

基于 ORB 特征的视觉里程计算法优化

林付春, 刘宇红, 周进凡, 马治楠, 何倩倩, 王曼曼, 张荣芬*

贵州大学大数据与信息工程学院, 贵州 贵阳 550025

摘要 针对在动态环境下受运动物体影响而不能准确进行机器人运动估计的问题, 提出一种基于 ORB 特征区域分割的视觉里程计算法。利用相邻区域特征点三维空间距离不变性, 对提取的特征点进行区域分割, 将图像中运动物体产生的特征点与静态背景的特征点分割开, 去除动态物体特征点的影响, 再进行相机的位姿估计, 从而去除场景中动态物体的干扰。实验结果表明, 基于 ORB 特征区域分割的视觉里程计算法能够实时地在动态和静态环境中进行相机的位姿估计, 具有很高的稳健性和精度。

关键词 机器视觉; 动态环境; 视觉里程计; 区域特征分割; 位姿估计

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.211507

Optimization of Visual Odometry Algorithm Based on ORB Feature

Lin Fuchun, Liu Yuhong, Zhou Jinfan, Ma Zhinan, He Qianqian,

Wang Manman, Zhang Rongfen*

College of Big Data and Information Engineering, Guizhou University, Guiyang, Guizhou 550025, China

Abstract Robot movement cannot be accurately estimated because of the impact of moving objects in a dynamic environment. Therefore, this study proposes a visual odometry algorithm based on ORB (Oriented FAST and Rotated BRIEF) feature regional segmentation. Further, using the distance invariance of the feature points in the adjacent regions in a three-dimensional space, the extracted feature points are segmented and the feature points generated by the moving objects in the image are separated from the feature points in the static background, and influences of dynamic object feature points are removed. Subsequently, the position of the camera can be estimated, thereby removing the interference caused by the dynamic objects in a scene. The experimental results show that the visual odometry algorithm based on ORB feature regional segmentation can perform real-time pose estimation in both dynamic and static environments with good robustness and high precision.

Key words machine vision; dynamic environment; visual odometer; regional feature segmentation; pose estimation

OCIS codes 150.0155; 150.1135; 150.1488

1 引言

近年来,随着互联网与人工智能技术的迅速发展,移动机器人技术也取得了飞跃的进展。但就室内移动机器人的定位导航和避障而言,由于室内 GPS 信号无法到达、激光雷达价格成本高昂,室内移动机器人很难实现智能化的自主导航避障^[1-2]。而采用基于视觉的同时定位与地图构建(SLAM)技术来估计相机的运动并构建室内地图,可实现移动

机器人室内自主导航避障的智能化^[3]。然而在现有的视觉 SLAM 中,移动机器人大部分是在静态、刚体、光照变化不明显和没有人为干扰的场景下实现实时运行并获得比较高的精度^[4]。当动态物体出现在摄像机的视觉图像中,特别是当摄像机采集的图像被动态物体占据大部分时,动态物体就会成为图像处理的干扰因素,严重影响相机位姿的估计,使得机器人无法准确获取自身位置信息和周围环境信息,从而不能在动态环境中构建实时性好、精度高的

收稿日期: 2019-04-03; 修回日期: 2019-04-25; 录用日期: 2019-05-06

基金项目: 贵州省科技计划(黔科合平台人才[2016]5707)、贵州省科技计划(黔科合基础[2019]1099)

* E-mail: rfzhang@gzu.edu.cn

地图。Nikoohemat 等^[5]通过轨迹计算将相邻时刻的激光数据融合在一起,并根据环境中的拓扑结构对三维空间的物体赋予语义信息,但该方法获取的语义激光数据不具有实时性,而且产生的语义信息过于简单,不能解决动态场景下 SLAM 的定位问题。Charles 等^[6]利用深度学习的三维点云分类模型,获取三维点云的语义信息,但目前三维点云分类训练集较少,导致模型泛化能力较弱,因而在实际应用中难以实现。曹军^[7]在动态环境下将视觉语义与激光雷达相结合进行 SLAM,但该方法需要将视觉与雷达相结合,硬件成本增加,且需计算语义信息,增加了计算量,此外,该方法对相机和激光雷达的标定参数十分敏感。王召东等^[8]采用深度学习进行语义分割,将图像中动态人的特征点去除,然后再进行相机位姿估计,该方法需要 GPU 支持,并且只能分割人的动态点,对其他动态物体失去作用。

