

# 基于环形特征匹配的双目视觉里程计

黄平, 曹镇\*, 王欢

哈尔滨工程大学智能科学与工程学院, 黑龙江 哈尔滨 150001

**摘要** 针对视觉里程计中采用基本特征点匹配算法易出现误匹配现象, 提出一种结合双向光流法的特征点环形匹配算法。该算法将双目图像与前后帧图像组成环形, 前后帧图像采用双向金字塔光流法跟踪特征点, 并剔除错误的特征点追踪结果。基本特征匹配算法通常采用快速最近邻匹配 (FLANN) 算法, 但其结果包含很多误匹配的点对。本文提出的匹配算法不仅能剔除误匹配特征点, 还能使特征点均匀分布于图像上。随后采用基于 3D-2D 点的 P3P (Perspective-3-Point) 算法结合 RANSAC (Random Sample Consensus) 得到初始位姿估计结果, 再利用 g2o (General Graph Optimization) 库对位姿估计结果进一步优化。定位实验结果表明, 结合双向光流法的特征点环形匹配算法具有更高的定位精度。

**关键词** 机器视觉; 视觉里程计; 环形特征匹配; 位姿估计

中图分类号 TP391

文献标志码 A

doi: 10.3788/AOS202141.1515002

## Stereo Visual Odometry Based on Ring Feature Matching

Huang Ping, Cao Zhen\*, Wang Huan

College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, Heilongjiang 150001, China

**Abstract** In this paper, regarding the phenomenon that the basic feature point matching algorithm is prone to mismatch in visual odometry, we proposes a ring matching algorithm of feature points combined with the bidirectional optical flow method. This algorithm forms a ring structure between the stereo image and the images in the front and rear frames. For the images in the front and rear frames, the bidirectional pyramid optical flow method is used to track feature points and eliminate mismatched feature points. The basic feature point matching algorithm usually adopts fast library for approximate nearest neighbors (FLANN), but the result contains many mismatched point pairs. The proposed matching algorithm can not only eliminate the mismatched feature points but also make the feature points evenly distributed on the images. Subsequently, the perspective-3-point (P3P) algorithm based on 3D-2D points is combined with random sample consensus (RANSAC) to obtain the initial pose estimation results. The general graph optimization (g2o) library is employed to further optimize the pose estimation results. The positioning experiments verify that the ring matching algorithm of feature points combined with the bidirectional optical flow method has higher positioning accuracy.

**Key words** machine vision; visual odometry; ring feature matching; pose estimation

**OCIS codes** 150.6910; 280.4788; 330.1880

## 1 引言

近年来人们注意到, 随着机器人自主程度的提

高, 机器人越来越贴近人们的生活, 其应用范围不断扩大, 而提高机器人自主性的首要任务是提高机器人的定位精度。因此, 学者们掀起了一股以视觉为

收稿日期: 2021-01-09; 修回日期: 2021-02-20; 录用日期: 2021-03-09

基金项目: 国家自然科学基金(61871143)、黑龙江省自然科学基金(LH2019F006)、哈尔滨市应用技术研究与开发项目(2017R-AQXJ095)

通信作者: \*75306846@qq.com

信息来源的视觉导航技术热潮。视觉里程计(VO)是视觉 SLAM 的前端技术<sup>[1]</sup>,近年来发展迅速,已经得到了广泛的研究和应用<sup>[2-6]</sup>。

要实现相邻帧信息之间的关联,需要提取图像的特征信息并进行特征匹配。在双目视觉里程计中,帧间的特征匹配往往被称为特征跟踪,左右图像之间的特征匹配被称为双目匹配,他们的实质都是将特征点一一对应,为了统一,将其统称为特征匹配<sup>[7]</sup>。在基于特征点的双目 VO 中双目相机仅仅用于深度计算,比如在具有里程碑意义的火星漫游车 Mars 项目中<sup>[8]</sup>,采用双目技术进行左右图像特征匹配时,仅将左目图像用于三角化测量,在检测和跟踪方面并没有充分利用双目图像信息,故其在匹配过程中易出现特征误匹配问题。Geiger 等<sup>[9]</sup>使用符合“环形”匹配模式的正确特征点来验证跟踪情况,即左右图像在任何两个连续帧之间进行特征匹配,从当前帧左图像中的所有特征点开始检测,特征点以循环方式匹配前一帧左图像、前一帧右图像、当前帧右图像,最后返回到当前帧左图像。如果经过此类环形匹配后左图像特征点不一致,则拒绝匹配。Badino 等<sup>[10]</sup>提出了一种多帧特征集成技术,该技术通过跟踪特征的整个轨迹来计算相机运动,构建了一个特征集合,该集合包含所有图像帧中所有跟踪特征的累积信息,可以减少估计运动的漂移量。Chien 等<sup>[11]</sup>将经典的 VO 问题扩展到更高的维度,提出多相机多帧立体平行追踪拓扑,为右眼引入独立追踪路径。该拓扑结构保持左右图像的并行特征,并通过左右匹配将它们桥接集成在一起,这使得两个摄像头的图像数据都得到了充分利用。实验结果证明双目追踪比单目追踪的效果要好。

视觉里程计的主要任务是根据相邻图像的信息估计相机运动。本文的视觉里程计是基于特征点来

估计的,因此特征点的选取将直接影响定位精度。为了解决特征跟踪易出现的特征点误匹配问题,本文充分利用双目相机多帧多图像的特点,提出一种结合双向光流法的特征点环形匹配算法,该算法不仅能剔除误匹配特征点,保留不易受环境影响的特征点,并且能使特征点均匀分布于图像上。采用基于 3D-2D 点的 P3P 算法结合 RANSAC(Random Sample Consensus)<sup>[12]</sup>算法去除大误差外点,得到初始位姿估计结果,再利用 g2o 库对位姿估计结果进一步优化。最后通过数据集进行算法实验,实验结果表明,相比于基本特征点匹配算法,采用结合双向光流法的特征点环形匹配算法有效减少了误匹配特征点,提高了定位精度。

