扩充单个点对中包含的信息, 有效减少求解单应矩阵所需点对数量。针对平面二自由度运动情况, 本文使用单个 SIFT 特征点对完成单应矩阵的估计, 并采用随机采样一致算法对结果进行优化, 最终分解单应矩阵得到相对位姿估计结果。在仿真实验及真实实验中与 2pt 方法和 5pt 方法进行对比, 证明了所提出的方法是有效的。

**关键词:** 位姿估计; 随机采样一致算法; 自动驾驶; 平面运动; 精密光测

**中图分类号:** O348.1 **文献标识码:** A **DOI:** 10.7520/1001-4888-23-023

## 0 引言

三维重建 (Structure from Motion, SFM) 和同步定位与建图 (Simultaneous Localization and Mapping, SLAM) 一直是计算机视觉中活跃的研究主题。目前这些技术已被成功应用于各类场景, 例如自动驾驶和视觉导航等领域。基于图像的相对位姿估计是 SLAM 和 SFM 系统的重要组成部分<sup>[1]</sup>, 一直被视为 SLAM 和 SFM 系统的基础算法, 因此, 提高相对姿态估计算法的精度、效率和鲁棒性仍然是计算机视觉领域内重点关注的问题。

目前的图像特征匹配算法在建立 2 幅图像之间的图像匹配点对过程中, 会不可避免地存在误匹配点对。因此相对位姿估计算法通常与鲁棒估计方法一起使用, 例如随机抽样一致 (Random Sample Consensus, RANSAC)<sup>[2]</sup>, 进而剔除图像匹配点对中的误匹配点对并恢复出 2 幅图像之间的相对位姿。由于在相同的野值比例条件下, RANSAC 的随机采样次数会随着相对位姿估计最小配置解所需图像匹配点对数量的增加而呈指数型增长, 因此, 研究获得相对姿态估计的最小配置解显得非常重要。

在传统的位姿估计中, 一般不对相机运动做出约束, 即相机在旋转和平移 2 个维度上均有 3 个自由度。在进行单应矩阵估计时, 为了进一步提高运算的效率, 需要减少求解所需的点对数量。在不增加单个点对所提供的约束条件的情况下, 减少单应矩阵中未知数的个数就成为了可行的方法, 在一般的单应

\* 收稿日期: 2023-02-03; 修回日期: 2023-05-08

基金项目: 国家自然科学基金项目(11902349); 湖南省自然科学基金项目(2020JJ5645)

通信作者: 关棒磊(1991-), 男, 博士, 副研究员。主要从事摄像测量、计算机视觉等方面研究。Email: guanbanglei12@nudt.edu.cn

孙祥一(1964-), 男, 博士, 研究员。主要从事摄像测量、计算机视觉等方面研究。Email: sunxiangyi\_hn@163.com

估计中,需估计的归一化单应矩阵为一个 $3 \times 3$ 的矩阵,矩阵中的9个元素是存在8个未知量的<sup>[3]</sup>。为了减少单应矩阵中未知数的个数,需要结合自动驾驶中的运动模式对相机运动模式加以约束。

在自动驾驶领域,汽车在公路上行驶的过程中,可以近似地认为汽车是在一个平面上做运动,如图1所示。在此类平面运动过程中,相机的视野中存在着大量的位于同一平面上的特征点,因此可以通过平面单应性来对前后2帧的相对位姿进行求解,同时,纯平面运动对单应矩阵中元素的关系提供了约束,进行单应矩阵估计的过程中可以使用更少的点对,进而减少RANSAC算法的迭代次数,提高算法的运行效率。

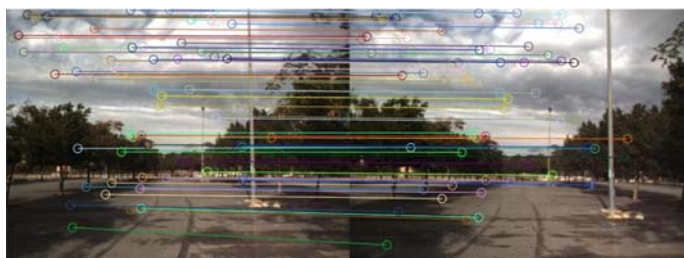


图1 车辆在公路上前后2帧实际视图及两视图间SIFT特征点(图片出自Málaga数据集<sup>[4]</sup>)

Fig.1 Actual view of the vehicle on the road between two consecutive frames and the SIFT feature points between the two views (images from the Málaga dataset<sup>[4]</sup>)

对于广泛存在的平面运动对单应矩阵给出的约束,本文提出了一种基于单个SIFT特征的相对位姿估计方法,主要贡献有:推导了基于平面运动情况下的平面单应矩阵的具体形式;基于推导出的单应矩阵,提出了平面运动情况下的相机相对位姿估计方法;通过仿真和真实实验证明了上述方法的有效性。

## 1 相关工作

减少RANSAC单次采样所需的点对个数是提高RANSAC计算效率的重要方法之一<sup>[5]</sup>。为了达到这个目的,需要寻找除常用的极线以外的其他约束方程。目前特征提取主要有SIFT<sup>[6]</sup>、ORB<sup>[7]</sup>、SURF<sup>[8]</sup>等方法。作为一种非常经典的特征描述子,SIFT特征<sup>[6]</sup>由点坐标、特征尺度和特征旋转3个部分组成,具有着平移、旋转、尺度和亮度上的不变性,是一种被广泛应用于各种场景下的图像特征描述子。

相机相对位姿经典的方法是基于基础矩阵的8pt方法<sup>[9]</sup>和基于本质矩阵的5pt方法<sup>[10]</sup>。前者利用2幅图像的8个匹配点对估计基础矩阵,进而分解得到相机位姿;后者利用2幅图像中的5个匹配点对估计本质矩阵,进而分解得到相机的位姿。针对平面场景,有学者提出4pt方法<sup>[11]</sup>,该方法利用2幅图像中的4个位于同一平面上的匹配点对,来估计平面的单应矩阵,进而恢复相机位姿。