为消除图像中大部分动态物体的影响,本文在传统 ORB(Oriented FAST and Rotated BRIEF)算法的基础上^[9],采用德洛内三角剖分法对提取的特征点进行区域分割,将图像中动态物体产生的特征点与静态背景的特征点分割开,采用静态背景特征点进行摄像机的位姿估计,从而实现相机在动态和静态环境下实时准确地获取相机位置和姿态估计。

2 视觉里程计原理

视觉里程计(VO)的任务是通过相邻关键帧图像估计相机运动的轨迹,构建环境局部地图。而在三维世界中,假设 k 时刻相机相对于世界坐标系的位姿为 C_k ,那么 k 与 $k+1$ 时刻之间的相机刚体位姿变换为 $T_{k,k+1} \in R_{4 \times 4}$, $T_{k,k+1} = \begin{bmatrix} R_{k,k+1} & t_{k,k+1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$, 其中 $R_{k,k+1} \in R_{3 \times 3}$ 表示旋转矩阵, $t_{k,k+1} \in R_{3 \times 1}$ 表示平移向量, $\mathbf{0}$ 表示三维 $\mathbf{0}$ 向量。 $k+1$ 时刻相机的位姿 $C_{k+1} = C_k(T_{k,k+1})^{-1}$ 。VO 利用相机拍摄的图像计算出相机位姿变换 T ,再由当前位姿与相机变换计算获得下一刻的运动,并可恢复场景的空间结构。只要把相邻时刻的运动“连”起来,就能构成相机的运动轨迹^[10-11]。

传统的 VO 算法包括图像预处理、特征点检测、特征点提取、特征点匹配、特征点筛选、多点透视成像(PnP)及迭代最近点(ICP)过程,其流程图如图 1 所示。为实现 VO 在动态环境中高精度实时运行,在传统 VO 算法的基础上,对提取的特征点进行区

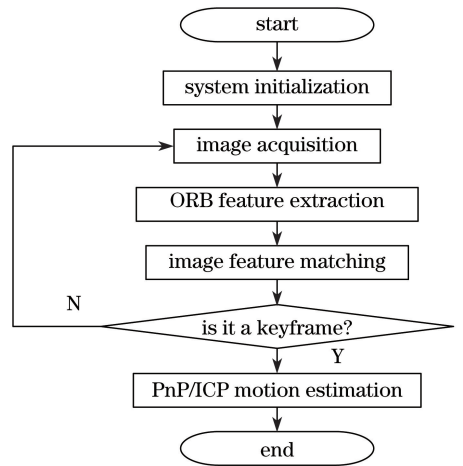


图 1 传统的 VO 算法流程图

Fig. 1 Flowchart of traditional VO method

域分割,将相机图像中静态区域特征点和动态区域特征点进行分离,最后采用静态区域特征点对相机进行位姿估计。优化后的 VO 算法流程图如图 2 所示。

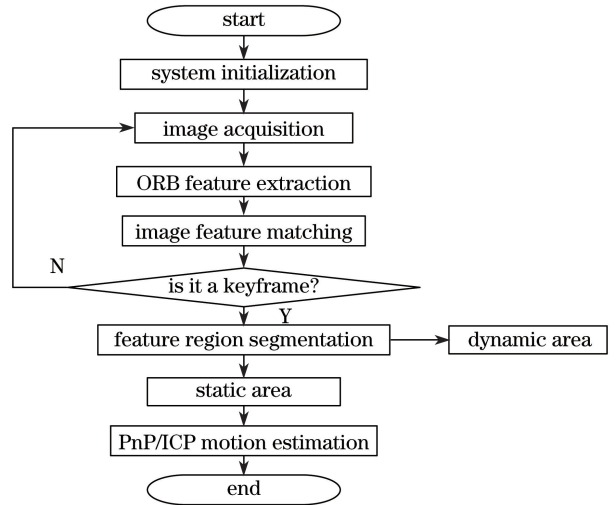


图 2 基于 ORB 改进的 VO 流程图

Fig. 2 Improved VO flowchart based on ORB

3 特征提取与匹配

3.1 特征点选择

现今主流的特征提取算法为 ORB 算法,尺度不变特征变换(SIFT)算法,加速稳健特征(SURF)算法,这 3 种算法都具有尺度不变、旋转不变的特性,能克服关键帧技术带来图像间较大变化的影响^[12-13]。SIFT 算法是 3 种算法中精度最高的算法,但其计算量也相当复杂,在 CPU 上不能实时运行。SURF 算法是 SIFT 算法的改进版,其实时性有了很大的提升,但利用其旋转不变性求主方向时太过于依赖局部区域像素的梯度方向,对特征匹配影响

