

DOI: 10.13382/j.jemi.B2104994

# 基于视觉里程计的室内位姿测量技术研究\*

张艳 王宇

(南京理工大学机械工程学院 南京 210094)

**摘要:**随着人口老龄化程度的不断加深,独居老人也在不断增多。在解决独居老人养老问题的应用上,室内定位是最基本和关键的问题。针对室内定位的需求,提出了双目视觉里程计定位算法,从相机成像模型、特征提取、特征匹配以及运动估计4个方面展开了研究。首先,采用双目摄像头作为传感器进行图像采集;然后通过提取ORB特征点来完成相邻图像之间的匹配关系,根据基于BRIEF描述子的立体匹配算法得到左右图像对应特征点的匹配关系;最后对相机的运动进行估计。设计了硬件和软件平台对提出的方法进行实验。实验证明,基于视觉里程计的室内定位技术能够准确地定位老人在室内的位置,能够实时地对老人进行安全监护。

**关键词:**室内定位;双目视觉;视觉里程计;特征匹配

**中图分类号:** TH702      **文献标识码:** A      **国家标准学科分类代码:** 460.4

## Design and implementation method of a visual odometer

Zhang Yan Wang Yu

(School of Mechanical Engineering, Nanjing University of Science &amp; Technology, Nanjing 210094, China)

**Abstract:** As the aging of the population continues to deepen, the number of elderly people living alone is also increasing. In the application of solving the problem of elderly living alone, indoor positioning is the most basic and critical issue. Aiming at the needs of indoor positioning, a binocular visual odometer positioning algorithm is proposed, which has been studied from four aspects: Camera imaging model, feature extraction, feature matching and motion estimation. First, the binocular camera is used as the sensor for image acquisition; then the matching relationship between adjacent images is completed by extracting ORB feature points, and the matching relationship between the corresponding feature points of the left and right images is obtained according to the stereo matching algorithm based on the BRIEF descriptor, the camera movement is estimated. The hardware and software platforms are designed to experiment with the proposed method. Experiments have proved that the indoor positioning technology based on visual odometer can accurately locate the position of the elderly indoors, and can safely monitor the elderly in real time.

**Keywords:** indoor positioning; binocular vision; visual odometer; feature matching

## 0 引言

由于老年人身体机能等方面的衰退以及受多种慢性疾病的影响,增强独居老人室内日常行为活动的安全监护,降低室内活动的意外风险,具有重要的社会意义。定位精度高、可靠性高的室内定位技术能够准确地确定老人在室内的位置,从而对老人的行踪进行实时的监护。

现有常用的室内定位技术主要包括无线 Wi-Fi、蓝牙、超声波定位、射频识别(RFID)、超宽带(UWB)等技术。Wi-Fi 和蓝牙等<sup>[1-2]</sup>依据信号强度定位的技术虽然实现成本低廉,但容易受噪声干扰且干扰较大,信号稳定性不好;超声波定位受多径效应和非视距传播影响很大,因而降低了其可用性;RFID 定位系统<sup>[3]</sup>抗干扰能力强,但是需要投入大量的硬件设备,布设识别器和天线需要大量的工程实践经验,难度大成本高,无法做到精确定位;

收稿日期: 2021-12-08      Received Date: 2021-12-08

\* 基金项目: 国家自然科学基金青年科学基金(61803203)、“基于无人机航迹优化的喷洒精准对靶方法研究”项目资助

UWB 技术<sup>[4]</sup>具有定位精度高,穿透性强的优点,然而其延时模块需要精确校准。基于计算机视觉的定位技术利用相机作为传感器采集图像获取视觉信息,用算法实现室内定位,该技术成本较低,精度高,且不会受到未知空间、环境和长距离的影响。

随着视觉技术的不断发展,依靠视觉的定位技术已被广泛应用。最早应用在视觉里程计上的是单目相机, Davision 等<sup>[5]</sup>提出了第 1 个实时的单目视觉 slam 系统 MonoSLAM,但单目视觉里程计需要外部信息进行约束才能获取空间尺度,且估算出的深度有较大的误差,这对姿态的估计非常不利。双目立体视觉通过立体匹配来计算获取深度信息,R-GBD 相机通过向被测物发送红外光或脉冲光来确定目标与自身的距离,解决了单目相机尺度不确定性的问题。但 R-GBD 相机受光照影响较大,且成本较高,而双目相机只需两个摄像头即可,成本较低,因此本文利用双目相机进行图像信息采集。

目前主流的双目定位方法主要有光流法和特征点法。光流法基于光照不变假设<sup>[6]</sup>,但实际中不可能完全符合光照不变性假设,导致其定位精度低。基于特征的方法主要包括点特征、边缘特征和块特征。其中点特征法在识别和抗噪声能力方面优于其他两种方法<sup>[7]</sup>。文献<sup>[8]</sup>使用点线结合的方法来降低若纹理情况对点特征的影响,然而加入线特征使匹配变得复杂困难,不利于系统的实时性。质量好的特征点具有可重复性和唯一性,如 SIFT 算法<sup>[9]</sup>、SURF 算法<sup>[10]</sup>、FAST<sup>[11]</sup>和 ORB 算法等<sup>[12]</sup>。相比前几种特征提取算法,ORB 特征利用改进的 FAST 角点,一方面提高了速度,另一方面增强了光照不变性、尺度不变性和旋转不变性,鉴于其良好的鲁棒性和实时性,本文采用 ORB 算法作为特征提取算法。