## 2 系统架构

本文提出的结合双向光流法的特征点环形匹配算法的整体框架如图 1 所示,即获得双目图像后,建立双目图像的图像金字塔,采用 ORB(Oriented FAST and Rotated BRIEF)特征检测算法检测图像金字塔中的特征点<sup>[13]</sup>,并采用金字塔光流法<sup>[14]</sup>跟踪特征点。利用本文提出的一种结合双向光流法的双目环形特征点匹配算法,将双目视觉里程计的特征匹配问题扩展到多帧多相机的情况,在该情况下,一组特征点在前后帧及左右图像之间通过光流计算被不断跟踪和管理,如果某一个特征点光流追踪失败,则表明该点质量较差,需将其从点集中剔除。此外,为了保证光流追踪过程中点集数量的稳定,每次删除质量较差的特征点后补充新的特征点。为保留下来的每个特征点设置一个圆形遮罩,以阻止 ORB 特征检测算法在已保留的特征点附近重复生成特征点,尽可能确保特征点均匀分布。本文方法充分利用双目相机的优势,计算各点的深度信息,得到空间

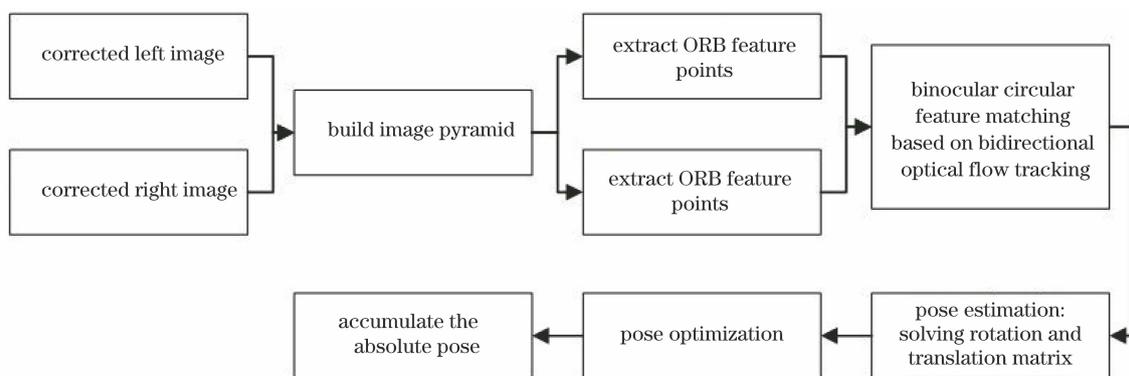


图 1 算法架构

Fig. 1 Algorithm architecture

坐标,最后估计载体位姿并优化位姿结果。

### 3 金字塔光流跟踪算法

经典的 L-K 光流法由于受到时间连续、空间一致和亮度守恒三个假设条件的制约,实际环境常常无法满足假设条件,易受到环境的影响<sup>[14]</sup>。构建图像金字塔这种方法,能提升对图像信息的拆解能力,便于感知图像信息更细微的变化。即使在某些运动速度过快的场景中,金字塔光流法依然能很好地跟踪特征点。由于本文在特征提取之前已经进行了图像金字塔构建操作,因此,选择金字塔光流法跟踪特征点进行算法实验。金字塔光流跟踪和 L-K 光流

跟踪的目的是一致的,首先计算出前一帧图像  $I$  像素点  $P$  的运动速度,也就是所说的像素点  $P$  的光流,然后在后一帧图像  $J$  的某个子区域上估计与之对应的像素点  $P$  的位置,从而实现特征跟踪。

在计算图像光流矢量时,采用图像金字塔的方法,可以避免图像中像素偏移矢量较大时,L-K 光流法中小运动假设不成立的情况。图像金字塔光流法跟踪算法处理小运动假设的主要思想是:将光流计算拆分至图像的每一层,金字塔光流跟踪过程由高层图像向相邻下一层方向进行,再将每一层中较小的光流矢量叠加,即可得到原始图像的光流值。计算过程如图 2 所示。

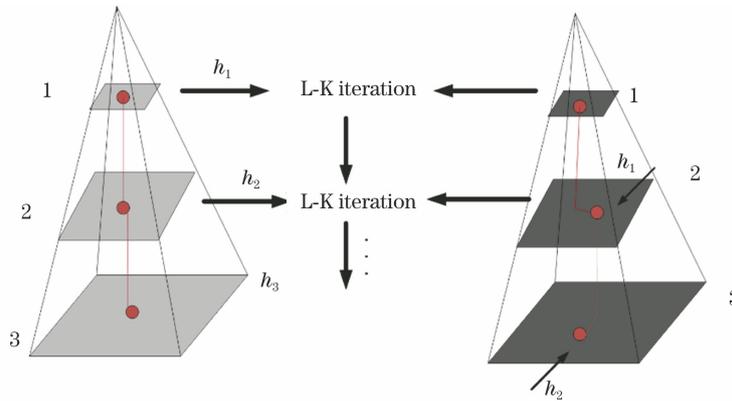


图 2 金字塔光流法

Fig. 2 Pyramid optical flow method

建立图像金字塔。设置图像  $I$  和图像  $J$  的金字塔为  $\{I_L, L=0, 1, \dots, m\}, \{J_L, L=0, 1, \dots, m\}$ 。定义图像中某像素点在第  $L$  层上的像素坐标为  $\mathbf{u}^{(L)} = (u_x^{(L)}, v_y^{(L)})$ ,按照金字塔计算方式,第  $L$  层的坐标值为

$$\mathbf{u}^{(L)} = \frac{\mathbf{u}}{2^L}, \quad (1)$$

式中:  $\mathbf{u}$  表示原始图像层像素点坐标;  $\mathbf{u}^{(L)}$  表示第  $L$  层上的像素坐标。

光流值会随着金字塔图像的缩小而减小。为了计算方便,最底层的光流  $\mathbf{g}^{(m)}$  初始值设为 0,即

$$\mathbf{g}^{(m)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (2)$$

假设  $L$  层的光流  $\mathbf{g}^{(L)}$  初始值为

$$\mathbf{g}^{(L)} = [g_x^{(L)}, g_y^{(L)}]^T. \quad (3)$$

取一个大小为  $(2w_x + 1, 2w_y + 1)$  的跟踪窗口,其中  $w_x, w_y$  为跟踪窗口在  $x, y$  轴方向的尺度基数,经典取值范围为  $1 \sim 7$ (整数),计算  $L$  层上的匹配误差  $\epsilon^{(L)}$ ,表达式为

$$\begin{aligned} \epsilon^{(L)}(\mathbf{d}^{(L)}, \mathbf{A}^{(L)}) &= \\ \epsilon^{(L)}(d_x^{(L)}, d_y^{(L)}, d_{xx}^{(L)}, d_{xy}^{(L)}, d_{yx}^{(L)}, d_{yy}^{(L)}) &= \\ \sum_{x=-w_x}^{w_x} \sum_{y=-w_y}^{w_y} [I_{\text{comp}}^{(L)}(x, y) - J_{\text{comp}}^{(L)}(x + dx, y + dy)]^2, \end{aligned} \quad (4)$$