很大^[14]。ORB算法利用FAST提取特征点,以灰度质心法实现旋转不变性,通过构建图像金字塔实现尺度不变性,精度和效率最高^[15]。同时ORB算法也是这3种算法中实时性最好的算法,其速度是SIFT的100倍,是SURF的10倍,综合考虑,采用ORB算法来进行特征提取。

3.2 ORB特征点

ORB特征由关键点和描述子组成。它的关键点称为“Oriented FAST”,是一种改进的FAST角点,它的描述子是BRIEF(Binary Robust Independent Elementary Features)。

FAST的特征点是一种角点,主要是以很快的速度对图像中局部像素灰度变化明显的地方进行检测。它的主要原理是,图像中如果一个像素与它邻域的像素值的差别较大,那么它很可能是一个角点,如图3所示。而对比于其他的角点检测算法,FAST仅仅需要对比周围像素亮度值的大小,计算很方便。

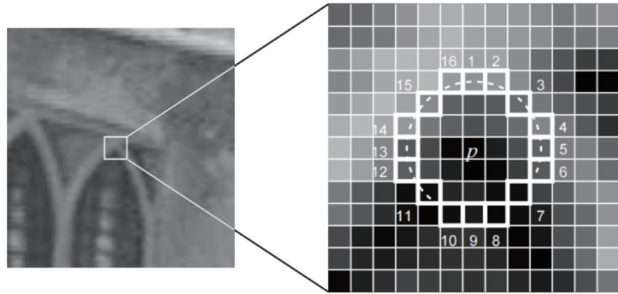


图3 FAST特征点

Fig. 3 FAST feature points

ORB算法特征点的检测过程为:

- 1) 在图像中选取像素 p , 假设它的亮度为 I_p 。
- 2) 设置一个阈值 T (比如 I_p 的 20%)。
- 3) 以像素 p 为中心, 选取半径为 3 的圆上的 16 个像素点^[4]。
- 4) 假如选取的圆上有连续 N 个点的亮度大于 $I_p + T$ 或小于 $I_p - T$, 那么像素 p 可以被认为 是特征点 (N 通常取 12, 即 FAST-12。其他常用的 N 取值为 9 和 11, 分别被称为 FAST-9, FAST-11)。
- 5) 重复 1)~4), 依次对图像中各个像素进行相同的操作。

FAST特征点只比较像素间的明暗差异,速度 比较快,但是提取的特征点数量也很多,不利于对 特征点的描述与匹配。所以,在ORB中对FAST 算法进行优化改进。首先,可以设置最终需提取

的角点数量 N , 计算FAST角点的Harris响应值, 对计算的Harris值进行排序, 选取最大的前 N 个 作为需要的特征点。其次, 针对FAST不具有方 向和尺度信息的缺点, ORB对FAST角点构建图 像金字塔, 实现特征点的尺度不变性, 同时采用灰 度质心法实现对FAST角点方向的旋转不变性。

灰度质心法的计算过程为:

- 1) 对一个图像块 A 进行矩定义, 即

$$m_{ts} = \sum_{x', y' \in A} x' y' I(x', y'), \quad t, s \in \{0, 1\}, \quad (1)$$

式中: m_{ts} 为图像块的矩; t, s 都取 0 时表示零阶矩, t 和 s 一个取值为 0, 另一个取值为 1 时, 表示一阶矩; $I(x', y')$ 为 (x', y') 的像素值; A 为图像块。

- 2) 计算出矩后, 利用矩计算出图像块的质 心, 即

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right), \quad (2)$$

式中, C 为图像块质心, m_{00}, m_{10}, m_{01} 为(1)式计算 的矩。

因此ORB中的FAST角点具有了尺度不变性和 旋转不变性, 大大提高了在图像特征点提取、特征 点描述及特征点匹配中的稳健性。

提取特征点之后, 采用改进的BRIEF二进制 描述子对特征点进行描述。BRIEF使用随机选点 进行比较, 实现速度非常快, 同时采用128维的二 进制向量表达, 易于存储, 适用于实时的图像 匹配^[16]。