通过立体匹配来获取同一目标在左右两幅图像上所对应的两个像素点,准确的确定其对应关系是获取目标点深度信息的关键。立体匹配算法可分为全局立体匹配算法和局部立体匹配算法<sup>[13]</sup>,全局立体匹配算法<sup>[14]</sup>利用图像的全局信息进行计算,计算量非常大,不适用于实时性要求高的里程计系统。局部匹配算法利用局部信息进行计算,文献<sup>[15]</sup>提出的基于区域灰度的局部立体匹配算法是较为经典的局部匹配算法,该算法匹配精度高,速度快,但视觉效果较差,抗干扰能力弱。文献<sup>[16]</sup>改进了 AD-Census 立体匹配算法,将梯度、距离阈值和颜色等结合进行交叉交叉域的代价聚合,一定程度上改善了弱纹理区域的匹配效果,但其图像预处理部分过于繁琐,不利于视觉定位的实时性。为提高立体匹配的精确程度以及兼顾视觉里程计的实时性要求,本文采用一种基于 BRIEF 描述子的局部立体匹配算法,以 BRIEF 描述子作为特征信息构造代价函数来进行评估,使深度信息的获取更具精确度。

综上,针对独居老人室内精确实时定位的需求,本文提出了一种基于特征点法的双目定位算法,该算法非常规地使用 BRIEF 描述子来构造代价函数,以获得准确的深度信息。算法通过跟踪 ORB 特征点来寻找两幅图像之间的特征对应关系,通过观测特征点的跟踪情况对可能出现的异常情况发出预警信息,并使用 PnP (perspective-n-point) 的 3D-2D 模型来进行老人的实时运动估计。使得老年人的位置和运动轨迹在得到及时监控的同时可以发出潜在风险的警告信息。

## 1 视觉里程计的设计

算法的主要流程如图 1 所示。在相机输入当前帧图像后,根据双目相机标定参数对图像进行立体校正,然后对校正后的左目图像进行关键点的提取。如果当前帧为系统启动以来的第 1 帧图像,则直接提取图像中的关键点,以当前左目相机的相机坐标系作为系统的世界坐标系来计算关键点的世界坐标,当前的载体位置和姿态矩阵为 $[R_{uc}, t_{uc}] = [I_3, 0]$ 。如果当前帧不是系统启动以来的第 1 帧,则对当前的左图像进行关键点追踪,确定像素点的对应关系,并以此估计当前载体位姿。输出当前位姿后,将图像中新出现的关键点进行立体匹配,计算其深度,恢复其相机坐标和世界坐标。之后采集下一帧图像,进行循环。

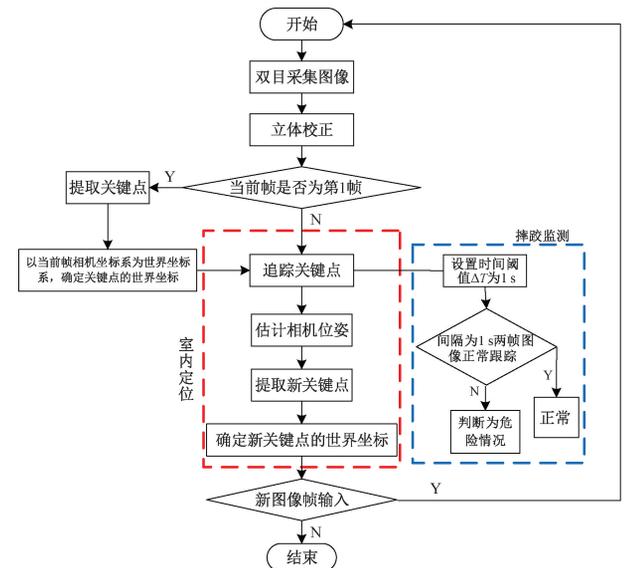


图 1 算法主要流程

Fig. 1 Algorithm main flow chart

在关键点跟踪过程中,设置时间阈值 $\Delta T$ 为 1 s,判断间隔时间为 1 s 的两帧图像之间关键点的跟踪情况,根据关键点跟踪情况来判断老人是否存在摔倒等危险。老人一般为线性缓慢运动,正常情况下间隔 1 s 的两帧图像有

较多共视点,即后一帧能够跟踪到较多的关键点;而老人发生摔倒等危险时,相机采集到的相隔 1 s 的两帧图像有较少甚至没有共视点,后一帧图像跟踪到的关键点也将很少或没有,出现关键点丢失的情况。

### 1.1 双目相机成像模型

理想状态下的双目相机模型为:两相机焦距相等,成像平面处于同一平面且光轴平行,图像中各行的像素精确对准<sup>[17]</sup>,如图 2 所示。

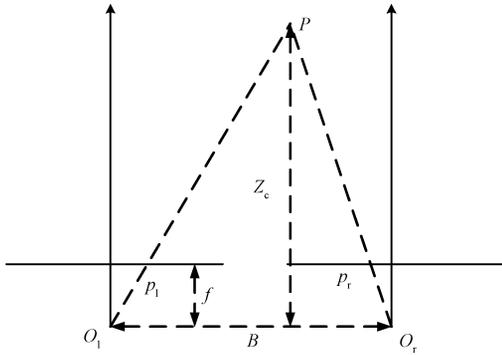


图 2 理想双目相机成像模型

Fig. 2 Ideal binocular camera imaging model

空间中一点的相机坐标为  $P^c(X_c, Y_c, Z_c)$ ,其中  $Z_c$  即为其深度。该点在左右相机的投影,即像素坐标分别为  $p_l(u_l, v), p_r(u_r, v)$ ,相机坐标系坐标分别为  $p'_l(x_l, y), p'_r(x_r, y)$ 。由相似三角形原理可求解该点的深度:

$$\frac{Z_c}{Z_c - f} = \frac{B}{B - x_l + x_r} \quad (1)$$

化简后有:

$$\frac{Z_c}{B} = \frac{f}{dx(u_l - u_r)} \quad (2)$$

实际情况中,两相机通常难以达到理想的共面、行对准等条件,在使用前需对其进行校正。一般用右侧相机相对左侧相机的旋转矩阵  $\mathbf{R}$  和位移向量  $\mathbf{t}$  表示两个相机之间的相对位姿。