式中:  $\mathbf{d}^{(L)}$  表示  $L$  层上的运动光流,  $\mathbf{A}^{(L)}$  表示  $L$  层上的仿射变换矩阵,  $I_{\text{comp}}^{(L)}$  和  $J_{\text{comp}}^{(L)}$  表示当前帧和下一帧原始图像  $I, J$  的  $L$  层采样图像。

$I_{\text{comp}}^{(L)}$  和  $J_{\text{comp}}^{(L)}$  可由这一层的光流和仿射变换矩阵初值  $\mathbf{g}^{(L)}$  和  $\mathbf{G}^{(L)}$  计算得到,计算方式为

$$\begin{cases} I_{\text{comp}}^{(L)}(x, y) = I^{(L)}(\mathbf{u}^{(L)}) \\ J_{\text{comp}}^{(L)}(x, y) = J^{(L)}[\mathbf{A}^{(L)}(\mathbf{G}^{(L)}\mathbf{u}^{(L)} + \mathbf{g}^{(L)}) + \mathbf{d}^{(L)}] \end{cases}, \quad (5)$$

式中:  $I^{(L)}(\cdot)$  为计算原始图像  $I$  的  $L$  层采样图像函数;  $J^{(L)}(\cdot)$  为计算原始图像  $J$  的  $L$  层采样图像函数。

计算这一帧图像中像素点的亮度值  $I(x, y) = I^{(L)}(\mathbf{u}^{(L)})$ ,由于位移量较小,认为前后两帧偏导数相

等,并求得该值在两个坐标轴方向的偏导为

$$\begin{cases} I_x = \frac{[I(x+1,y) - I(x-1,y)]}{2} \\ I_y = \frac{[I(x,y+1) - I(x,y-1)]}{2} \end{cases}, \quad (6)$$

前后帧图像同一个位置坐标的灰度值的差为

$$\delta I(x,y) = I(x,y) - J(x,y). \quad (7)$$

空间梯度矩阵计算公式为

$$\mathbf{G} = \sum_{x=-w_x}^{w_x} \sum_{y=-w_y}^{w_y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}, \quad (8)$$

图像之间的灰度误差向量可表示为

$$\bar{\mathbf{b}} = \sum_{x=-w_x}^{w_x} \sum_{y=-w_y}^{w_y} \begin{bmatrix} I_x & \delta I \\ I_y & \delta I \end{bmatrix}. \quad (9)$$

上一层的光流计算结果,作为下一层光流初始估计值,记作  $\mathbf{g}^{(L)}$ ,通过最小化匹配误差计算该层的光流  $\mathbf{d}^{(L)}$ 。也可以说是用  $\mathbf{d}^{(L)}$  和  $\mathbf{g}^{(L)}$  估计下一层光流的初始值  $\mathbf{g}^{(L-1)}$ 。

假设在当前层上的光流  $\mathbf{d}^{(L)}$  和变换矩阵  $\mathbf{A}^{(L)}$  已经被计算出来,并作为计算结果传递到下一层,得到下一层的初始值估计为

$$\begin{cases} \mathbf{g}^{(L-1)} = 2(\mathbf{g}^{(L)} + \mathbf{G}^{(L)} \mathbf{d}^{(L)}) \\ \mathbf{G}^{(L-1)} = \mathbf{A}^{(L)} \mathbf{G}^{(L)} \end{cases}, \quad (10)$$

式中:  $\mathbf{g}^{(L-1)}$  为  $L-1$  层光流的值,  $\mathbf{G}^{(L-1)}$  为  $L-1$  层的空间梯度矩阵。

将  $\mathbf{g}^{(L-1)}$  和  $\mathbf{G}^{(L-1)}$  作为初始值,重新循环上面的步骤,直到最上一层,计算出光流  $\mathbf{d}$  和仿射变换矩阵  $\mathbf{A}$ 。

最后计算仿射光流,表达式为

$$\bar{\boldsymbol{\eta}} = \mathbf{G}^{-1} \bar{\mathbf{b}}, \quad (11)$$

式中:  $\mathbf{G}^{-1}$  为空间梯度逆矩阵;  $\bar{\mathbf{b}}$  为图像之间的灰度误差向量。(11)式的展开式为

$$[\eta_x \quad \eta_y \quad \eta_{xx} \quad \eta_{xy} \quad \eta_{yx} \quad \eta_{yy}]^T = \mathbf{G}^{-1} \bar{\mathbf{b}}, \quad (12)$$

式中:展开向量为各方向的像素值残差。更新跟踪结果,当  $\sqrt{n_x^2 + n_y^2} < \delta$  时完成这一层的迭代,阈值  $\delta$  经典值为 0.03。

循环迭代至原始图像层,光流值的最终值则为每一层计算的分段光流和。

## 4 基于双目的特征点匹配算法

双目图像之间需进行特征点匹配,且需查找出误匹配点并剔除。假设提取特征后得到的两个特征点的集合为  $P = \{p_1, p_1, \dots, p_n\}$  和  $Q = \{q_1, q_2, \dots, q_n\}$ , 匹配特征点就是在  $P$  集合中找到能与  $Q$  中元素一一对应的元素,通常是否对应的判定标准就是特征