目前汽车、无人机、手机等移动平台通常都配备了惯性测量单元(Inertial Measurement Unit, IMU)。IMU可以提供俯仰角、偏航角和滚转角。由于IMU测量的俯仰角和偏航角的精度比滚转角精度高,因此在实际使用过程中,通常仅使用IMU测量得到的俯仰角和偏航角参与计算,因此旋转上的自由度从3减少到了1。在5pt方法<sup>[10]</sup>的基础上,FRAUNDORFER<sup>[12]</sup>提出了一种使用IMU的2个角度与3个点对完成对相机的相对位姿的估算方法。在真实的环境中存在着大量的平面场景,例如:地面、墙面、门等,因此,有大量学者在估算相对位姿的时候也考虑平面信息。SAURER等<sup>[13]</sup>将平面分为水平、垂直平面和一般平面,并给出了具体计算方法。如果特征点对位于地面上,使用2个点就可以估算相机的相对位姿;如果特征点对位于垂直平面上,使用2.5个点就可以估算相机的相对位姿;如果特征点对位于一般平面上,使用3个点就可以估算相机的相对位姿。特别针对工业环境,工业机器人的运动被固定为一自由度平移和一自由度旋转条件下,ORTIN等<sup>[14]</sup>提出了一基于牛顿迭代的2点算法和线性3点算法。由于在用SIFT提取特征点的时候不仅提供了点坐标信息,同时还提供了特征旋转和特征尺度信息,因此可以为一个特征点对引入更多的约束方程。DING等<sup>[15]</sup>在已知平面单应性约束的

基础上加入了 SIFT 特征描述子的约束提出了最小求解器,与传统的方法相比需要更少的点数。

## 2 基于单个 SIFT 特征的相对位姿估计方法

车辆在公路上行驶,可以近似为平面运动。这种情况下,运动的自由度为 3,即 1 个旋转上的自由度,2 个平移上的自由度。1 个旋转上的自由度指的是旋转轴为运动平面法向量方向的旋转,2 个平移上的自由度是任意方向上的平移尺度自由度和该方向与指定坐标轴的夹角大小的角度自由度。由于本文使用的是单目相机,因此无法恢复位姿变化过程中的尺度信息<sup>[11]</sup>。至此,所要求解的未知量仅有旋转角和平移角。

接下来对单应矩阵的具体形式进行分析。

### 2.1 单应矩阵的运动建模

以相机前向为  $z$  轴,右向为  $x$  轴,相机中心为原点, $y$  轴方向通过右手系确定,建立相机坐标系<sup>[16]</sup>。相机坐标系如图 2 所示。

地面的单应矩阵  $\mathbf{H}$  可以表示为

$$\mathbf{H} = \mathbf{R} - \frac{\mathbf{t}}{d} \mathbf{N}^T \tag{1}$$

式中:  $\mathbf{R}$  为旋转矩阵;  $\mathbf{t}$  为在相机坐标系下的平移向量;  $\mathbf{N}$  为平面法向量; 上标 T 代表着对向量的转置;  $d$  为相机到平面的距离。定义在世界坐标系下的旋转矩阵  $\mathbf{R}$  和平移向量  $\mathbf{t}_0$  分别为

$$\mathbf{R} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \tag{2}$$

$$\mathbf{t}_0 = [\sin\alpha \ 0 \ \cos\alpha]^T \tag{3}$$

式中:  $\theta$  为旋转角;  $\alpha$  为平移角。旋转角和平移角的定义如图 3 所示。

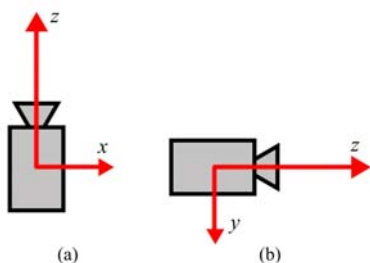


图 2 (a)俯视图下的相机坐标系;  
(b)左视图下的相机坐标系

Fig. 2 (a) camera coordinate system in top view;  
(b) camera coordinate system in left view

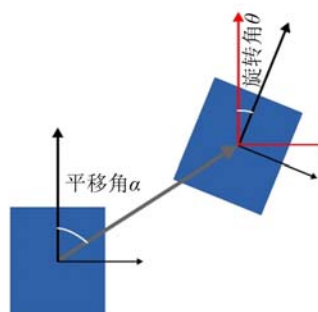


图 3 相机平面运动旋转角和平移角的定义

Fig. 3 Definition of rotation and translation angles for camera plane motion

式(3)中  $\mathbf{t}_0$  是在世界坐标系中的平移向量,为了将在世界坐标系下表示的平移向量  $\mathbf{t}_0$  转换到相机坐标系中进行表示,需要做如下变换:

$$\mathbf{t} = \mathbf{R}\mathbf{t}_0 \tag{4}$$

式中,  $\mathbf{t}$  为在相机坐标系下表示的平移向量。

平面法向量  $\mathbf{N}$  可以表示为

$$\mathbf{N} = [0 \ 1 \ 0]^T \tag{5}$$

至此式(1)单应矩阵表达式中所有可被相机旋转和平移表示的量均已获得,将式(2)至式(5)代入到式(1)中,就可以将  $\mathbf{H}$  表示为

$$\mathbf{H} = \begin{bmatrix} \cos\theta & \frac{1}{d}\sin(\theta-\alpha) & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & -\frac{1}{d}\cos(\theta-\alpha) & \cos\theta \end{bmatrix} \quad (6)$$

式中,  $d$  依旧是单应矩阵中的未知量, 但是并不需要对  $d$  进行求解, 利用图像中的匹配点对将单应矩阵估计出以后, 将  $h_2$  和  $h_8$  两项相除就可以消去  $d$ 。

## 2.2 单应矩阵的求解

对  $\mathbf{H}$  建立约束求解未知量。单应矩阵  $\mathbf{H}$  的点对应关系为

$$\mathbf{H}\mathbf{p}_1 \sim \mathbf{p}_2 \quad (7)$$

式中  $\mathbf{p}_1$  和  $\mathbf{p}_2$  分别是 2 幅图像中对应点的像素坐标, 可以表示为

$$\mathbf{p}_1 = [u_1 \quad v_1 \quad 1]^T \quad (8)$$

$$\mathbf{p}_2 = [u_2 \quad v_2 \quad 1]^T \quad (9)$$

为方便表述, 将式(6)的单应矩阵形式简化如下:

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \quad (10)$$

将式(8)和式(9)代入到式(7)中, 可得<sup>[11]</sup>

$$h_1 u_1 + h_2 v_1 + h_3 - u_2 u_1 h_7 - u_2 v_1 h_8 - u_2 h_9 = 0 \quad (11)$$

$$h_4 u_1 + h_5 v_1 + h_6 - v_2 u_1 h_7 - v_2 v_1 h_8 - v_2 h_9 = 0$$

为了能够对单应矩阵  $\mathbf{H}$  进行求解, 需要再引入一条与 SIFT 特征旋转与尺度有关的约束<sup>[3]</sup>:

$$-s_2 c_1 h_1 - s_1 s_2 h_2 + c_1 c_2 h_4 + s_1 c_2 h_5 + (u_2 s_2 c_1 - v_2 c_1 c_2) h_7 + (u_2 s_1 s_2 - v_2 s_1 c_2) h_8 = 0 \quad (12)$$