对于空间点  $P^w(X_w, Y_w, Z_w)$ ,在左右相机坐标系的坐标分别为  $P^{cl}(X_{cl}, Y_{cl}, Z_{cl}), P^{cr}(X_{cr}, Y_{cr}, Z_{cr})$ 。则  $P^{cl}$  和  $P^{cr}$  的关系可以表示为:

$$P^{cr} = \mathbf{R}P^{cl} + \mathbf{t} \quad (3)$$

而  $P^{cl}$  和  $P^{cr}$  相对与世界坐标系坐标  $P^w$  的关系可以表示为:

$$\begin{aligned} P^{cl} &= \mathbf{R}_{wcl}P^w + \mathbf{t}_{wcl} \\ P^{cr} &= \mathbf{R}_{wcr}P^w + \mathbf{t}_{wcr} \end{aligned} \quad (4)$$

其中,  $\mathbf{R}_{wcl}, \mathbf{t}_{wcl}$  和  $\mathbf{R}_{wcr}, \mathbf{t}_{wcr}$  为左右相机所在位置的外参。

联立三式可解得:

$$\begin{aligned} \mathbf{R} &= \mathbf{R}_{wcr}\mathbf{R}_{wcl}^T \\ \mathbf{t} &= \mathbf{t}_{wcr} - \mathbf{R}\mathbf{t}_{wcl} \end{aligned} \quad (5)$$

由于误差的存在,每对点解算到的  $\mathbf{R}, \mathbf{t}$  会有所差别。在相机标定时令每组图像计算出的  $\mathbf{R}, \mathbf{t}$  的中位数作为初值,进行极大似然估计,通过最小化重投影误差,得到精确的标定结果。

$$\mathbf{E} = \sum_{i=1}^n \sum_{j=1}^m \| p_{ij} - p_{ij}^{\wedge}(K, k_1, k_2, k_3, p_1, p_2, \mathbf{R}_{wcli}, \mathbf{t}_{wcli}, \mathbf{R}_{wcri}, \mathbf{t}_{wcri}, \mathbf{R}, \mathbf{t}, P_j) \|^2 \quad (6)$$

其中,  $\mathbf{E}$  为重投影误差,  $K$  为相机内参,  $k_1, k_2, k_3$  为径向畸变参数,  $p_1, p_2$  为切向畸变参数。

在估计出右侧相机相对左侧相机的旋转矩阵  $\mathbf{R}$  和位移向量  $\mathbf{t}$  后,利用所得的  $\mathbf{R}$  和  $\mathbf{t}$  对双目相机进行立体校正。将  $\mathbf{R}$  按式(7)约束拆分成两个矩阵  $\mathbf{r}_l, \mathbf{r}_r, \mathbf{r}_l$  代表左摄像机旋转一半的旋转矩阵,  $\mathbf{r}_r$  代表右摄像机旋转 1/2 的旋转矩阵,它们的旋转方向相反。

$$\begin{aligned} \mathbf{R} &= \mathbf{r}_r^{-1}\mathbf{r}_l \\ \mathbf{r}_r\mathbf{r}_l &= \mathbf{I} \end{aligned} \quad (7)$$

使左右成像平面按  $\mathbf{r}_l, \mathbf{r}_r$  分别进行旋转,从而达到共面,效果如图 3 所示。

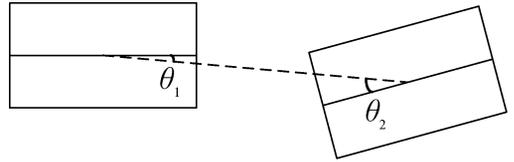


图 3 旋转后的两成像平面

Fig. 3 Two imaging planes after rotation

此时两平面的基线并未平行,利用位移向量  $\mathbf{t}$  来构造变换矩阵  $\mathbf{R}_{rect} = [\mathbf{e}_1^T, \mathbf{e}_2^T, \mathbf{e}_3^T]^T$  使得两平面基线平行,其中  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  的定义如式(8)所示。

$$\begin{cases} \mathbf{e}_1 = \frac{\mathbf{t}}{\|\mathbf{t}\|} \\ \mathbf{e}_2 = \frac{[-t_y, t_x, 0]^T}{\sqrt{t_x^2 + t_y^2}} \\ \mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2 \end{cases} \quad (8)$$

则立体校正时的左右相机的变换矩阵可以表示为:

$$\begin{aligned} \mathbf{R}_{left} &= \mathbf{R}_{rect}\mathbf{r}_l \\ \mathbf{R}_{right} &= \mathbf{R}_{rect}\mathbf{r}_r \end{aligned} \quad (9)$$

### 1.2 特征提取

本文通过提取 ORB 特征点来完成相邻图像之间的匹配关系,ORB 特征点由关键点和描述子两部分组成,故其特征检测算法也可分为两部分,分别是关键点提取和描述子构建。其中 ORB 关键点提取算法采用的是由 FAST 算法改进而得的 oFAST 算法,该算法引入图像金字

塔以及灰度质心法,使得 FAST 特征具有尺度不变性和旋转不变特性。

oFAST 算法主要步骤如下:

1) 如图 4 所示,以点  $P$  为圆心确定一个半径为 3 个像素点的圆,该圆的圆周经过 16 个像素格,分别记为  $P_1 \sim P_{16}$ ,记点  $P$  亮度值为  $I_p$ 。

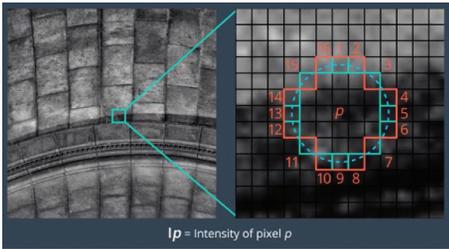


图 4 像素示意图

Fig. 4 Pixel diagram

2) 设置阈值  $t$ ,在  $P_1 \sim P_{16}$  像素中,首先检测圆上的第 1、5、9、13 pixel 的亮度,当这 4 个像素中有 3 个像素的亮度同时大于  $I_p + t$  或小于  $I_p - t$  时,当前像素才被认为有可能是一个角点,否则直接排除。在判断该像素可能为一个角点后,继续检测,若检测过程中出现连续 12 个像素点的亮度值全大于  $I_p + t$  或者全小于  $I_p - t$  的情况,则将  $P$  点记为候选特征点。

3) 对以上两步进行循环,对所有像素进行相同操作。

4) 去除局部密集的候选特征点,计算每个候选点的 FAST 得分,计算公式为:

$$V = \max \left( \sum_{x \in S_b} |I_n - I_p| - t, \sum_{x \in S_l} |I_n - I_p| - t \right) \quad (10)$$

其中,  $I_n$  为圆周上像素点的值,  $S_b$  为亮度值大于  $I_p + t$  的像素点集合,  $S_l$  为灰度值小于  $I_p - t$  的像素点集合。比较相邻候选特征点的  $V$  值,保留  $V$  值大的候选点为关键点,去除  $V$  值小的候选点。

5) 构建高斯金字塔,添加关键点的尺度不变性。设置金字塔层数  $n = 8$ ,比例因子  $s = 1.2$ 。将原图像按比例因子缩放,每层图像像素值计算如式(11)所示。

$$I' = I_p / 1.2^k \quad (k = 1, 2, \dots, 8) \quad (11)$$

6) 利用灰度质心法构建方向向量,添加关键点的旋转不变性。

在关键点邻域的图像块中,定义图像块的矩为:

$$m_{pq} = \sum_{x,y \in B} x^p y^q I(x,y) \quad p, q = \{0, 1\} \quad (12)$$

通过图像块的矩确定图像块的质心:

$$C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (13)$$

连接图像块的几何中心  $O$  与质心  $C$ ,得到一个方向向量  $\vec{OC}$  于是特征点的方向可以定义为:

$$\theta = \arctan(m_{01}/m_{10}) \quad (14)$$

在提取 oFAST 关键点后,计算每个点的描述子。ORB 使用改进的 BRIEF 特征描述。BRIEF 是一种二进制描述子,它的描述向量由许多个 0 和 1 组成,这里的 0 和 1 编码了关键点附近两个像素(例如  $p$  和  $q$ )的大小关系:如果  $p$  比  $q$  大,则取 1,否则取 0。若以关键点为圆心,一定像素为半径取 128 个这样的  $p, q$ ,就可以得到 128 维的由 0、1 组成的向量。BRIEF 按一定的方式随机选点,速度非常快,存储方便,适用于实时的图像匹配。在 ORB 中,rbrief 描述子在 BRIEF 描述子的基础上加入旋转因子进行改进。

### 1.3 特征点匹配

对于前后两帧图像的特征点匹配,若直接采用比较每个特征点之间汉明距离的暴力算法,计算量太大,不适用于实时性要求高的里程计系统。近似最近邻算法集成在 FLANN 开源库中,适用于实时性要求高的场合,但会在一定程度上出现误匹配。采用随机采样一致性(RANSAC)算法可在一定程度上消除近似最近邻算法带来的误匹配,其核心机理为使用有效方式完成单应性矩阵的计算。单应性矩阵为描述平面上两幅图片对应点映射关系的转换矩阵,定义如下:

$$\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (15)$$

其中,  $(u \ v \ 1)^T$  和  $(u' \ v' \ 1)^T$  为像素点坐标,

$\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix}$  为单应性矩阵  $H$ 。

一对点可以确定两个方程,至少需要 4 对匹配点来确定  $H$  阵。实际中初匹配点的对数远大于 4,为确定出最有利于  $H$  阵精度的 4 对匹配点,可以采用随机采样一致性检验求解。主要步骤如下:

1) 以一定方式随机抽取 4 对匹配点,对模型进行拟合(即估算单应性矩阵  $H$ );

2) 由于误差的存在,数据点存在一定波动,假设误差范围为  $\delta$ ,以第 1 步  $H$  阵为基准,计算剩余匹配点误差,找出误差范围内的点,并统计点的个数  $n$ ;

3) 重新随机选取 4 对点,重复第 1) 步、第 2) 步的操作,直到结束迭代;

4) 找出最大  $n$  满足的单应性矩阵  $H$ 。

对于左右两相机图像对应特征点的匹配,即立体匹配,本文采用基于 BRIEF 描述子的局部立体匹配算法。传统 BM(block matching)算法如图 5 所示,在立体校正后,以左目相机图像为参考,设一关键点像素坐标为  $(u_l, v)$ ,其灰度为  $I(u_l, v)$ 。以该关键点为中心,将它周围  $M \times N$  的区域窗口  $W$  作为匹配单元。由双目成像模型可知,