点描述子之间的距离。

### 4.1 基本特征点匹配算法

目前常用的特征点匹配方法有暴力匹配(BF)算法和快速最近邻匹配(FLANN)算法。暴力匹配算法的原理为:计算每个特征点与需要匹配的特征点间的距离,将计算的距离进行从小到大的排序。由于本研究采用的 BRIEF(Binary Robust Independent Elementary Feature)描述子是二进制格式,故没有必要采用欧氏距离来计算。对于二进制格式的描述子来说,采用 Hamming 距离计算最快也最方便。两个字码中不同位的数目称为 Hamming 距离,换句话说,它就是将一个字符串变换成另外一个字符串所需要替换的字符个数,要统计有多少个位不同,只需将异或运算加诸于两个字码就可以,并在结果中统计 1 的个数。暴力匹配需要将  $P$  数据集中的每个点与  $Q$  数据集中的每个点作比较,算法复杂度为  $O(n^2)$ ,耗时长。快速最近邻特征点匹配算法需要建立数据索引,适用于高维数据,建好了 KD 树后,根据已知的待匹配的特征描述符向量,从根节点开始查找包含该特征描述符的对应空间节点,从这个节点开始计算特征描述符之间的距离,记录当前最近邻点与次近邻点,然后向上回溯,在每个父节点处更新当前最近邻和次近邻向量,并判断节点的另一个子节点是否存在与之距离更小的点,如果存在,则将其转入子节点中继续搜索。当回溯到根节点时,可以找到最终的最近邻点、次近邻特征点,算法的复杂度为  $O(\log n)$ ,相比于 BF 匹配法,FLANN 极大地降低了算法复杂度。

不论是 BF 还是 FLANN,他们是最基本的特征匹配操作,仅仅进行粗匹配时,特征点对中含有很多误匹配的点对和匹配质量不好的点对,故还需对匹配特征点对进行筛选。常用的特征点筛选方案有三种。第一种方案是基于最近距离,即两个特征点描述符的欧氏距离或 Hamming 距离最近,则视这两个特征点为最佳匹配;第二种方案是基于阈值,即两个描述符的距离小于一定阈值,则视这两个特征点匹配成功;第三种方案是基于最近邻域比率法,这种方法使用最近和次近距离的比值作为阈值提取匹配的特征点,通常阈值设定为 0.7。

根据经验,本文在基本匹配方面选择匹配效果最好的 FLANN 结合最近邻域比率法,但这种方式仍然不够严谨,还需在基本的匹配方法之上设计更为严格的匹配和筛选算法。

## 4.2 结合双向光流法的特征点环形匹配算法

在提取特征点之后,仅进行特征点单向匹配还是会有误匹配残留,因此,本文提出一种结合双向光流法的特征点环形匹配算法,算法示意图如图 3 所示。环形指的是特征点在左右图像和前后帧 4 幅图像之间形成闭环;双向光流法指的是前一帧和后一帧图像之间通过光流相互跟踪,其中后一帧向前一帧进行光流跟踪时称之为反向光流跟踪。反向光流跟踪这种处理方式是为了剔除一些错误的光流跟踪,比如说处在直线上的特征点,由于光流跟踪只是在局部跟踪,故跟踪结果极有可能在直线上滑动,基于反向跟踪技术能够避免这种情况出现。光流跟踪的基本假设包括:灰度不变假设和相同运动假设。灰度不变假设是指同一个空间点的像素灰度值,在各个图像中是固定不变的;相同运动假设是指某一个窗口内的像素具有相同的运动。特征点跟踪要求能够满足上述两条假设,如果不能全部满足就会导致跟踪失败。而跟踪失败的点就是需要从点集中删除的质量较差的点,因为它们容易产生误匹配,进而影响后续位姿估计的精度。

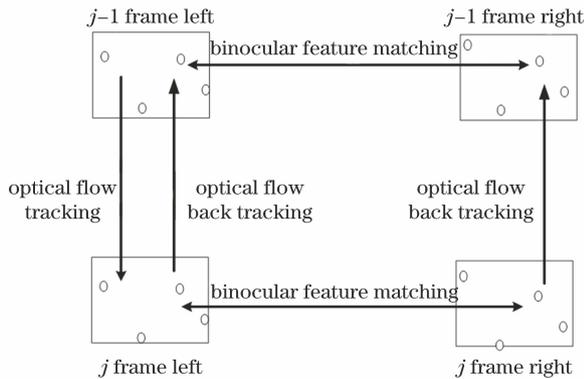


图 3 双向光流法环形匹配示意图

Fig. 3 Ring matching with bidirectional optical flow method

具体过程如下:以第  $j$  帧和第  $j-1$  帧的左右图像(共 4 幅)作为环形匹配基础,以某一帧图像称作第  $j$  帧为例,如果第  $j$  帧是第一帧,对于第一帧图像,由于没有参考帧,我们认为所有检测到的特征均为跟踪成功的特征,故直接提取双目图像 ORB 特征点并利用基础的 FLANN 结合距离最近邻比率匹配双目特征点。如果判断不是第一帧,则执行以下特征点跟踪机制。

1) 进行特征点正向光流跟踪。以相机左图像为基准,使用图像金字塔光流跟踪法(以下简称光流跟踪)得到上一帧特征在当前帧的位置,检测每个特征点像素坐标是否超出图像边界,保存跟踪成功且

未超出图像边界的左图像特征点,保存下来的坐标标志位设为 1,被筛掉的特征点标志位设为 0。这里需注意:标志位的引入使得与标志位相关的量同时更新,即标志位为 0 时特征点删除,标志位为 1 时特征点保留。结合标志位信息,可知未保存的特征点的标志位值为 0,所以对应的世界三维坐标的标志位也是 0,把标志位是 0 的世界三维坐标删除,经过光流跟踪后,不仅保存跟踪成功的特征点,还要更新前一帧左右图像特征点集和对应世界三维坐标点集。运用跟踪成功的左图像的特征点与右图像的特征点进行双目匹配,记录标志位,保留匹配成功的特征点。利用标志位得到左右匹配后保留的当前帧左右图像特征点集,并更新前一帧左右图像特征点集,记为  $F_0$  和  $F_1$ 。

2) 进行特征点反向光流跟踪。用光流跟踪将当前帧左右图像特征点向前一帧跟踪,此时得到的前一帧的特征点坐标为  $F'_0$  和  $F'_1$ ,如果  $f'_0 \in F'_0$  和  $f_0 \in F_0$  之间的坐标点欧氏距离小于 0.5 且  $f'_1 \in F'_1$  和  $f_1 \in F_1$  之间的坐标点欧氏距离小于 0.5,取满足这些条件的特征点对应的当前帧的特征点,根据标志位得到对应世界系的坐标集合,记作  $d_{30\_track}$ ,最终将这些特征点及其世界系坐标集合用于 3D-2D 方式估计位姿。

视觉里程计需要连续的图像序列。由于在双目匹配和前后帧跟踪的过程中会不断剔除特征点,因此完成一次环形匹配后,特征点会减少很多。如果不添加新的特征,几次循环后算法将无法进行下去,故需要为图像增加新的特征点。