式中,  $s$  与  $c$  分别为特征旋转的正弦值与余弦值, 其下角标表示对应的图像。

将式(11)和式(12)联立可改写为  $\mathbf{G}\mathbf{h}_0 = 0$  的形式,  $\mathbf{G}$  为系数矩阵,  $\mathbf{h}_0 = [h_1, h_2, h_3, \dots, h_9]^T$  为单应矩阵中的未知元素组成的列向量。

考虑到式(6)中单应矩阵各元素的表达式, 可以将单应矩阵中各元素之间的关系表示如下:

$$h_4 = h_6 = 0, h_5 = 1, h_1 = h_9, h_3 = -h_7 \quad (13)$$

进而可以将方程组  $\mathbf{G}\mathbf{h}_0 = 0$  整理为只含有  $h_1, h_2, h_7, h_8$  未知量的非齐次方程组, 表示为  $\mathbf{M}\mathbf{h} = \mathbf{b}$ , 其中  $\mathbf{M}$  为  $3 \times 4$  的系数矩阵,  $\mathbf{b}$  为非齐次项,  $\mathbf{h} = [h_1, h_2, h_7, h_8]^T$  为未知数向量。

对于 4 个未知数的方程, 式(11)和式(12)一共给出了 3 个约束, 剩余约束条件由三角函数约束给出, 至此约束条件数量大于方程中未知数的数量, 可以参考 SAURER 等<sup>[13]</sup>求解方程组的方法对非齐次方程组进行求解。

对  $\mathbf{M}$  矩阵进行 SVD 分解:

$$\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (14)$$

式中:  $\mathbf{U}$ 、 $\mathbf{D}$ 、 $\mathbf{V}$  是矩阵  $\mathbf{M}$  做 SVD 分解后得到的 3 个矩阵, 其中  $\mathbf{U}$ 、 $\mathbf{V}$  为两正交矩阵,  $\mathbf{D}$  为对角矩阵, 主对角线值为  $\mathbf{M}$  矩阵的奇异值。

令  $\mathbf{v}$  为  $\mathbf{V}$  的最后一列的列向量, 那么解向量  $\mathbf{h}$  可以表示为

$$\mathbf{h} = \mathbf{V}\mathbf{y} + \lambda\mathbf{v} \quad (15)$$

式中:  $y_i = \frac{\mathbf{U}^T \mathbf{b}}{D_{ii}}$ ;  $\lambda$  为一个未知的比例系数, 可以利用三角函数约束关系进行确定<sup>[13]</sup>。

在  $\mathbf{h}$  的表达式中, 第 1 个元素  $h_1$  可以表示为

$$h_1 = (\mathbf{V}\mathbf{y})_{(1,1)} + \lambda v_1 \quad (16)$$

同理, 第 2 个元素  $h_2$  可以表示为

$$h_2 = (\mathbf{V}\mathbf{y})_{(2,1)} + \lambda v_2 \quad (17)$$

由于在解向量  $\mathbf{h}$  中,  $h_1$  代表旋转角的正弦值大小,  $h_2$  代表旋转角的余弦值大小, 根据三角函数关系  $h_1^2 + h_2^2 = 1$ , 将式(16)和式(17)代入三角函数关系中可得关于未知数  $\lambda$  的一元二次方程, 通过求解该一元二次方程, 就可以得到  $\lambda$  的 2 个解, 将解代回式(16)和式(17)可得  $\mathbf{h}$  的 2 个解。

在  $\mathbf{h}$  的 2 个解中, 旋转角相差  $180^\circ$ , 考虑到实际情况中 2 个相邻关键帧之间的实际转角不会非常大, 因此假设相机旋转在  $-90^\circ$  到  $90^\circ$  之间, 剔除掉另外一个值, 得到  $\mathbf{h}$  的解。最后将  $\mathbf{h}$  的值还原到单应矩阵  $\mathbf{H}$  上, 并做归一化得到单应矩阵  $\mathbf{H}$  的解。

为了消除图像中误匹配点的影响, 使用 RANSAC 算法<sup>[5]</sup>进行计算。由于通过约束减少了未知数的数量, 从而减少了位姿估计时需要的点对数量, 根据本节对单应矩阵的模型建立和求解过程的分析, 估计单应矩阵仅需要一对点对, 这也意味着使用 RANSAC 算法进行计算的时间大幅缩短, 从而提高了实时性。

### 3 实验

#### 3.1 仿真实验

在仿真实验中, 生成 50 个随机的地面点以及 50 个空间中随机平面上的点, 考虑到本文主要面向的自动驾驶领域中车辆是朝前开的, 因此在仿真实验中仅对相机的前向运动和随机运动进行测试, 侧向运动由于明显不符合实际情况故而不进行测试。在前向运动中, 令相机在相机光轴方向上运动。在随机运动中, 令相机在前向角为  $[-25^\circ, 25^\circ]$  的区间内进行随机运动。根据第 2 节中的描述, 求解本文所提出的模型仅需要一个 SIFT 特征点对, 因此将本文提出的方法简称为 1SIFT 方法。为评估 1SIFT 方法的性能, 需要与其他方法进行对比, 实验中所采取的对比方法分别为 2pt 方法<sup>[13]</sup>和 5pt 方法<sup>[10]</sup>。令图像噪声以 0.1pixel 的幅度从 0pixel 逐渐增加到 2pixel, 对比参与测试的各个算法之间的性能。由于 SIFT 特征旋转的误差对算法的精度也存在影响, 因此, 在 1SIFT 方法的仿真实验中, 额外引入一个特征旋转误差。定义 6 个特征旋转误差等级, 分别为  $0^\circ$ 、 $0.2^\circ$ 、 $0.4^\circ$ 、 $0.6^\circ$ 、 $0.8^\circ$  和  $1^\circ$ , 在同一特征旋转误差等级下, 对不同点对的特征旋转误差额外乘上一个标准正态分布随机数。为方便表述, 上述在 6 个特征旋转误差等级下的 1SIFT 方法与参与对比的 2pt 方法与 5pt 方法, 分别称为: 1SIFT- $0^\circ$ 、1SIFT- $0.2^\circ$ 、1SIFT- $0.4^\circ$ 、1SIFT- $0.6^\circ$ 、1SIFT- $0.8^\circ$ 、1SIFT- $1^\circ$ 、2pt 和 5pt 方法, 在同一像素误差等级下, 同一个算法运行 1000 次, 取 1000 次的中值作为结果。