右图像中与该关键点对应的点  $(u_r, v)$  必定在该点的左侧。以右图中与左图像中关键点相同坐标的点为起始点,沿所在行向左滑动窗口  $W$ , 依次比较每个像素, 计算其相似度。将相似度的值最大的点作为匹配点。

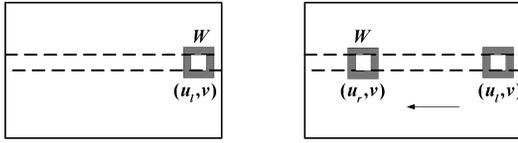


图5 BM 算法原理

Fig. 5 BM algorithm principle diagram

由于遮挡等因素的存在,左侧图像中的关键点并不一定能在右侧图像中找到匹配点,在匹配过程中,采用绝对误差和  $S_1$  作为相似性测度函数,用来衡量两点及其周围窗口的匹配程度,如式(16)所示。

$$S_1 = \sum_{i,j \in W} | I_l(u_l + i, v + j) - I_r(u_r + i, v + j) | \quad (16)$$

其中,  $I_l$  和  $I_r$  分别为左右像素点的亮度值。但由于光照等原因,其像素点的亮度值易于受到外界干扰从而造成误差影响匹配精度。在提取 ORB 特征点过程中,已经获得 BRIEF 描述子,将 BRIEF 向量代替亮度值作为特征信息来构造代价函数,能够有效提高匹配精度。其相似性测度函数  $S_2$  可表示如式(17)所示:

$$S_2 = \sum_{i,j \in W} | L(u_l + i, v + j) - R(u_r + i, v + j) | \quad (17)$$

其中,  $L, R$  主要指左右像素的 BRIEF 特征,当相似性测度函数数值处于峰值时,匹配会处于结束状态,所得点则是匹配点。

### 1.4 运动估计

通过对关键点进行追踪,得到关键点的匹配关系后,根据这些匹配关系恢复相机运动,经过立体匹配,能够得到关键点的三维相机坐标信息  $P_c$ 。对于已知外参的第  $k-1$  帧图像,可以据此恢复关键点的世界坐标。对关键点进行跟踪得到关键点在后续图像帧的像素坐标,根据像素坐标和世界坐标系之间的对应关系,可以恢复相机在第  $k$  帧的位姿。因此,就将相机运动恢复转化为了 2D-3D 的多点透视问题即 PnP 问题,本文采取非线性优化方法进行求解。

将像素坐标(观测到的投影位置,即匹配得到的)与 3D 空间点按照当前估计的位姿进行投影得到的坐标位置相比较从而得到误差项,即重投影误差,非线性优化的过程即最小化重投影误差的过程。如图 6 所示,通过特征点的匹配可知  $p_1$  和  $p_2$  是同一个空间点  $P$  在前后两帧图像上的投影,但并不知道相机的位姿。代入初始值,  $P$  在后一帧图像上的投影  $p_2^{\wedge}$  与实际的  $p_2$  之间有一定的距离  $e$ 。所以调整相机的位姿,使得这个距离  $e$  变小。由于在这个过程中有很多个这样的点,最后每个点的误差往

往都不会精确为 0。

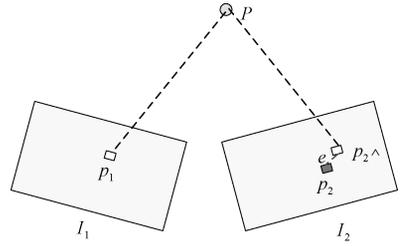


图6 重投影误差示意图

Fig. 6 Schematic diagram of reprojection error

图 6 中,空间点  $P$  齐次坐标为  $P = [X, Y, Z, 1]^T$ ,其在图像  $I_1$  中投影的像素坐标为  $p_1 = [u_1, v_1]^T$ ,在图像  $I_2$  中的重投影像素坐标为  $p_2^{\wedge} = [u_2', v_2']^T$ ,而空间点  $P$  在图像  $I_2$  中的观测值  $p_2 = [u_2, v_2]^T$ ,  $e = p_2 - p_2^{\wedge}$  表示重投影误差。理想的重投影过程用式(18)表达如下:

$$s_2 u_2 = K \exp(\xi^{\wedge}) P \quad (18)$$

其中,  $s_2$  表示空间点  $P$  在图像  $I_2$  所在相机坐标系的深度,  $K$  表示相机内参数,  $\exp(\xi^{\wedge})$  表示相机从图像  $I_1$  到图像  $I_2$  的姿态变换阵,也可以用  $T$  表示,  $\xi$  表示  $T$  对应的李代数。重投影值与真实值存在的误差定义如式(19)所示:

$$e_2 = u_2 - \frac{1}{s_2} K \exp(\xi^{\wedge}) p \quad (19)$$

一个相机姿态下观测到的特征点往往不止一个,假设有  $N$  个特征点,则构成求相机姿态  $\xi$  的最小二乘问题:

$$\xi^* = \arg \min_{\xi} \frac{1}{2} \sum_{i=1}^N \| e_i \|^2 = \arg \min_{\xi} \frac{1}{2} \sum_{i=1}^N \| (u_i - \frac{1}{s_i} K \exp(\xi^{\wedge}) p_i) \|^2 \quad (20)$$

## 2 实验结果和分析

本文实验平台采用笔记本电脑(联想小新 Air15),运行环境是 VirtualBox 虚拟机下的 Ubuntu16.04 操作系统。采用 Intel 的 RealSense 摄像头 D435i,如图 7 所示,该摄像头为全局快门,帧率为 20 fps,图像分辨率为 1280×720,相机基线为 5 cm。实验场景为室内,通过手持传感器,在实验场景中运动,进行位姿估计。