本研究在提取新的特征点时,以保留下来的每个特征点坐标为中心,每个特征点使用一个半径为 30 pixel 的圆形面罩将其遮挡住,致使被覆盖区域不产生新的特征,仅在圆形以外提取新的特征点。图 4 为面罩遮挡示意图。程序上将圆形面罩区域内像素点的灰度值设为 0,这样基于角点计算的 ORB 特征点提取算法无法在圆形面罩内生成,从而保证特征点选取的平均化。除此之外,每次保留下来的特征点都会被分配自己的 ID,在设置时对每个特征点跟踪帧数进行记录并按降序排序,算法会倾向于保留跟踪时间长的特征点。这样做的好处是:不仅保留了经过双向光流法的环形匹配得到的优质特征点,使前后帧特征点具有历史相关性,缩短了检测到相同位置特征点所需的额外时间,还使得特征点均匀分布于图像上。

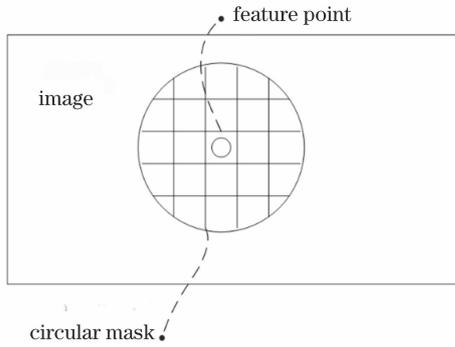


图 4 提取新特征点的面罩遮挡示意图

Fig. 4 Extracting new feature point with masks

### 5 位姿估计与优化

经过结合双向光流法的环形匹配算法对特征点进行选择与记录,得到保留下来的特征点的坐标,并求得它们在世界系下的三维坐标。在已知空间三维点坐标和它们对应的平面投影坐标的条件下,构建 3D-2D 点投影关系。P3P 是 3D-2D 最常用的位姿估计方法,输入数据为 3 对 3D-2D 匹配点,世界坐标系 3D 点记为  $A, B, C$ ,平面 2D 点记为  $a, b, c$ 。另设一对验证点  $D-d$ ,用于从最后可能的解中选出正确解。

从图 5 可以看到,3 对匹配点间存在相似关系,即  $\triangle oab \sim \triangle oAB, \triangle obc \sim \triangle oBC, \triangle oac \sim \triangle oAC$ ,  $o$  表示相机中心。

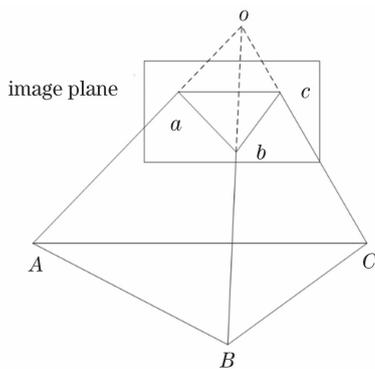


图 5 3D-2D 点投影

Fig. 5 3D-2D point projection

利用余弦定理有

$$\begin{cases} |oA|^2 + |oB|^2 - 2|oA| \cdot |oB| \cos\langle a, b \rangle = |AB|^2 \\ |oB|^2 + |oC|^2 - 2|oB| \cdot |oC| \cos\langle b, c \rangle = |BC|^2 \\ |oA|^2 + |oC|^2 - 2|oA| \cdot |oC| \cos\langle a, c \rangle = |AC|^2 \end{cases} \quad (13)$$

对(13)式全部除以  $|oC|^2$ , 并设  $x = \frac{|oA|}{|oC|}, y = \frac{|oB|}{|oC|}$ , 可得

$$\begin{cases} x^2 + y^2 - 2xy \cos\langle a, b \rangle = \frac{|AB|^2}{|oC|^2} \\ y^2 + 1 - 2y \cos\langle b, c \rangle = \frac{|BC|^2}{|oC|^2} \\ x^2 + 1 - 2x \cos\langle a, c \rangle = \frac{|AC|^2}{|oC|^2} \end{cases} \quad (14)$$

记  $v = \frac{|AB|^2}{|oC|^2}, uv = \frac{|BC|^2}{|oC|^2}, wv = \frac{|AC|^2}{|oC|^2}$ , 有

$$\begin{cases} x^2 + y^2 - 2xy \cos\langle a, b \rangle - v = 0 \\ y^2 + 1 - 2y \cos\langle b, c \rangle - uv = 0 \\ x^2 + 1 - 2x \cos\langle a, c \rangle - wv = 0 \end{cases} \quad (15)$$

把(15)中第一个式子中的  $v$  放到等式一边,代入另外两个等式,可得

$$\begin{cases} (1-u)y^2 - ux^2 - \cos\langle b, c \rangle y + 2uxy \cos\langle a, b \rangle + 1 = 0 \\ (1-w)x^2 - wy^2 - \cos\langle a, c \rangle x + 2wxy \cos\langle a, b \rangle + 1 = 0 \end{cases} \quad (16)$$

3 个夹角的余弦值为

$$\begin{cases} \cos\angle AoC = \cos\langle a, c \rangle = \frac{\vec{oa} \cdot \vec{oc}}{|\vec{oa}| \cdot |\vec{oc}|} \\ \cos\angle AoB = \cos\langle a, b \rangle = \frac{\vec{oa} \cdot \vec{ob}}{|\vec{oa}| \cdot |\vec{ob}|} \\ \cos\angle BoC = \cos\langle b, c \rangle = \frac{\vec{ob} \cdot \vec{oc}}{|\vec{ob}| \cdot |\vec{oc}|} \end{cases} \quad (17)$$

以  $oa$  为例,图像坐标点到相机光心距离的公式为

$$|oa| = \sqrt{(u_a - u_o)^2 + (v_a - v_o)^2 + f^2} \quad (18)$$

式中:  $f$  为焦距;  $u_a, v_a$  为点  $a$  在图像坐标系下的坐标;  $u_o, v_o$  为点  $o$  在图像平面上正投影的坐标。

将  $|oa|$  代入(17)式,可得

$$\cos\langle a, b \rangle = \frac{(u_a - u_o)(u_b - u_o) + (v_a - v_o)(v_b - v_o) + f^2}{\sqrt{(u_a - u_o)^2 + (v_a - v_o)^2 + f^2} \cdot \sqrt{(u_b - u_o)^2 + (v_b - v_o)^2 + f^2}} \quad (19)$$