得到描述子后采用最简单的特征匹配方法, 即暴力匹配, 对图像进行匹配。计算当前帧的每 一个特征点与参考帧的每一个特征点的汉明距 离, 然后排序, 选取最近的一个特征点作为匹配 点。同时对匹配点间设定阈值, 由参考文献[4]可 知, 在工程应用中, 根据经验, 通常汉明距离小于 所有特征点最小距离的2倍时, 特征点才进行匹 配, 再对匹配的特征点进行筛选, 从而有效地减小 匹配的误差。

4 特征区域分割

当相机处于静态环境时, 传统VO算法可以很 好地进行相机的位姿估计。当相机采集的图像内容 被场景中运动物体占据, 即提取的特征点在运动物 体上时, VO对相机位置的估计就会出现很大的误 差。基于此, 利用德洛内三角剖分算法对动态特征 点与静态特征点进行分割。

4.1 德洛内三角剖分

德洛内三角剖分算法是计算机视觉与计算机图形学的桥梁。在计算机视觉中,三角剖分可用于标记空间目标特征或运动场景跟踪,或两个不同摄像机的场景匹配,以获得深度信息^[17]。

为能很好地体现出特征点在真实空间的位置关系,利用相机坐标系下的三维空间位置投影坐标的特征点组成点集。德洛内三角剖分图形算法,将点集中的特征点组成一个个三角形连接关系,使每一个三角形的外接圆都不包含点集中的其他点,并且使每一个三角形最小内角的角度值最大,从而建立只有相邻特征点才会被连接的稀疏网络结构^[18]。特征点被当作网络结构的顶点,相邻特征点的连接称为边。利用建立的稀疏网络,计算相邻特征点空间距离是否发生变化,即边的长度是否发生变化。一对特征点的坐标为 $\mu_1(x_1, y_1, z_1), \mu_2(x_2, y_2, z_2)$, $x_1, y_1, z_1, x_2, y_2, z_2$ 分别为特征点的坐标值,其距离的度量采用欧氏距离 $D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$, 其中, z 为两个特征点的距离, $z = fb/d$, f 为相机焦距, b 为

基线, d 为双目视差。 x 为特征点横坐标值, $x = \frac{z(u - c_x)}{f_x}$, y 为特征点纵坐标值, $y = \frac{z(v - c_y)}{f_y}$, (u, v) 为特征点在图像平面上的像素坐标, (c_x, c_y) 为相机光轴穿过图像平面的主点坐标, (f_x, f_y) 分别为 x 轴和 y 轴方向上的相机焦距。相机在三维空间场景中进行欧氏变换,其任意两点间的距离不会随相机的运动发生变化。若相邻两个特征点有一个是动态特征点,而另一个是静态特征点,当特征点和相机发生相对变化时,两个特征点之间的距离就会发生变化。相邻特征点间距变化特性示意图如图 4 所示。

在与参考帧对比中,当前帧中,静态背景中的特征点 Q'_1, Q'_2 的距离 $|Q'_1Q'_2|$ 未发生变化,而特征点 Q'_1 与特征点 Q'_3 , 特征点 Q'_2 与特征点 Q'_3 的距离却发生了比较大的变化,即 $|Q'_1Q'_3|, |Q'_2Q'_3|$ 的值发生了比较大的变化,去掉 Q'_1 与 Q'_3 相连边和 Q'_2 与 Q'_3 相连的边后,特征点 Q'_1, Q'_2, Q'_3 被分为 $\{Q'_1, Q'_2\}, \{Q'_3\}$ 两个集合。其中,物体运动后的 Q'_1, Q'_2, Q'_3 对应参考帧 Q_1, Q_2, Q_3 特征点。