### 2.1 ORB 特征点的提取、匹配及丢失实验

实验过程中,在室内场景中取 3 组相邻两帧图像进行 ORB 特征点的提取和匹配,比较加入 RANSAC 算法前后的误匹配数量,如表 1 所示。

由表 1 可知,加入 RANSAC 算法前后 3 组数据的误匹配数量显著下降,图 8(a)为第 1 组数据中只有近似最近邻算法(即未删除误匹配)的效果,图 8(b)为第 1 组数



图 7 RealSense D435i

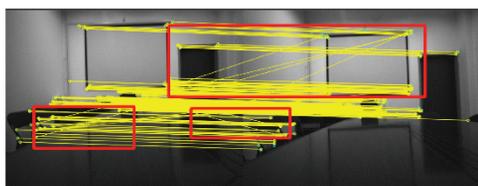
Fig. 7 RealSense D435i

据中在最近邻算法的基础上加入 RANSAC 算法的效果。在图 8(a)中可以看到用方框圈出的区域内存在较多的匹配错误,而在 8(b)的同一区域则不存在这些误匹配。因此可以看出结合了 RANSAC 算法匹配效果更好,从而有利于视觉里程计精度的提高。

表 1 加入 RANSAC 算法前后的误匹配数量

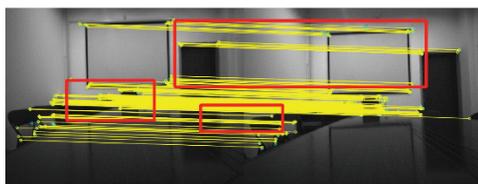
Table 1 The number of mismatches before and after adding the RANSAC algorithm

实验组数	仅使用近似最近邻算法的误匹配数	加入 RANSAC 算法的误匹配数
1	93	2
2	236	13
3	117	2



(a) 未删除误匹配

(a) Mismatched renderings were not deleted



(b) 删除误匹配

(b) Delete the wrong match

图 8 删除误匹配前后匹配效果

Fig. 8 The matching renderings before and after deleting the false match

手持相机进行缓慢线性运动,采集间隔为 1 s 的前后两帧图像,观察两帧图像特征点跟踪情况,3 组实验结果如表 2 所示。

由表 2 知,后一帧跟踪到的特征点数均大于前一帧跟踪到点数的 85%。第 1 组实验过程如图 9 所示,右侧

为间隔为 1 s 的后一帧图像,能够跟踪到较多的关键点。

表 2 缓慢线性运动中特征点跟踪情况

Table 2 Feature point tracking in slow linear motion

实验组数	前一帧图像的特征点数	后一帧图像跟踪到的特征点数
1	500	437
2	392	351
3	423	388



图 9 缓慢运动时特征点跟踪情况

Fig. 9 Feature point tracking when moving slowly

手持相机快速挥动模拟老人摔倒情况,采集间隔为 1 s 的前后两帧图像。观察两帧图像特征点跟踪情况,3 组实验结果如表 3 所示。

表 3 快速挥动相机运动中特征点跟踪情况

Table 3 Feature point tracking in fast waving camera movement

实验组数	前一帧图像的特征点数	后一帧图像跟踪到的特征点数
1	486	0
2	311	1
3	347	1

由表 3 知,后一帧跟踪到的特征点数几乎都接近 0,第 1 组实验过程如图 10 所示,右侧图像中几乎没有被跟踪到的关键点。说明在这种情况下间隔为 1 s 的两幅图像几乎没有共视点,不符合老人正常运动情况,此时则可以判断老人可能出现危险情况。



图 10 快速挥动相机时特征点跟踪情况

Fig. 10 The tracking of feature points when the camera is quickly swung

### 2.2 室内实际场景实验

在室内环境进行定位实验,主要进行直线往返运动和任意轨迹运动,从而对本文设计的视觉里程计的定位效果进行一个评估。

#### 1) 室内直线运动

实验过程中,手持相机保持水平高度不变直线行走

于距离固定的两点之间,取 6 组数据,定位起点和终点直线距离分别为 0.5、1、2、3、5、7 m。本实验中定义系统初始化成功后第一帧图像的相机坐标系为世界坐标系,相机坐标系定义如下:相机镜头面向方向作为  $z$  轴正方向,相机右边方向为  $x$  轴正方向, $y$  轴与  $x$ 、 $z$  构成右手坐标系向下,如图 11 所示。起点坐标为  $(0,0)$ ,终点坐标分别为  $(0,-0.5)$ 、 $(0,1)$ 、 $(0,2)$ 、 $(0,3)$ 、 $(0,-5)$ 、 $(0,7)$ ,单位为米。实验过程中采集的部分图像如图 12 所示。



图 11 相机坐标系示意图

Fig. 11 Schematic diagram of camera coordinate system

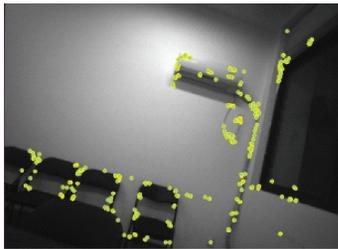


图 12 实验过程采集效果图

Fig. 12 Collection effect diagram during the experiment

对实验结果进行处理,定位结果如表 4 所示。由表 4 可以看到当载体进行不同长度的直线往复运动时,当运动长度在 2 m 以内,定位误差在 10 cm 以内。在 7 m 以内随着运行长度的增加,定位误差在 70 cm 以内且没有显著偏离。第 2、5 次实验相机运动轨迹分别如图 13 所示。