同理可得  $\cos\langle a, c \rangle, \cos\langle b, c \rangle$ 。

在(16)式中,图像坐标系下的坐标值已知,  $u, w$  可以通过  $A, B, C$  在世界坐标系下的坐标计算出

来;余弦角已知;只有  $x, y$  是未知的,随相机移动而变化;故该方程组是关于  $x, y$  的一个二元二次方程组。该方程组的求解要用到吴消元法,每组  $x, y$  最

多有 4 个实根,对于每组  $x, y$ ,均求得对应的  $v$ ,随后可以推导出  $A, B, C$  和光心之间距离,即

$$\begin{cases} |oC| = \frac{|AB|}{\sqrt{v}} \\ |oA| = x \cdot |oC| \\ |oB| = y \cdot |oC| \end{cases} \quad (20)$$

假设特征点在图像平面坐标系上的横轴坐标为  $x$ ,在相机坐标系上的横轴坐标为  $X_c$ ,相机坐标系与图像平面坐标系的转换关系如图 6 所示,根据坐标系转换关系的相似三角形规则,可得

$$\frac{|op|}{|oP|} = \frac{x}{X_c} \quad (21)$$

故点  $P$  的三个坐标值可表示为

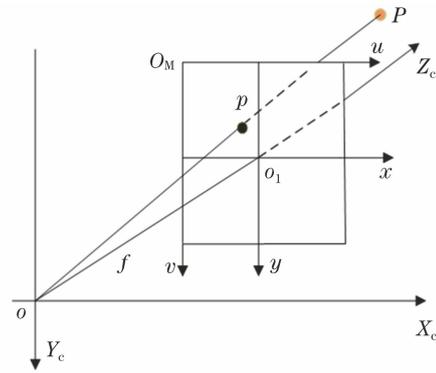


图 6 相机坐标系与图像坐标系关系

Fig. 6 Relationship between camera coordinate system and image coordinate system

$$\begin{cases} X_c = \frac{x \cdot |oP|}{|op|} = \frac{x \cdot |oP|}{\sqrt{x^2 + y^2 + f^2}} = \frac{\frac{x}{f_x} \cdot |oP|}{\sqrt{\left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2 + 1}} \\ Y_c = \frac{y \cdot |oP|}{|op|} = \frac{y \cdot |oP|}{\sqrt{x^2 + y^2 + f^2}} = \frac{\frac{y}{f_y} \cdot |oP|}{\sqrt{\left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2 + 1}} \\ Z_c = \frac{z \cdot |oP|}{|op|} = \frac{z \cdot |oP|}{\sqrt{x^2 + y^2 + f^2}} = \frac{|oP|}{\sqrt{\left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2 + 1}} \end{cases} \quad (22)$$

根据世界点在世界系下坐标  $P_w$  和相机系下坐标  $P_c$  的一一对应关系,求解位姿变换矩阵,即满足

$$P_c = RP_w + t, \quad (23)$$

式中  $R$  为旋转矩阵; $t$  为平移矩阵。

PnP 算法对噪声数据和 3D-2D 异常对应点相对敏感,如果对应点的对应关系异常,则得到的旋转、平移矩阵也必然不准确。为了解决外点问题,本研究将 RANSAC 与 P3P 结合,用于初步估计位姿。由于同一个空间点往往会被一部相机在不同的时间内多次观测,同一部相机在每个时刻观测到的点也不止一个,故构建最小二乘优化问题时,需对估计值进行捆绑调整(BA),以最小化重投影误差作为计算条件。具体实现时,将基于图优化理论解决非线性最小二乘优化问题的图优化库 g2o 作为优化工具,图优化库中的图由若干顶点和连接这些顶点的边组成,本系统设置位姿为顶点,边为误差项,优化前设置好库中指定的求解器和迭代算法,优化的过程就是调整满足边约束的顶点使得误差

最小。参数都选好之后,直接调用优化函数即可优化位姿。

## 6 实 验

实验采用的双目图像数据来自卡尔斯鲁厄 KITTI 数据集,该数据集由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创办,已被世界各国学者广泛采用。实验图像是经过校正和剪裁的双目图像,内外参数已知,非常适合 VO 实验。

### 6.1 特征匹配实验

图 7 和图 8 为 KITTI 00 序列第 107 帧双目图像特征点的匹配结果。图 7 表示 FLANN 结合最近邻域比例的方法去除误匹配(以下称为基本双目匹配)的结果,得到 104 对匹配关系,从匹配结果来看存在一些非常明显的误匹配连线,错误匹配结果已用粗实线标出,正确的匹配点对的连线应该是相互平行的,而错误匹配点对的连线与大部分平行连线相交。而采用本文提出的结合双向光流法的环形匹

配算法得到图 8 的匹配连线结果,共 60 对匹配成功的特征点,经过双向光流法结合双目匹配筛选误匹

配点,使得双目图像前后帧的约束性更强,经严格的筛选后,几乎看不出误匹配连线。



图 7 基本双目匹配结果

Fig. 7 Basic stereo matching results



图 8 本文提出的匹配算法

Fig. 8 Proposed matching algorithm

图 9 为根据本文提出的匹配算法得到的特征点分布,该图像拍摄于车辆右转弯处,讨论转弯处图像的特征点分布利于分析特征点跟踪时长,图上圆形特征点代表此特征点被跟踪的帧数较多,方形特征点表示被跟踪的帧数较少。



图 9 环形匹配保留的特征点

Fig. 9 Feature points retained by ring matching

从图 9 可以看出,保留下来的特征点较均匀地分布在图像上,跟踪帧数较多的圆形特征点主要集中在图中标记的方框内,左侧方框内景深较大,处在转弯之前直线行驶过程中的视野范围内,相比于直行路段景深较小的地方,圆形的特征点较多,表明景深大的地方原始特征点属于跟踪多帧的原始特征点;右侧方框内的特征点处于整个图像的中心位置,在视野变换的过程中不易受到影响,因此跟踪到较多的特征点;跟踪帧数较少的方形特征点多数分布在图像右侧,这是由于车辆右转弯,图像右侧属于新视野范围,该结论符合逻辑。总之,筛选保留下来的特征点质量良好。