为评估算法的精度, 需要将各算法的估计位姿分别与仿真环境中位姿真值进行对比, 得到旋转误差和平移误差。将旋转误差定义如下<sup>[17]</sup>:

$$E_r = \arccos \left\{ \frac{\text{tr}(\mathbf{R}_{gr}^T \mathbf{R}_{estimated}) - 1}{2} \right\} \quad (18)$$

式中:  $E_r$  为旋转误差;  $\mathbf{R}_{gr}$  为旋转矩阵真值;  $\mathbf{R}_{estimated}$  为估计得到的旋转矩阵。

由于 1SIFT 方法使用单目相机, 无法得到尺度信息, 因此在误差评估过程中不考虑平移的尺度, 仅对平移的角度误差进行评估, 平移误差定义如下<sup>[17]</sup>:

$$E_t = \arccos \left\{ \frac{\mathbf{t}_{gr}^T \mathbf{t}_{estimated}}{\|\mathbf{t}_{gr}\| \|\mathbf{t}_{estimated}\|} \right\} \quad (19)$$

式中:  $E_t$  为平移角度误差;  $\mathbf{t}_{gr}$  为平移向量真值;  $\mathbf{t}_{estimated}$  为估计得到的平移向量。

经过仿真环境中的运算与对比, 实验结果如图 4 与图 5 所示。

根据图 4 与图 5 中的实验结果不难发现, 所提出的 1SIFT 方法在前向运动与随机运动的情况下, 全噪声等级均取得了比 2pt 方法与 5pt 方法更加优秀的精度。在低噪声等级的情况下, SIFT 特征旋转误差的提高会显著降低算法的精度, 但随着图像噪声等级的增加, 影响精度的主要因素从特征旋转误差变化为图像噪声后, 所定义的 6 个特征旋转误差等级下, 1SIFT 方法并不会因为特征旋转误差的变化使得算法精度出现显著的变化。

由于 1SIFT 方法是使用单应矩阵进行计算的, 因此进一步对环境非平面性影响进行了评估。生成 50 个随机的地面点以及 50 个空间中随机平面上的点, 仅对相机的随机运动进行测试, 将图像噪声固

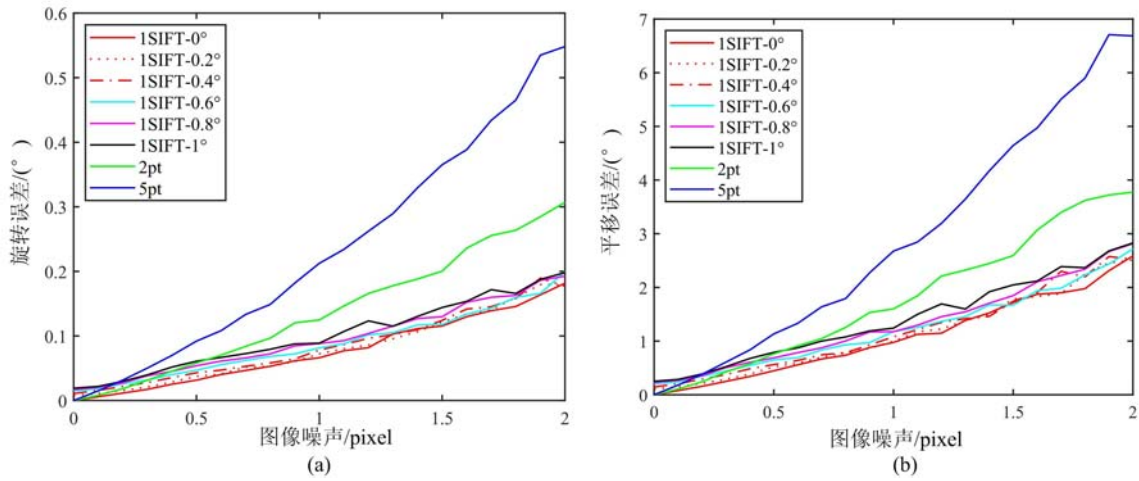


图 4 (a)前向运动情况下旋转误差随图像噪声等级的变化情况；

(b)前向运动情况下平移误差随图像噪声等级的变化情况

Fig. 4 (a)the change of rotation error with image noise level under forward motion;

(b)the change in translation error with image noise level under forward motion

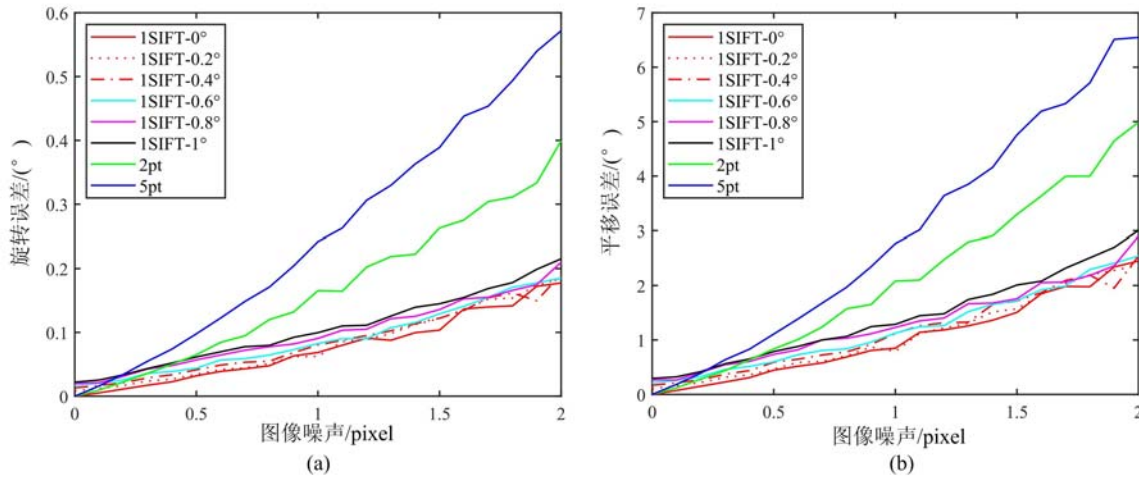


图 5 (a)随机运动情况下旋转误差随图像噪声等级的变化情况；

(b)随机运动情况下平移误差随图像噪声等级的变化情况

Fig. 5 (a)the change of rotation error with image noise level under random motion;

(b)the change in translation error with image noise level under random motion

定为 2pixel, SIFT 特征旋转误差固定为 0。定义离面度为场景中不在地面上的特征点与所有特征点的比例,令地面点以 10 个的幅度从 100 个到 0 个递减,即观察场景特征点的离面度以 0.1 的幅度从 1 逐步减少到 0,评估旋转误差与平移误差的方法与旋转与平移误差对比实验中一致。在仿真环境中评估 1SIFT 方法位姿估计的精度,结果如图 6 所示。