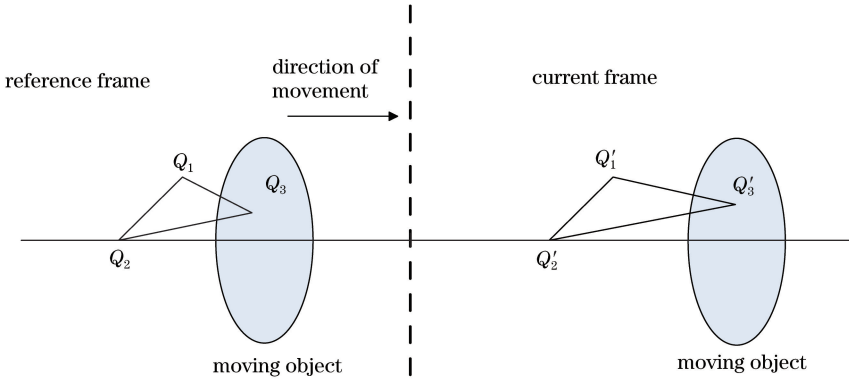


图 4 相邻特征点间距变化特性示意图

Fig. 4 Diagram of variation of adjacent feature point spacing

4.2 静态背景区域提取

利用德洛内三角剖分建立图形结构后,在参考帧中遍历所有的边,计算每一个边的空间距离,同时计算当前帧中对应特征点之间的三维空间距离。如果计算得到的当前帧和参考帧对应特征点的距离变化超过一定的阈值,则把该边删除。然后通过深度优先搜索算法(DFS),获得图中的各个连通区域^[19]。通常来说,静态背景在图像上占据的空间面积要比动态物体在图像上占据的空间面积大,特征点数目也比较多。因此,首先计算区域面积,选取面积最大的 3 个区域作为候选区域,同时为特征点数目设置一个阈值。然后计算区域内的特征点数目是

否达到设定阈值,若未达到阈值,则将该区域舍去;若达到阈值,选取面积最大的区域作为静态背景区域,用该区域的特征点来对相机进行运动估计。

5 关键帧提取与状态估计

5.1 关键帧技术

采用关键帧技术一方面是为了减小 VO 前端累计误差,另一方面是为了更好地对特征区域进行分割。关键帧技术增加了当前帧与参考帧的距离,使得运动物体与静态背景发生的相对运动变明显,特征点间的距离也变大,会让区域分割的效果变得更好^[20]。当估计的相机位姿相对于关键帧的变化超

过设置的阈值时,将当前帧设置为关键帧。采用两帧之间的距离度量相机位姿与关键帧变化,即计算两帧之间的位移与旋转的范数和。两个关键帧之间要保证拥有足够大的重叠面积和足够多的共同特征点。为提高算法的稳健性,若对其中的一帧位姿估计失败,设定其前一帧为关键帧。

5.2 相机运动估计

在经过区域分割得到静态背景的特征点后,对特征点采用非线性的 PnP 方法进行相机运动估计。将一帧图像的 3D 特征点重投影到另一帧图像上,最小化像素平面的误差,以此来进行相机位姿优化估计^[21]。如图 5 所示, P 点为世界坐标系中的特征点, P_1 为 P 点在参考帧图像上形成的像素坐标, P_2 为 P 点在当前帧中对应匹配特征点形成的像素坐标, e 为通过投影计算出的误差值。

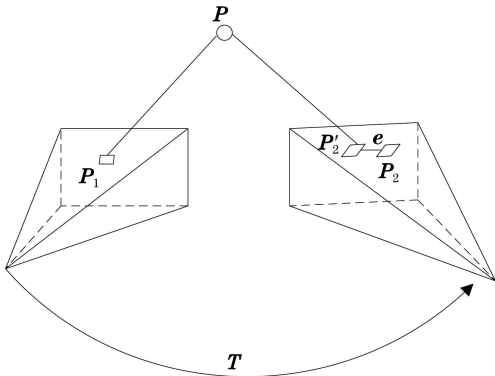


图 5 相机最小重投影误差图

Fig. 5 Minimum re-projection error of camera

把误差求和构建最小二乘法问题寻找相机位姿。目标函数为

$$\xi = \arg \min_T \frac{1}{2} \sum_{i=1}^n \| P_i - \tau(P_i, T) \|^2, \quad (3)$$

式中: i 的取值范围为 $1 \sim n$, n 为整数; P_i 为第 i 个特征点三维世界的坐标; T 为相机变换的旋转矩阵; $\tau(P_i, T)$ 为相机变换的计算函数,具体表达式为 $\tau(P_i, T) = KTP_i$,其中 K 为相机内部参数。

6 实验及结果分析

本实验在 CPU 为 INTEL 酷睿 I7-7500U、2.7 GHz、12 G 运行内存的个人计算机上进行。软件平台为 Ubuntu16.04 系统,带有 Opencv3.2 版本,并且编译了 PCL、Eigen 等库。

实验采用 ORB 算法进行特征检测,获取的图像特征点如图 6 所示,对得到的特征点进行筛选并进行匹配,如图 7 所示。采用德洛内三角剖分对当前帧与参考帧建立三角网络,如图 8 所示。

