

DOI: 10.13382/j.jemi.B2003040

# 基于运动矢量的改进视觉 SLAM 算法\*

孙龙龙<sup>1,2</sup> 江明<sup>1,2</sup> 焦传佳<sup>1,2</sup>

(1. 安徽工程大学 电气工程学院 芜湖 241000; 2. 安徽工程大学 高端装备先进感知与智能控制教育部重点实验室 芜湖 241000)

**摘要:**针对移动机器人运行场景中出现运动物体时,视觉同时定位与地图构建(SLAM)算法位姿估计误差大且构建地图不一致的问题,提出了一种基于特征点运动矢量的改进视觉 SLAM 算法。首先,引入基于特征点运动矢量的运动点检测算法。通过结合初始相机位姿,计算图像特征点的运动矢量,并使用期望最大化方法求解运动矢量角度的高斯混合模型参数,通过结合前一帧的运动点检测结果,从而区分当前图像中的运动特征点;其次,基于运动点检测结果,对当前帧相机位姿进行优化;再次,通过设置图像预处理环节,剔除运动点占比较大和与前一帧相似性较高的图像,提高闭环检测算法的计算效率;最后,使用剔除动态点后的图像特征点对场景进行描述,并改进单个节点处图像间相似性得分计算函数,经过闭环确认后,得到正确闭环。数据集实验表明,所提算法具有较高的位姿估计精度和较好的鲁棒性,同时能有效检测场景中闭环的存在,且建图效果较好。

**关键词:**同时定位与地图构建;运动矢量;运动点检测;闭环检测

**中图分类号:** TP242      **文献标识码:** A      **国家标准学科分类代码:** 510.40;510.80

## Improved visual SLAM algorithm based on the motion vector

Sun Longlong<sup>1,2</sup> Jiang Ming<sup>1,2</sup> Jiao Chuanjia<sup>1,2</sup>(1. School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China;  
2. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Anhui Polytechnic University, Wuhu 241000, China)

**Abstract:** Aiming at the problem that the simultaneous localization and mapping (SLAM) algorithm has a large pose error and inconsistent map construction when a moving object appears in mobile robot's operating scene, an improved visual SLAM algorithm based on feature point motion vector is proposed. Firstly, the algorithm of motion points based on feature point motion vector is introduced. The motion vector can be calculated by combining the initial camera pose, and the Gaussian mixture model parameters of its angle are solved by using the expectation maximization method. And the motion point detection result of the previous frame is used to distinguish motion points in the current image. Secondly, the camera pose will be optimized based on results of the motion point detection. Then the image is pre-processed, and images with a number of motion points and higher similarity to the previous frame will be eliminated, which can improve the calculation efficiency of loop closure detection. Finally, the scene is described by using feature points after excluding dynamic features, and the similarity score calculation function of two images at a single node is improved. After loop closure confirmation, the correct loop is obtained. The datasets experimental results show that the improved algorithm has better robustness and higher accuracy in the pose estimation. And it can effectively detect the existence of loops in the scene and has a good mapping.

**Keywords:** simultaneous localization and mapping; motion vector; motion point detection; loop closure detection

## 0 引言

随着机器人技术的发展,移动机器人逐渐实现了自主运动,而机器人同时定位与地图构建(simultaneous localization and mapping, SLAM)技术被认为是真正实现机器人完全自主运动核心的部分<sup>[1]</sup>。最初,通常使用激光 SLAM 技术,即通过激光雷达获取周围的障碍物信息,但受限于激光雷达昂贵的价格,难以进行广泛的应用<sup>[2]</sup>。而基于相机的视觉 SLAM 技术相比激光 SLAM<sup>[3]</sup>具有成本低、功耗低、小型化等优点,因而迅速在机器人领域得到了广泛的应用<sup>[4]</sup>。