由图 6 可知,在离面度到达 0.8 之前,旋转误差与平移误差均随着离面度的提升有小幅度的增加。但是当离面度到达 0.8 之后,由于场景整体很难找到一个平面来得到精度较高的单应矩阵,旋转误差和平移误差均出现了极大的增加,因此当观察场景为复杂非平面的情况时,所提出的方法会因为很难得到平面而使得位姿估计的精度变得非常差。

进一步对算法的数值稳定性进行分析。

定义数值稳定性的评价标准为<sup>[18]</sup>

$$\xi_R = \| \mathbf{R}_{gt} - \mathbf{R}_{estimated} \|_F \tag{20}$$



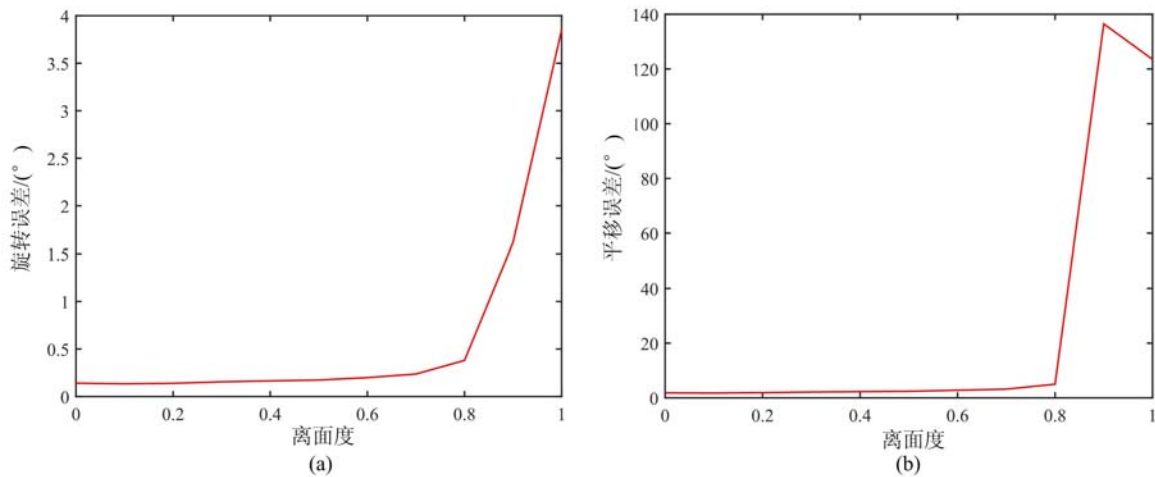


图 6 (a)1SIFT 方法的旋转误差随观察场景离面度的变化情况；  
(b)1SIFT 方法的平移误差随观察场景离面度的变化情况

Fig. 6 (a) the rotation error of the 1SIFT method changes with the degree of separation of the observation scene;  
(b) the change of the translation error of the 1SIFT method with the degree of separation of the observation scene

$$\xi_i = \left\| \frac{\mathbf{t}_{gt}}{\|\mathbf{t}_{gt}\|} - \frac{\mathbf{t}_{estimated}}{\|\mathbf{t}_{estimated}\|} \right\|_2 \quad (21)$$

式中,  $\xi_R$  和  $\xi_t$  分别为评估旋转和平移数值稳定性的指标。

在仿真环境中, 将像素噪声和 SIFT 特征旋转噪声均设置为 0, 将算法重复运行 1000 次, 分别记录下旋转误差数值稳定性  $\xi_R$  和平移误差数值稳定性  $\xi_t$ , 绘制为正态分布的概率密度曲线, 对比方法为 2pt 和 5pt 方法, 结果如图 7 所示。

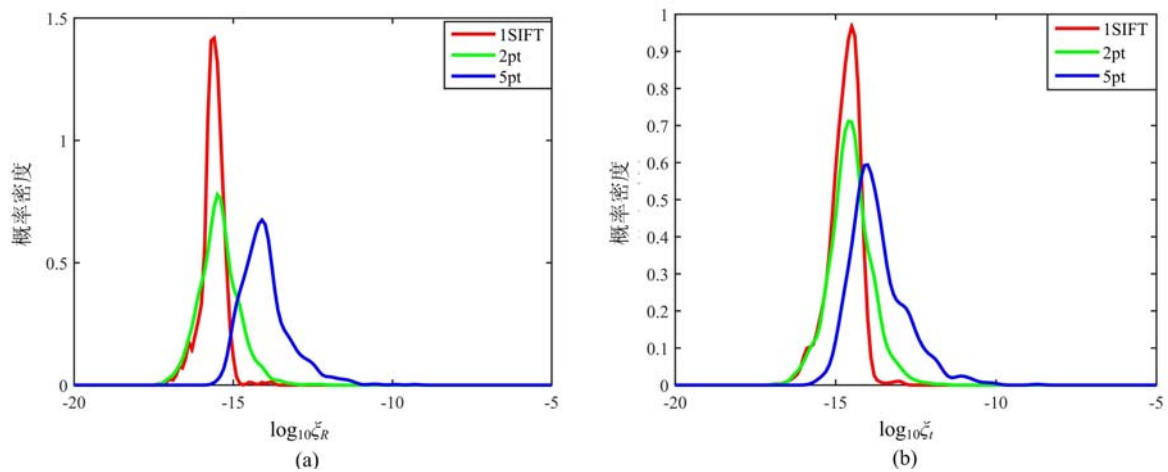


图 7 (a)3 种对比方法的旋转误差数值稳定性分析结果;(b)3 种对比方法的平移误差数值稳定性分析结果

Fig. 7 (a) numerical stability analysis results of rotation error of three comparison methods;  
(b) numerical stability analysis results of translation error of three comparison methods

图 7 中, 横轴代表旋转与平移误差以 10 为底的对数值, 纵轴代表对应不同误差值的概率密度。通过对数值稳定性的分析对比, 可以发现所提出的方法最大概率密度处的误差值与 2pt 方法相仿, 优于 5pt 方法, 但是其峰值远高于 2pt 方法与 5pt 方法, 证明了所提出方法具有更为良好的数值稳定性。