图 10 和图 11 分别是无面罩遮挡和有面罩遮挡保留特征点条件下的提取新特征点的实验结果。从



图 10 无面罩遮挡条件下提取新特征点

Fig. 10 Extracting new feature points without mask

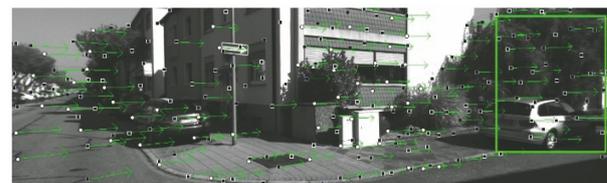


图 11 有面罩遮挡条件下提取新特征点

Fig. 11 Extracting new feature points under masks

圆形特征点数目来看,图 11 中跟踪帧数较多的圆形特征点数目较多,该实验结果现象与本文设计的面罩遮挡会根据特征点 ID 进行跟踪次数降序排序并倾向保存跟踪帧数多的特征点算法逻辑相符,跟踪帧数多的圆形特征点更稳定,使帧与帧之间信息产生更强的关联性,这样的特征点对后续进行位姿估计具有积极影响。从图像边界来看,方框处于图像信息变化较大的位置,图中箭头表示特征点的光流方向,由于车子向右旋转,方框内多数是新到来的信息,可以看出图 11 新的特征点比图 10 多,这说明该算法在提取新特征点时能够提取到较多的新信息;从整个图像上的特征点分布情况来看,有面罩遮挡

比无面罩遮挡时特征点分布均匀。总的来说,本章提出的特征匹配与筛选策略能提高特征点的稳定性并能保证特征点均匀分布于图像上。

以 KITTI Odometry grayscale 数据集 sequence00 序列作为输入数据,分别采用 FLANN 结合最近邻域比例匹配算法(基本匹配算法)和本文提出的双向光流法环形匹配算法进行双目视觉里程计实验,得到如图 12 所示的特征点数目波形图。横轴表示图片帧数,纵轴表示保留下来用于位姿估计的特征点数目,带方格线(上波纹)表示采用基本特征点匹配

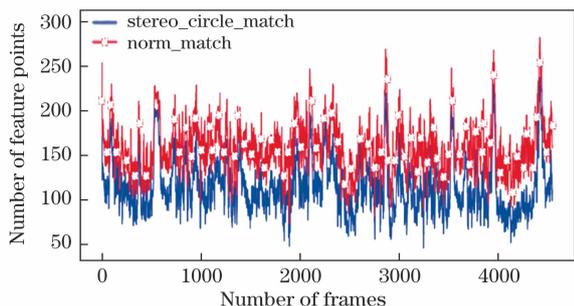


图 12 特征点数目波形图

Fig. 12 Comparison of the number of feature points

筛选算法在双目视觉里程计中保存的特征点数目,实线(下波纹)表示本文提出的算法特征点数目波形图,从波形图可以看出本文特征点匹配算法能去除一定的误匹配点。

### 6.2 KITTI 数据集运行结果

综合双目视觉里程计中的各模块技术,进行双目视觉里程计系统实验,并给出算法在 KITTI 数据集中的运行结果。

图 13 表示应用 KITTI 数据集序列运行双目视觉里程计的轨迹结果图。其中,每行并列图片中的左图为轨迹俯视图,右图为三个坐标系的值分解图,右图的横轴表示图像帧数,纵轴表示在各个方向上的值。图中 vokitti\_00 表示视觉里程计运用基本匹配方法得到的轨迹结果, stereo\_circle 表示视觉里程计采用本文匹配方法得到轨迹结果, 00 线表示真实参考轨迹。结果表明本文算法更贴近真实参考轨迹,基本匹配算法的绝对轨迹误差值为 64.528 m,本文匹配算法得到的绝对轨迹误差为 40.54 m,真实轨迹全程 3723.485 m,本文提出的匹配算法比基本匹配算法精度提高了接近 0.7%。

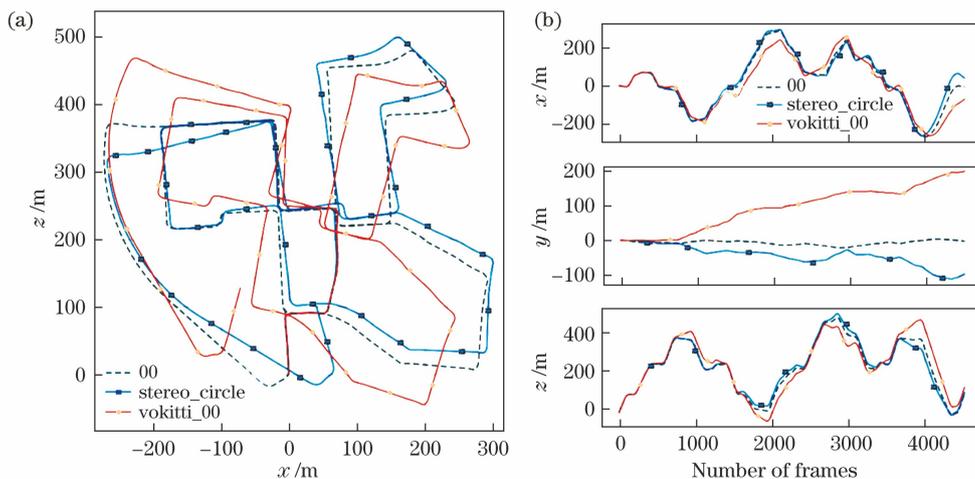


图 13 算法轨迹图。(a)俯视图;(b)xyz 轴坐标值分解图

Fig. 13 Algorithm trajectory. (a) Top view; (b) xyz axis coordinate decomposition value

为了从精度方面评估算法的优劣性,采用评估估计轨迹与真实轨迹之间的均方根误差 (RMSE)、均值 (MEAN)与标准差(STD) 来反映本文算法的精确性。

本研究对 KITTI 视觉里程计数据集中的前 11 序列(00-10)进行了测试与对比。选择 ORB-SLAM2<sup>[15]</sup>与基于 Laplace 分布的双目视觉里程

计<sup>[16]</sup>(以下简称算法 A)与本文算法作对比。从表 1 中可以看到,本文算法在 11 个测试序列中有 9 个序列的 RMSE 数值优于其他两种对比算法。在 04 和 05 序列中 ORB-SLAM2 的 RMSE 值略优于本文算法,但在 STD 值稍逊于本文算法。综上可知,本文提出的算法对双目视觉里程计定位精度的提高有明显的作