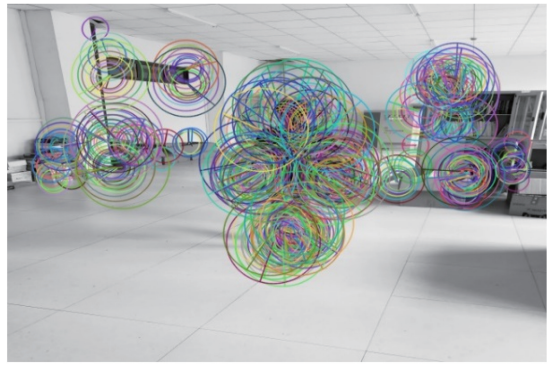


图 6 ORB 特征点提取

Fig. 6 ORB feature point extraction



图 7 ORB 特征点匹配

Fig. 7 ORB feature point matching

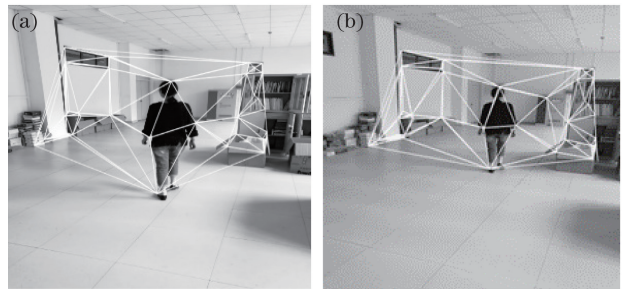


图 8 德洛内三角剖分建立三角网络图。(a)参考帧三角网络;(b)当前帧三角网络

Fig. 8 Triangular networks established by Delaunay triangulation. (a) Triangle network of reference frame; (b) triangle network of current frame

采用公开数据集 KITTI 对本文优化的 SLAM 算法进行评估。采用本文优化的算法在部分 KITTI 数据集图像序列上进行测试,实验中对本文算法与传统 ORB-SLAM 算法的测试结果进行比较,通过文献[22]对其算法复现,通过相机关键帧轨迹与真实的相机关键帧轨迹的对比得到部分序列 01、05、07、09 的均方根误差,如表 1 所示。

KITTI 数据集中包含道路上的车辆和行人等动态物体。本文算法在相机位姿轨迹上得到了一定程度的提升, X 表示在第一个序列中传统 ORB-SLAM 失效,而本文算法的误差为 15.66 m,对于其他图像序列关键帧轨迹均方根误差,本文算法都有提升,体现优化后算法有更高的准确性。

表1 相机轨迹均方根误差

Table 1 Root mean square error of camera track

Image sequence	KITTI01	KITTI05	KITTI07	KITTI09
Scene scale / (m×m)	1157×1827	479×426	191×209	465×568
Number of key frames	1101	2761	1101	1591
ORB-SLAM error / m	X	8.23	3.36	7.62
Error of proposed algorithm / m	15.66	7.42	3.14	6.58

将本文算法和 ORB-SLAM 在部分 EuRoC 数据集图像序列上进行测试,比较本文算法与传统 ORB-SLAM 算法对相机位姿估计的精度。本文优

表2 相机位姿估计误差表

Table 2 Error table of camera pose estimation

Algorithm	RMSE / m	Mean / m	Median / m	Min / m	Max / m
ORB-SLAM	0.087274	0.082128	0.075679	0.021258	0.164835
Proposed algorithm	0.060309	0.062565	0.067717	0.024182	0.163003

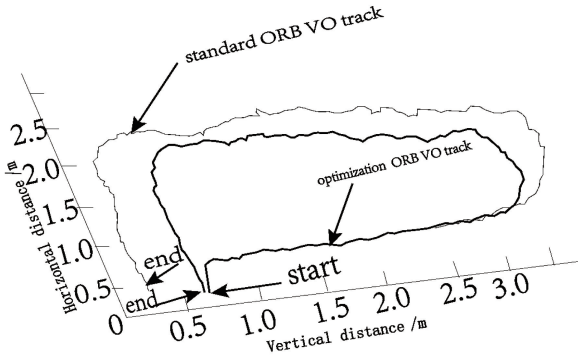


图9 ORB前端相机轨迹

Fig. 9 ORB front-end camera track

本次回环实验中相机在大约 1 min 的时间内运动了 12 m 的距离。在实验中,未经验过动态区域特

表3 不同场景下相机运动误差

Table 3 Camera motion errors in different scenes

Algorithm station	Average error of improved ORB		Average error of standard ORB		Reduction ratio of average error	
	Average distance error / m	Average angle error / (°)	Average distance error / m	Average angle error / (°)	Reduction ratio of distance error / m	Reduction ratio of angle error / (°)
Scene 1	0.08	2.1	0.71	11.8	88.73	82.20
Scene 2	0.13	4.2	0.95	18.3	86.32	77.05
Scene 3	0.52	9.3	1.66	20.6	68.67	54.84