经过若干年的发展,目前机器人视觉 SLAM 的整体框架包括前端视觉里程计,后端非线性优化,闭环检测和地图构建<sup>[5]</sup>。而 SLAM 问题的求解主要可以分为 2 种方法,其中传统的方法沿用机器人领域的滤波方法,如扩展卡尔曼滤波(EKF)<sup>[6]</sup>、扩展信息滤波(EIF)<sup>[7]</sup>、粒子滤波(PF)<sup>[8]</sup>等,但由于滤波器模型自身所具有的缺陷,因而难以应用到大规模环境地图的创建中。相比滤波方法,基于图优化的方法在减少计算量和实际环境适用性等方面均具有较大的优势,因而目前通常使用基于图优化的方法求解 SLAM 问题。其中, Klein 等提出的 PTAM(parallel tracking and mapping)算法是第一个使用图优化作为后端的方法,实现了跟踪与建图的并行化<sup>[9]</sup>,但且存在跟踪与丢失等缺陷。Henry 等<sup>[10]</sup>提出了基于深度相机的 SLAM 算法,该算法首先提取给定图像的 SIFT 特征点,并采用迭代最近点(iterative closest point, ICP)算法求解 2 幅图像间相机位姿变换,最后通过使用图优化工具 TORO 进行全局优化。Engelhard 等<sup>[11]</sup>采用 SURF 特征代替 SIFT 特征进行特征匹配,并结合像素对深度信息,获得对应空间匹配点对集合,通过采用随机采样一致性(RANSAC)算法与 ICP 算法求取相邻帧间相机初始位姿,最后使用 Hogman 位姿图优化方法求解全局最优位姿。Mur-Artal 等<sup>[12]</sup>基于 PTAM 算法架构,提出了基于 ORB 特征点的单目 SLAM 系统,继承了 PTAM 的后端非线性优化实现方案与关键帧处理机制,并在 2017 年进行了扩展<sup>[13]</sup>,是目前 SLAM 系统中最为完善、易于使用的系统之一。然而上述 SLAM 系统均是假设机器人运行环境是静态的或图像信息的变化只依赖于相机自身的运动,缺乏对场景中突然出现运动物体如行人、移动的桌椅等的考虑。

当机器人运行场景中存在运动物体时,将给 SLAM 系统引入误差,从而得到错误的位姿信息,进而影响所构建地图的准确性。而为了解决动态环境下的 SLAM 问题,必须引入运动物体检测(MOD)方法<sup>[14]</sup>区分图像特征点中的静态特征和运动特征。其中, Bibby 等<sup>[15]</sup>提出的

SLAMIDE 算法通过期望最大化(EM)算法更新场景中的特征点运动模型,并使用可逆模型选择机制将动态物体引入到 SLAM 过程。Alcantarilla 等<sup>[16]</sup>将场景流引入到视觉里程计中,通过场景流的变化情况从而检测出场景图像中运动物体的存在。文献[17]基于当前帧的场景流构造角度直方图,并通过高斯混合函数近似角度直方图,使用 EM 算法求解对应高斯混合参数,从而达到检测动态特征的目的。虽然这类方法能够有效检测动态特征,减小位姿估计误差,但缺乏关于场景中存在的运动物体对闭环检测影响的考虑。

针对场景中出現运动物体时,给 SLAM 系统带来的位姿估计错误与地图构建误差等问题,本文通过引入基于特征点运动矢量的运动点检测算法,提出了一种改进视觉 SLAM 算法。一方面,通过引入运动点检测算法,对视觉里程计进行改进,提高了位姿估计的精度。通过结合初始相机位姿,引入了特征点运动矢量的计算模型,并使用 EM 方法求解对应的高斯混合模型参数,通过结合阈值以区分图像中的静态和动态特征点,并判断是否需要优化初始相机位姿。另一方面,基于运动点检测结果,对闭环检测算法进行了改进。通过设置图像预处理,剔除了运动点占比较大和与前一帧相似性较高的图像,有效提高了闭环检测算法的计算效率;通过使用剔除动态特征后的图像特征点对场景进行描述,同时改进了单个节点处两幅图像的相似性得分计算函数,并通过时间一致性和对极几何约束对候选闭环进行校验,可以有效提高闭环检测的准确率。