由于 1SIFT 算法中使用了 RANSAC 算法, 因此减少计算过程中使用点对的数量理论上可以有效降低 RANSAC 算法的迭代次数, 进而提高算法执行的效率。RANSAC 算法迭代次数的公式为<sup>[5]</sup>

$$N = \frac{\ln(1 - p)}{\ln(1 - (1 - v)^m)} \quad (22)$$

式中:  $N$  为计算出的 RANSAC 算法迭代次数;  $v$  为 RANSAC 算法中输入数据的外点所占的比例;  $m$  为求解所需的点的数量;  $p$  为使用 RANSAC 方法后可以排除所有外点的概率。根据式(22), 在  $p$  和  $v$  保持不变的情况下, 如图 8 所示, 随着求解所需点的数量  $m$  的增加, 迭代次数  $N$  会大幅度增加。

为了定量地分析 RANSAC 迭代次数的变化情况, 对所提出的 1SIFT 方法与 2pt 和 5pt 方法进行对比实验。由于 2pt 算法与 5pt 算法精度不受 SIFT 特征旋转的影响, 为保证对比实验的公平性, 实验中不引入 SIFT 特征旋转误差, 仅引入图像噪声误差。使仿真环境下的图像噪声从 0 按 0.1pixel 的幅度逐步增加到 2pixel, 对比在同一噪声等级下不同方法的 RANSAC 迭代次数, 3 个方法中 RANSAC 迭代的阈值与迭代次数上限均一致, 实验结果如图 9 所示。

图 9 中横轴为实验中所设定的 21 个误差等级, 纵轴是由指数形式表示的 RANSAC 迭代次数。通过对比实验发现, 所提出的 1SIFT 方法, 由于将求解所需的点对数量降低为 1, 因此大幅度地减少了 RANSAC 的迭代次数, 对比 2pt 方法的迭代次数减少约一个数量级。图像噪声大于 0.4pixel 后, 5pt 方法的 RANSAC 迭代次数停止上升的原因在于实验中所设置的 RANSAC 迭代次数上限为 10000, 当噪声等级大于 0.4pixel 时, 5pt 方法运行 RANSAC 算法会达到设定的迭代次数上限进而停止迭代, 因此 5pt 方法后续的曲线没有继续上升。

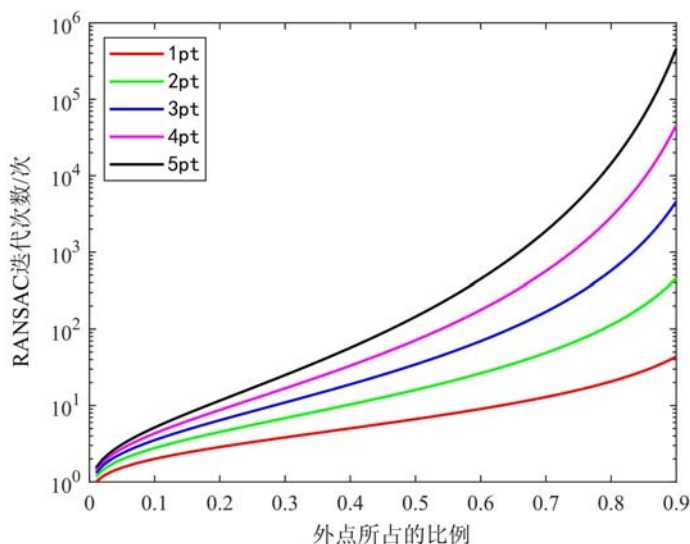


图 8 99% 概率排除所有外点条件下的 RANSAC 算法迭代次数随外点比例与求解所需点对个数的变化情况

Fig. 8 The 99% probability excludes the change of the number of iterations of the RANSAC algorithm with the proportion of the exterior points and the number of point pairs required for the solution

### 3.2 真实实验

在真实实验中, 使用真实场景的数据集数据来进行算法的评估分析。由于在仿真实验中发现所提出的方法非常不适用于非平面场景, 因此在真实实验中采用 Málaga 数据集<sup>[4]</sup>用于算法的评估。Málaga 数据集使用双目相机、激光雷达、IMU、GPS 等对城市道路情况进行测量拍摄, 由于采集数据的地点相对比较开阔, 因此比较符合所提出的算法对观测场景平面度的要求。

考虑到 Málaga 数据集并没有给出相机的位姿真值, 实验中参考 DING 等<sup>[15]</sup>的论文中所采用的实验配置, 使用车上载有的 GPS 和 IMU 数据计算出相对可靠的真值。将数据集中给出的 GPS 位置简单地连接起来, 通过线性插值得到真值信息。考虑到相机的采集帧率为 20Hz, 而 GPS 的采集帧率仅有 1Hz<sup>[4]</sup>, 为了降低插值误差影响和相机小幅度运动对误差的放大作用, 每隔 10 张图片定义 1 张关键帧, 仅使用关键帧的图片进行位姿估计。对于拍摄于 GPS 数据前的图片, 全部弃之不用。由于数据集中还存在 GPS 数据与相机的卫星时间戳没有对齐的情况, 因此在插值过程中, 插值函数的自变量设置为卫星时间。对比方法同样为 2pt<sup>[15]</sup>和 5pt<sup>[10]</sup>方法, 为便于展示, 将同一序列下 3 个方法中最优的运行结果进行加粗, 对比结果见表 1。



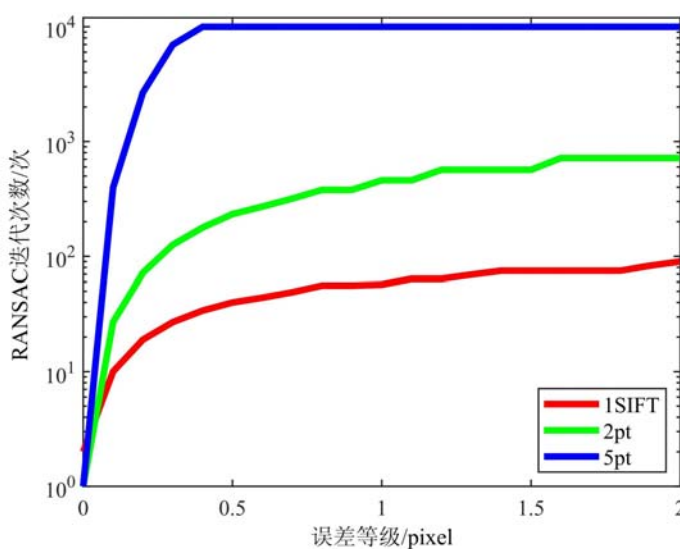


图 9 3 种方法在不同噪声等级下的 RANSAC 迭代次数变化情况

Fig. 9 Changes in the number of RANSAC iterations under different noise levels of the three methods

表 1 3 种方法在 Málaga 数据集 15 个序列下的相对旋转误差对比(单位:°)

Tab.1 Comparison of the relative rotational errors of the three methods for the Málaga dataset with 15 sequences(unit: °)