表 4 直线运动测量定位误差

Table 4 Linear motion measurement positioning error

运动过程	终点定位坐标	定位误差/m
第 1 次直线运动	$(0.002\ 6, -0.489\ 3)$	0.011 03
第 2 次直线运动	$(0.038\ 32, 0.968\ 7)$	0.049 5
第 3 次直线运动	$(0.143\ 5, 1.9)$	0.052 2
第 4 次直线运动	$(0.643\ 3, 2.973\ 5)$	0.643 8
第 5 次直线运行	$(-0.499\ 6, -4.928)$	0.504 8
第 6 次直线运动	$(0.612\ 4, 6.911)$	0.618 9

为进一步验证本文算法的有效性,将文献[16]提出的双目定位算法和较为经典的单目 ORBSLAM 算法[18]进行比较,以  $(0,0)$  为原点,  $(0,0.9)$ 、 $(0,2.5)$ 、 $(0,8)$  为

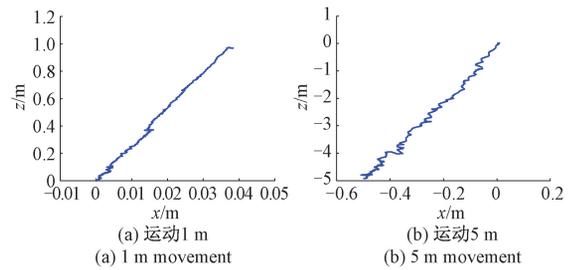


图 13 第 2、5 次直线运动相机轨迹估计结果

Fig. 13 The second and fifth linear motion camera trajectory estimation results

终点分别进行 3 组实验,比较 3 种算法的效果,结果如表 5 所示。

表 5 几种不同算法的效果对比

Table 5 Comparison of the effects of several different algorithms

算法	实际终点坐标/m	测得定位坐标/m	平均每帧处理时间/ms	误差/m
文献[16]算法	$(0,0.9)$	$(0.004\ 2, 0.872)$	59.06	0.028
	$(0,2.5)$	$(0.021\ 1, 2.552)$		0.056
	$(0,8)$	$(0.215\ 1, 8.265)$		0.341
单目 ORBSLAM	$(0,0.9)$	$(0.006\ 2, 0.954\ 1)$	30.57	0.054
	$(0,2.5)$	$(0.115\ 7, 2.783)$		0.306
	$(0,8)$	$(0.831\ 9, 9.271)$		1.519
本文算法	$(0,0.9)$	$(0.002\ 1, 0.851\ 4)$	38.61	0.049
	$(0,2.5)$	$(0.101\ 3, 2.329)$		0.199
	$(0,8)$	$(0.757\ 4, 7.503)$		1.071

由表 5 可知,文献[16]提出的基于改进 AD-Census 立体匹配算法的双目视觉里程计定位精度虽然优于本文算法,但由于增加的图像预处理操作和改进的立体匹配算法平均每帧数据处理时间可达 59.06 ms,一般相机帧率为 20 fps,可见此类精度的提升对定位的实时性存在一定的影响。单目视觉里程计相较本文算法不需要进行双目特征的提取和匹配,平均每帧数据处理时间略有下降,但单目视觉的尺度问题会给定位精度带来较大的影响。本文提出的算法其精度的提升对运算时间稍有影响,但小于平均每帧数据采集时间,因此符合实时定位的需求。

2) 室内任意轨迹运

在室内手持相机进行任意轨迹运动后回到起始点,采集的部分图像进行 ORB 特征点提取结果如图 14 所示。取 3 组数据,处理后定位结果如表 6 所示,定位起点均为  $(0,0,0)$ ,由表可知平均定位误差为 0.164 3 m。图 15 给出了第 1 组数据的里程计位姿估计结果轨迹,定位终点为  $(-0.037\ 65, -0.125\ 1, -0.010\ 99)$ ,单位为 m,定位误差为 0.131 1 m。

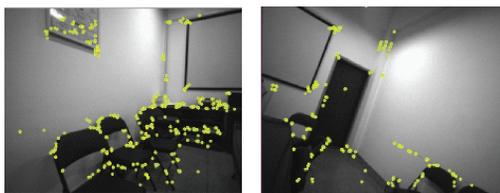


图 14 实验过程采集效果

Fig. 14 Effect picture collected during the experiment

表 6 任意轨迹运动测量定位误差  
Table 6 Arbitrary trajectory motion  
measurement positioning error

实验组数	定位终点坐标	定位误差/m
1	(-0.037 65, -0.125 1, -0.010 99)	0.131 1
2	(0.111 9, -0.275 7, 0.04)	0.189 5
3	(0.117 3, -0.121 8, 0.032 46)	0.172 2

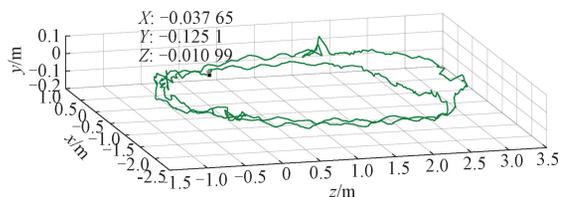


图 15 载体运动轨迹估计结果三维视图

Fig. 15 Three-dimensional view of the estimation result of carrier motion

在任意轨迹运动实验中  $x$  轴方向平均定位误差为  $0.0913\text{ m}$ ,  $z$  轴方向平均定位误差为  $0.0278\text{ m}$ , 误差均在厘米级别,  $y$  轴方向平均定位误差为  $0.1742\text{ m}$ , 为主要误差来源, 对于  $y$  轴方向上位置的波动幅度, 主要是由人体在行走过程中的上下波动引起的。结合载体在室内的直线往复运动和任意轨迹运动的实验结果可以看出, 本文设计的双目定位算法对室内环境具有良好的定位能力。