表 1 本文算法、ORB-SLAM2 以及算法 A 估计的轨迹与真实轨迹之间 RMSE、MEAN、STD 的对比  
Table 1 Comparison of RMSE, MEAN, STD between the trajectory estimated by ours, ORB-SLAM2, and Algorithm A and the real trajectory

Sequence	RMSE /m			MEAN /m			STD /m		
	Ours	ORB-SLAM2	Algorithm A	Ours	ORB-SLAM2	Algorithm A	Ours	ORB-SLAM2	Algorithm A
00	<b>1.292</b>	1.319	5.248	<b>1.151</b>	1.189	4.696	0.587	<b>0.571</b>	2.343
01	<b>10.062</b>	10.581	33.938	<b>9.573</b>	10.018	28.257	<b>3.097</b>	3.406	18.797
02	<b>5.660</b>	6.169	11.365	<b>4.798</b>	5.142	10.332	<b>3.002</b>	3.408	4.733
03	<b>0.631</b>	0.684	1.031	<b>0.534</b>	0.5990	0.909	<b>0.336</b>	0.346	0.486
04	0.224	<b>0.215</b>	0.495	0.215	<b>0.203</b>	0.426	<b>0.065</b>	0.072	0.207
05	0.715	<b>0.677</b>	4.207	0.648	<b>0.602</b>	3.061	<b>0.299</b>	0.308	2.885
06	<b>0.636</b>	0.760	2.839	<b>0.596</b>	0.710	2.533	<b>0.222</b>	0.271	1.272
07	<b>0.534</b>	0.970	3.655	<b>0.502</b>	0.896	3.079	<b>0.183</b>	0.371	1.971
08	<b>3.210</b>	3.882	13.001	<b>2.925</b>	3.514	12.555	<b>1.322</b>	1.652	2.594
09	<b>1.664</b>	2.750	4.668	<b>1.520</b>	2.236	3.561	<b>0.676</b>	1.601	3.018
10	<b>0.976</b>	1.028	2.817	<b>0.884</b>	0.925	2.628	<b>0.413</b>	0.449	1.013

## 7 结 论

提出一种结合双向光流法的双目视觉里程计方法,该方法有效地解决了视觉里程计中采用基本特征点匹配算法易出现误匹配的问题,并在利用数据进行定位实验时得到较好的定位结果,验证了所提算法的有效性和优越性。

### 参 考 文 献

- [1] Nister D, Naroditsky O, Bergen J. Visual odometry [C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, June 27-July 2, 2004, Washington, DC, USA. New York: IEEE Press, 2004: 1315094.
- [2] Lin F C, Liu Y H, Zhou J F, et al. Optimization of visual odometry algorithm based on ORB feature[J]. Laser & Optoelectronics Progress, 2019, 56(21): 211507.  
林付春, 刘宇红, 周进凡, 等. 基于 ORB 特征的视觉里程计算法优化[J]. 激光与光电子学进展, 2019, 56(21): 211507.
- [3] Zou B, Lin S Y, Yin Z S. Semantic mapping based on YOLOv3 and visual SLAM[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201012.  
邹斌, 林思阳, 尹智帅. 基于 YOLOv3 和视觉 SLAM 的语义地图构建[J]. 激光与光电子学进展, 2020, 57(20): 201012.
- [4] Zhang T X, Zhou J, Liao H L, et al. Simultaneous localization and mapping strategy of graph optimization based on three-dimensional laser[J]. Laser & Optoelectronics Progress, 2019, 56(20): 201502.
- [5] Lu S D, Tu M Y, Luo X Y, et al. Laser SLAM pose optimization algorithm based on graph optimization theory and GNSS [J]. Laser & Optoelectronics Progress, 2020, 57(8): 081024.  
陆世东, 涂美义, 罗小勇, 等. 基于图优化理论和 GNSS 激光 SLAM 位姿优化算法[J]. 激光与光电子学进展, 2020, 57(8): 081024.
- [6] Zhang Z J, Cheng X J, Cao Y J, et al. Application of 3D reconstruction of relic sites combined with laser and vision point cloud[J]. Chinese Journal of Lasers, 2020, 47(11): 1110001.  
张子健, 程效军, 曹宇杰, 等. 结合激光与视觉点云的古遗迹三维重建应用[J]. 中国激光, 2020, 47(11): 1110001.
- [7] Yang Q X, Wang L, Yang R G, et al. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 31(3): 492-504.
- [8] Olson C F, Abi-Rached H, Ye M, et al. Wide-baseline stereo vision for Mars rovers [C]//Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), October 27-31, 2003, Las Vegas, NV, USA. New York: IEEE Press, 2003: 1302-1307.
- [9] Geiger A, Ziegler J, Stiller C. StereoScan: dense 3D reconstruction in real-time [C]//2011 IEEE Intelligent Vehicles Symposium (IV), June 5-9, 2011, Baden-Baden, Germany. New York: IEEE Press, 2011: 2015002-10.

- 963-968.
- [10] Badino H, Yamamoto A, Kanade T. Visualodometry by multi-frame feature integration [C] // 2013 IEEE International Conference on Computer Vision Workshops, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 222-229.
- [11] Chien H J, Klette R. Substantial improvement of stereo visual odometry by multi-path feature tracking [C] // 2017 International Conference on Image and Vision Computing New Zealand (IVCNZ), December 4-6, 2017, Christchurch. New York: IEEE Press, 2017: 1-6.
- [12] Raguram R, Frahm J M, Pollefeys M. A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus [M] // Forsyth D, Torr P, Zisserman Z. Computer vision-ECCV 2008. Lecture notes in computer science. Cham: Springer, 2008, 5303: 500-513.
- [13] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C] // 2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 2564-2571.
- [14] Hsieh C K, Lai S H, Chen Y C. 2D expression-invariant face recognition with constrained optical flow [C] // 2009 IEEE International Conference on Multimedia and Expo, June 28-July 3, 2009, New York, NY, USA. New York: IEEE Press, 2009: 1058-1061.
- [15] Mur-Artal R, Tardós J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras [J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [16] Fan H Q, Wu J H. Stereo visual odometry based on Laplace distribution [J/OL]. Acta Automatica Sinica: 1-11. [2021-02-20]. <https://doi.org/10.16383/j.aas.c190860>.
- 范涵奇, 吴锦河. 基于 Laplace 分布的双目视觉里程计 [J/OL]. 自动化学报: 1-11. [2021-02-20]. <https://doi.org/10.16383/j.aas.c190860>.