方法	序列 1	序列 2	序列 3	序列 4	序列 5	序列 6	序列 7	序列 8
1SIFT	0.7425	<b>1.0110</b>	<b>1.1878</b>	<b>2.1955</b>	<b>0.8363</b>	<b>0.6638</b>	<b>0.8998</b>	<b>1.0234</b>
2pt <sup>[15]</sup>	0.8945	1.5734	3.1298	5.2850	1.3079	1.2883	1.7493	1.7360
5pt <sup>[10]</sup>	<b>0.6694</b>	1.4491	2.9607	7.9199	0.9950	2.7341	1.4613	2.0747
方法	序列 9	序列 10	序列 11	序列 12	序列 13	序列 14	序列 15	
1SIFT	<b>1.1351</b>	<b>1.8259</b>	<b>0.6506</b>	<b>0.7701</b>	<b>0.9243</b>	<b>0.5017</b>	<b>1.3138</b>	
2pt <sup>[15]</sup>	1.6666	3.2160	0.9539	1.1709	1.8426	1.3363	2.0725	
5pt <sup>[10]</sup>	1.3807	4.8171	0.8042	0.9399	3.4784	5.9957	10.1376	

如表 1 所示,在 Málaga 数据集上共 15 个序列中分别对 3 个方法进行了运行,分别获得 3 个方法估计出的旋转角度,通过与使用 GPS 数据计算得到的真值进行对比,可以得到旋转角度误差,旋转误差的具体计算方法与仿真实验部分一致。通过对旋转误差对比实验的结果进行分析,可以发现:在大多数情况下,1SIFT 方法相比其他 2 种对比方法可以取得更为优秀的旋转精度。

图 10 和图 11 是使用 1SIFT 方法计算前的 SIFT 点匹配情况和使用 1SIFT 方法计算后的内点匹配情况。由于在计算 RANSAC 模型距离这一部分,使用了算法估计得到的旋转矩阵和平移向量恢复出本质矩阵,通过本质矩阵信息计算模型距离,而并没有使用单应矩阵计算模型距离,这使得 1SIFT 方法的内点并不仅限于地面点,对于真实的空间环境具有更好的适应性。通过前后匹配情况的对比,可以发现 1SIFT 方法可以有效地剔除掉错误的 SIFT 匹配点。

图 12 是 3 种方法在 Málaga 数据集第 10 序列和第 13 序列运行结果的 CDF 曲线,曲线越靠上,则说明其结果越集中于小误差的部分。为了保证曲线的平滑度和可靠度,需要选取包含图片数据更多的序列,因此选择了序列 10(17310 张图片)和序列 13(31442 张图片)2 个序列用来进行 CDF 的评估。通过对图 12 中曲线的分析,可以发现,所提出的 1SIFT 方法略好于 2pt 方法<sup>[13]</sup>和 5pt 方法<sup>[10]</sup>。

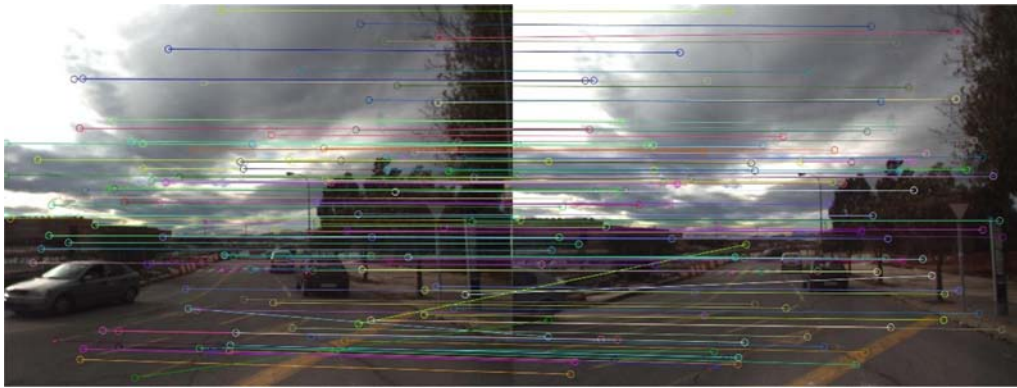


图 10 相邻 2 帧间的 SIFT 点匹配情况

Fig. 10 SIFT point matching between two adjacent frames

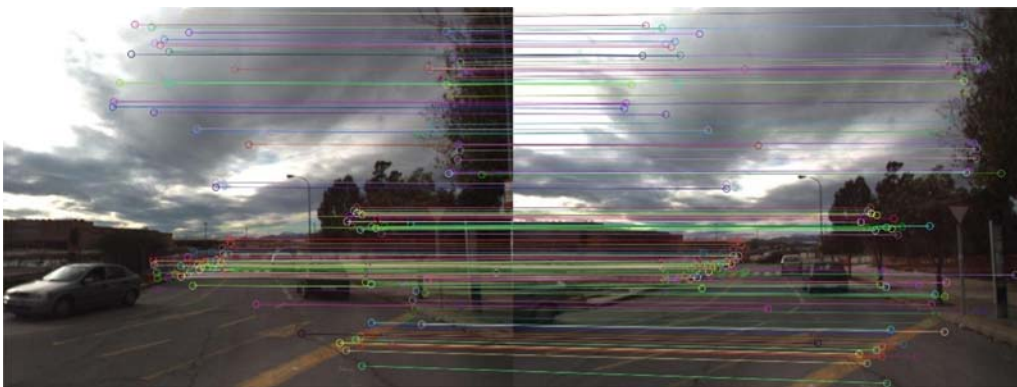


图 11 经过 1SIFT 方法处理后的内点匹配情况

Fig. 11 Inlier point matching after 1SIFT method processing

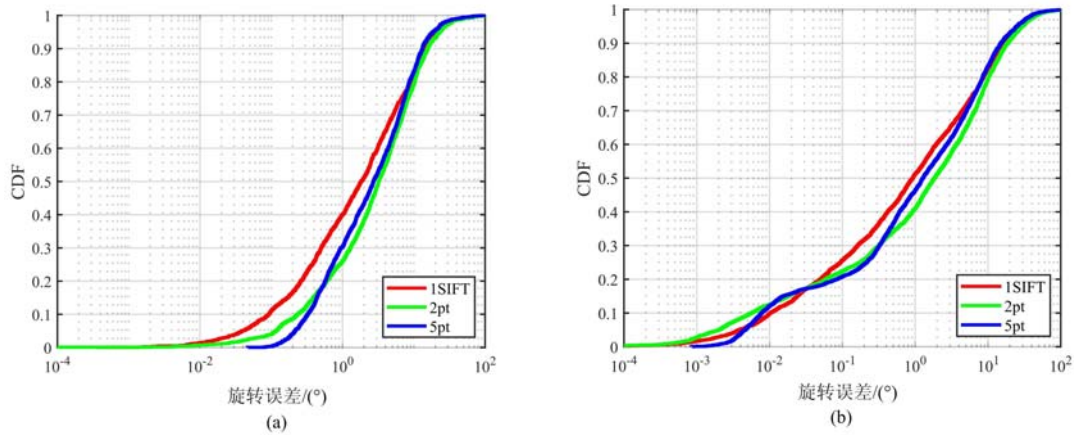


图 12 (a)3 种参与对比方法的旋转误差 CDF 曲线情况(数据来自 Málaga 数据集第 10 序列);