### 3 结 论

针对独居老人室内监测的迫切需求, 提出了一种基于双目视觉的定位算法, 通过特征提取、特征匹配和运动估计, 最终能够准确地定位老人在室内的位置, 重点对特征匹配进行了改进。一方面, 采用 RANSAC 算法消除了近似近邻法中存在的误匹配; 另一方面, 为了提高立体匹配的精度, 提出了一种基于 BRIEF 描述子的代价函数作为特征信息。在此基础上, 利用一定时间间隔内两帧图像的特征点跟踪情况来判断老年人是否处于摔倒危险中。通过 3 组实验验证了该方法的可行性。通过特征匹配实验, 可以直观地看出 RANSAC 算法能够有效地消除

误匹配; 通过特征匹配实验中行走和跌倒情况的对比, 验证了两幅图像在设计的时间阈值内的跟踪效果和跌倒危险性评估的可行性; 此外, 通过不同情况下的室内定位实验以及将本文算法与同类算法作比较, 验证了改进方法的有效性和准确性。但仍需根据本文的场景进行进一步的改进和优化, 如增加辅助定位等方法来进一步提高测量精度。

### 参考文献

- [1] 刘健. 基于 Wi-Fi 的室内定位技术研究[J]. 信息通信, 2018(2): 259-260.  
LIU J. Research on indoor positioning technology based on Wi-Fi [J]. Information and Communication, 2018(2): 259-260.
- [2] LI H. Low-cost 3d bluetooth indoor positioning with least square[J]. Wireless Personal Communications, 2014, 78(2): 1331-1344.
- [3] ALGHAMDI S, VAN SCHYNDEL R, KHALIL I. Accurate positioning using long range active RFID technology to assist visually impaired people[J]. Journal of Network and Computer Applications, 2014, 41: 135-147.
- [4] 于海洋, 曹俊, 纪怀猛. 基于 UWB 的室内导航系统研究[J]. 数字技术与应用, 2020, 38(3): 138-139, 142.  
YU H Y, CAO J, JI H M. Research on indoor navigation system based on UWB [J]. Digital Technology and Application, 2020, 38(3): 138-139, 142.
- [5] DAVISON A J, REID I D, MOLTON N D, et al. MonoSLAM: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.
- [6] 郭瑞峰, 贾榕. LK 光流法的多信息融合目标跟踪算法研究[J]. 现代电子技术, 2019, 42(18): 55-59.  
GUO R F, JIA R. Research on multi-information fusion target tracking algorithm based on LK optical flow method [J]. Modern Electronic Technology, 2019, 42(18): 55-59.
- [7] 管一凡. 融合双目视觉和惯性的定位技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.  
GUAN Y F. Research on positioning technology combining binocular vision and inertia [D]. Harbin: Harbin Institute of Technology, 2020.
- [8] MEI X, SUN X, ZHOU M, et al. On building an accurate stereo matching system on graphics hardware [C]. 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE, 2011: 467-474.
- [9] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

- [10] BAY H, TUYTELAARS T, GOOL L V. Surf: Speeded up robust features[C]. European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2006: 404-417.
- [11] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection [C]. European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2006: 430-443.
- [12] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF [C]. 2011 International Conference on Computer Vision. IEEE, 2011: 2564-2571.
- [13] SCHARSTEIN D, SZELISKI R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. International Journal of Computer Vision, 2002, 47(1): 7-42.
- [14] WANG D, LIM K B. A new segment-based stereo matching using graph cuts [C]. 2010 3rd International Conference on Computer Science and Information Technology. IEEE, 2010, 5: 410-416.
- [15] DI STEFANO L, MARCHIONNI M, MATTOCCIA S. A fast area-based stereo matching algorithm[J]. Image and Vision Computing, 2004, 22(12): 983-1005.
- [16] 赵冬青, 畅雅雯, 单彦虎, 等. 双目测量系统的室内定位与重建[J]. 激光杂志, 2022, 43(1): 19-23.
- ZHAO D Q, CHANG Y W, SHAN Y H, et al. Indoor localization and reconstruction of binocular measurement system[J]. Laser Magazine, 2022, 43(1): 19-23.

- [17] 赵文恺, 李刚. 基于双目相机的视觉里程计[J]. 计算机工程与设计, 2020, 41(4): 1133-1138.
- ZHAO W K, LI G. Visual odometry based on binocular camera[J]. Computer Engineering and Design, 2020, 41(4): 1133-1138.
- [18] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: a versatile and accurate monocular SLAM system [J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.

### 作者简介



张艳, 2020年于安徽建筑大学获得学士学位, 现为南京理工大学测试计量技术及仪器专业硕士研究生。

E-mail: zy321@njjust.edu.cn

**Zhang Yan** received her B. Sc. degree in 2020 from Anhui Jianzhu University, now she is a M. Sc. candidate in Measuring and Testing Technologies and Instruments at Nanjing University of Science and Technology.



王宇(通信作者), 现为南京理工大学副研究员, 主要研究方向为 MEMS 惯性导航技术、组合导航技术。

E-mail: wangyu78@njjust.edu.cn

**Wang Yu** (Corresponding author) is an associate researcher at Nanjing University of Science and Technology. His main research interests include MEMS inertial navigation technology, integrated navigation technology.