由表3数据可以发现,在 scene 1 和 scene 2 中,改进的 ORB 视觉里程计算法对比标准 ORB 算法,距离和角度平均误差率减小了 80% 以上。即使在 scene 3 下,改进的 ORB 视觉里程计算法的距离和角度误差率减少了 68.67% 和 54.84%。改进的

化算法与传统 ORB-SLAM 算法得到的关键帧相机轨迹与真实轨迹的误差如表 2 所示。

由表 2 可知,在没有动态物体的环境中,本文优化算法的 RMSE 和 Mean 值比 ORB-SLAM 算法小,即优化后算法的精度优于传统 ORB-SLAM 算法。

真实实验室场景中,一个人手拿双目相机在实验室中运动,相机的正前方有一个人同样在运动,系统实时采集视频图像,并对获取的数据进行处理。将相机运动的起点和终点设置为同一个地点,从而人为地设置闭环。采用经典 ORB 前端和经过优化的 ORB 前端进行对比,对比的相机轨迹如图 9 所示。通过计算相机轨迹起点和终点的距离与角度偏差,来衡量本文优化算法的精度。

征分割的 ORB 视觉里程计系统中,相机的起点和终点产生 0.75 m 的距离误差和 11.8° 的角度误差。经过动态区域分割的 ORB 视觉里程计系统中,相机的起点和终点产生 0.08 m 的距离误差和 2.1° 的角度误差。在本文实验平台环境下,距离和角度的误差分别减小了 89.33% 和 82.20%,基本实现动态场景下实时准确的相机位姿估计。

为检验本文算法在不同规模动态场景中的效果,分别在场景一(小规模动态场景,相机前有 1 个人运动)、场景二(中规模动态场景,相机前有 2 个人运动)、场景三(大规模动态场景,相机前有 5 个人运动)3 种场景中各进行 5 次实验。本文算法的距离平均误差和角度平均误差如表 3 所示。

ORB 视觉里程计算法在室内动态环境下对相机姿态估计有更好的效果,具有更好的有效性和稳健性。

7 结 论

提出基于 ORB 特征区域分割的视觉里程计算

法,该算法首先将动态特征点与静态特征点分开,然后仅仅利用静态特征点对相机位姿进行估计,最后通过最小化重投影误差对相机位姿进行优化。本文算法在数据集上测试的结果表明,在动态环境中优化后的算法比传统算法精度高。在真实实验场景测试的结果表明,在动态环境中本文算法比标准ORB-SLAM的精度至少提高50%。综上,本文优化视觉里程计对机器人在动态环境中进行地图构建有一定程度的优化,对提高移动机器人的智能化水平具有重要意义。