(b)3 种参与对比方法的旋转误差 CDF 曲线情况(数据来自 Málaga 数据集第 13 序列)

Fig. 12 (a) rotation error CDF curves of the three participating comparison methods (data from series 10 of the Málaga dataset); (b) rotation error CDF curves of the three participating comparison methods (data from sequence 13 of the Málaga dataset)

### 4 结论

本文提出了一种基于单个 SIFT 特征的平面运动相对位姿估计方法。该方法基于自动驾驶和精密光测的实际情况,将实际运动情况近似为平面运动,并将平面运动的 2 个自由度的平移表述为 1 个自由度的平移角和 1 个自由度的尺度,由于单目相机无法恢复尺度,因此仅对 1 个自由度的旋转角和 1 个自

由度的平移角进行求解,同时引入了 SIFT 特征用于增加单个点对中所包含的约束数量,将平面位姿估计所需要的点对数量减少为 1,降低了 RANSAC 算法的迭代次数,提高了效率,并且在仿真和真实实验中分别与其他方法进行了对比和测试,结果表明,所提出的方法是相对稳定可靠的。

#### 参考文献:

- [1] 王刚, 郭鹏宇, 尚洋, 等. 平面镜反射辅助视场下相机位姿估计问题的 Cramer-Rao 下界[J]. 实验力学, 2016, 31(1):39-46 (WANG Gang, GUO Pengyu, SHANG Yang, et al. On the Cramer-Rao lower bound of camera pose estimation using planner mirror reflection for expanding the camera's FOV[J]. Journal of Experimental Mechanics, 2016, 31(1):39-46 (in Chinese))
- [2] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6):381-395.
- [3] BARATH D, KUKELOVA Z. Homography from two orientation- and scale- covariant features[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019:1091-1099.
- [4] BLANCO-CLARACO J L, MORENO-DUENAS F A, GONZÁLEZ-JIMÉNEZ J. The Málaga urban dataset: high-rate stereo and LiDAR in a realistic urban scenario[J]. The International Journal of Robotics Research, 2014, 33(2):207-214.
- [5] DERPANIS K G. Overview of the RANSAC algorithm[J]. Image Rochester NY, 2010, 4(1):2-3.
- [6] NG P C, HENIKOFF S. SIFT: predicting amino acid changes that affect protein function[J]. Nucleic Acids Research, 2003, 31(13):3812-3814.
- [7] RUBLEE E, RABAU D V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision, 2011:2564-2571.
- [8] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[C]//Computer Vision-ECCV 2006: 9th European Conference on Computer Vision, 2006: 404-417.
- [9] HARTLEY R I. In defense of the eight-point algorithm[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(6):580-593.
- [10] NISTÉR D. An efficient solution to the five-point relative pose problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6):756-770.
- [11] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision[M]. Cambridge: Cambridge University Press, 2003.
- [12] FRAUNDORFER F, TANSKANEN P, POLLEFEYS M. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles[C]//European Conference on Computer Vision, 2010:269-282.
- [13] SAURER O, VASSEUR P, BOUTTEAU R, et al. Homography based egomotion estimation with a common direction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2):327-341.
- [14] ORTIN D, MONTIEL J M M. Indoor robot motion based on monocular images[J]. Robotica, 2001, 19(3):331-342.
- [15] DING Y, BARATH D, KUKELOVA Z. Homography-based egomotion estimation using gravity and sift features [C]//Proceedings of the Asian Conference on Computer Vision, 2020.
- [16] 叶雪娟, 甘叔玮, 张小虎, 等. 绳索摆角高精度视觉测量系统设计及其在悬吊式重力补偿系统中的应用[J]. 实验力学, 2021, 36(6):735-745 (YE Xuezhou, GAN Shuwei, ZHANG Xiaohu, et al. Design of high-precision rope swing angle visual measurement system and its application in suspended gravity compensation system[J]. Journal of Experimental Mechanics, 2021, 36(6):735-745 (in Chinese))
- [17] DING Y, YANG J, PONCE J, et al. An efficient solution to the homography-based relative pose problem with a common reference direction[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 1655-1664.
- [18] DING Y, YANG J, KONG H. An efficient solution to the relative pose estimation with a common direction[C]//2020 IEEE International Conference on Robotics and Automation (ICRA), 2020:11053-11059.

## Relative pose estimation method based on single SIFT features

TAN Ze, GUAN Banglei, SUN Xiangyi

(Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, College of Aerospace Science and Engineering,  
National University of Defense Technology, Changsha 410073, Hunan, China)

**Abstract:** Position estimation is one of the fundamental problems in precision optical measurement and autonomous driving. For practical applications such as autonomous driving, where the camera moves on a plane and the degrees of freedom of the camera position are three, this paper proposes a camera relative pose estimation method based on a single SIFT feature. Since the monocular camera cannot recover the translation scale, the degrees of freedom of the camera motion are reduced to two degrees of freedom with only rotation and translation angles. By observing the ground, ground homography information containing camera motion and plane normal vectors can be obtained. Therefore, camera motion can be restored by extracting homonymous ground points to estimate the homography matrix. In order to reduce the number of RANSAC interactions and improve the efficiency of the algorithm, SIFT features are introduced to the pose estimation, which include the coordinates of the homonymous points in the two images, as well as their feature rotations and feature scales. So that the information contained in a single point pair can be expanded, and the number of point pairs required for solving the homography matrix can be reduced efficiently. For the case of planar two-degree-of-freedom motion, this paper uses a single SIFT feature point pair to complete the estimation of homography matrix, and then uses Random Sample Consensus algorithm to optimize the results, and finally decomposes the homography matrix to obtain the relative position estimation results. The proposed method is proved to be effective by comparing it with the 2pt method and 5pt method in simulation and real experiments.

**Keywords:** pose estimation; RANSAC; autonomous driving; planar motion; optical measurement