参 考 文 献

- [1] Xiong W, Xu Y L, Cui Y Q, *et al.* Geometric feature extraction of ship in high-resolution synthetic aperture radar images [J]. *Acta Photonica Sinica*, 2018, 47(1): 0110001.
熊伟, 徐永力, 崔亚奇, 等. 高分辨率合成孔径雷达图像舰船目标几何特征提取方法 [J]. *光子学报*, 2018, 47(1): 0110001.
- [2] Huo J, Li Y H, Yang M. Measurement and error analysis of moving target pose based on laser projection imaging [J]. *Acta Photonica Sinica*, 2017, 46(9): 0912001.
霍炬, 李云辉, 杨明. 激光投影成像式运动目标位姿测量与误差分析 [J]. *光子学报*, 2017, 46(9): 0912001.
- [3] Li M. Research on visual odometer method of outdoor mobile robot based on stereo camera [D]. Nanjing: Southeast University, 2015.
李孟. 基于立体相机的室外移动机器人视觉里程计方法研究 [D]. 南京: 东南大学, 2015.
- [4] Gao X, Zhang T. Visual SLAM XIV: from theory to practice [M]. Beijing: Electronics Industry Press, 2017.
高翔, 张涛. 视觉 SLAM 十四讲: 从理论到实践 [M]. 北京: 电子工业出版社, 2017.
- [5] Nikoohemat S, Peter M, Elberink S O, *et al.* Exploiting indoor mobile laser scanner trajectories for semantic interpretation of point clouds [C] // ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, September 18-22, 2017, Wuhan, China. New York: IEEE, 2017, IV-2/W4: 355-362.
- [6] Charles R Q, Su H, Mo K C, *et al.* PointNet: deep learning on point sets for 3D classification and segmentation [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 77-85.
- [7] Cao J. Research on laser SLAM sensing method based on visual semantics in dynamic scene [D]. Guangzhou: Guangdong University of Technology GDUT, 2018.
曹军. 动态场景下融合视觉语义的激光 SLAM 感知方法研究 [D]. 广州: 广东工业大学, 2018.
- [8] Wang Z D, Guo C. An improved ORB_SLAM2 in dynamic scene with semantic segmentation [J]. *Journal of Dalian Maritime University*, 2018, 44(4): 121-126.
王召东, 郭晨. 一种动态场景下语义分割优化的 ORB_SLAM2 [J]. *大连海事大学学报*, 2018, 44(4): 121-126.
- [9] 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, 2011: 2564-2571.
- [10] Lin Z L, Zhang G L, Yao E L, *et al.* Stereo visual odometry based on motion object detection in the dynamic scene [J]. *Acta Optica Sinica*, 2017, 37(11): 1115001.
林志林, 张国良, 姚二亮, 等. 动态场景下基于运动物体检测的立体视觉里程计 [J]. *光学学报*, 2017, 37(11): 1115001.
- [11] Li Z, Zhou W H, Liu J L. Stereo visual odometry from disparity space in dynamic environments [J]. *Journal of Zhejiang University(Engineering Science)*, 2008, 42(10): 1661-1665.
李智, 周文晖, 刘济林. 动态场景下基于视差空间的立体视觉里程计 [J]. *浙江大学学报(工学版)*, 2008, 42(10): 1661-1665.
- [12] Bay H, Tuytelaars T, van Gool L. SURF: speeded up robust features [M] // Leonardis A, Bischof H, Pinz A. European conference on computer vision-ECCV 2006. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2006, 3951: 404-417.
- [13] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [14] Li J, Cao Y P, Chen C, *et al.* On-line three-dimension measurement method based on SURF algorithm [J]. *Acta Photonica Sinica*, 2017, 46(9): 0912002.
李建, 曹益平, 陈澄, 等. 基于 SURF 算法的在线三维测量方法 [J]. *光子学报*, 2017, 46(9): 0912002.

- [15] Li Y D, Xu X P, Wang J Q. Feature extraction based on pyramid match kernel algorithm with adaptive partitioning [J]. *Acta Photonica Sinica*, 2017, 46(12): 1210001.
李艳获, 徐熙平, 王佳琪. 基于自适应分块金字塔匹配核的特征提取算法[J]. *光子学报*, 2017, 46(12): 1210001.
- [16] Calonder M, Lepetit V, Strecha C, *et al.* BRIEF: binary robust independent elementary features[M] // Daniilidis K, Maragos P, Paragios N. *European conference on computer vision-ECCV 2010. Lecture notes in computer science*. Berlin, Heidelberg: Springer, 2010, 6314: 778-792.
- [17] Lee D T, Schachter B J. Two algorithms for constructing a Delaunay triangulation [J]. *International Journal of Computer & Information Sciences*, 1980, 9(3): 219-242.
- [18] Peng Z. Research on vision-based self-motion estimation and environment modeling method in dynamic environment [D]. Zhejiang: Zhejiang University, 2017.
彭真. 动态环境下基于视觉的自运动估计与环境建模方法研究[D]. 浙江: 浙江大学, 2017.
- [19] Gong J H. Depth priority algorithm and its improvement [J]. *Modern Electronics Technique*, 2007, 30(22): 90-92.
龚建华. 深度优先搜索算法及其改进[J]. *现代电子技术*, 2007, 30(22): 90-92.
- [20] Zhu D M, Wang Z Q. Extraction of keyframe from motion capture data based on motion sequence segmentation[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2008, 20(6): 787-792.
朱登明, 王兆其. 基于运动序列分割的运动捕获数据关键帧提取[J]. *计算机辅助设计与图形学学报*, 2008, 20(6): 787-792.
- [21] Gargallo P, Prados E, Sturm P. Minimizing the reprojection error in surface reconstruction from images [C] // 2007 IEEE 11th International Conference on Computer Vision, October 14-21, 2007, Rio de Janeiro, Brazil. New York: IEEE, 2007: 9849037.
- [22] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147-1163.