## 1 基于特征点运动矢量的改进视觉里程计

视觉里程计能够通过相邻帧间的图像估计相机运动,并构建场景局部地图,而针对场景中出現运动物体时,视觉里程计位姿估计误差较大的问题,本文通过引入基于特征点运动矢量的运动点检测环节,从而对视觉里程计进行改进。改进视觉里程计流程如图 1 所示。

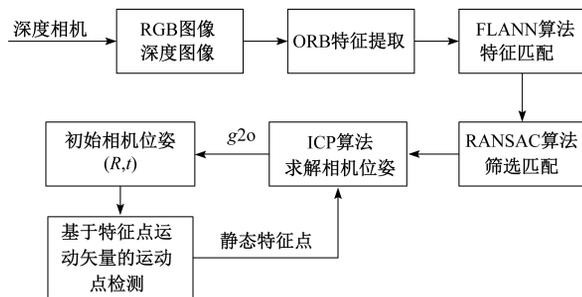


图 1 改进视觉里程计流程

Fig. 1 The process of improved visual odometry

### 1.1 特征点运动矢量计算模型

理论情况下,相邻图像中静态物体对应的匹配特征点对在相机坐标系下的坐标值经过位姿变换后,其坐标差值为0。但若相邻图像中存在运动物体时,由于物体处于运动状态,因而即使匹配特征点对的坐标值经过位姿变换后,其坐标差值也不为0,且对应差值相对较大。

基于上述考虑,本文通过构建特征点运动矢量计算模型,并将求解得到的帧间初始相机位姿引入到特征点运动矢量的计算模型中,从而描述相邻图像对应场景中物体的相对运动情况。根据初始相机位姿,计算前一帧图像中的匹配特征点在当前帧相机坐标系下的坐标值,并通过与当前帧图像对应特征点的相机坐标值进行比较,求解特征点运动矢量。

如图2所示,假设有相邻图像 $I_1, I_2$ ,其中图像 $I_1$ 到图像 $I_2$ 的相机位姿变换为 $(R, t)$ ,对应的相机中心分别为 $O_1, O_2$ 。假设存在一个空间点 $P$ ,其在图像 $I_1$ 中投影点为 $p_1$ ,在图像 $I_2$ 中投影点为 $p_2$ ,有:

$$p_1 = [u_1, v_1]^T, p_2 = [u_2, v_2]^T \quad (1)$$

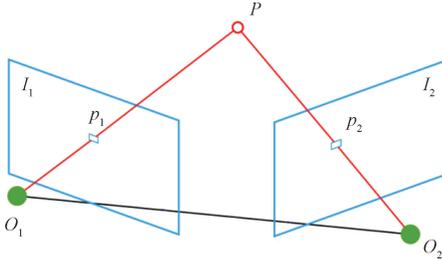


图2 特征点运动矢量模型

Fig. 2 The feature point motion vector model

通过深度图像可以获取像素点 $p_1, p_2$ 对应的深度值,假设分别为 $s_1, s_2$ ,则在图像 $I_2$ 对应的相机坐标系下坐标值分别为:

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = s_1 \cdot R \cdot \mathbf{K}^{-1} \cdot \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} + t \quad (2)$$

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} = s_2 \cdot \mathbf{K}^{-1} \cdot \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \quad (3)$$

其中, $\mathbf{K}$ 为相机内参矩阵,由相机焦距 $f_x, f_y$ 和偏移 $c_x, c_y$ 决定,有:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

则特征点对应的运动矢量可以表示为:

$$\mathbf{M} = \begin{bmatrix} \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix} - \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} \end{bmatrix} \quad (5)$$

如式(5)所示,通过求解匹配特征点对在当前图像对应相机坐标系下的坐标差值,从而表示当前帧中特征点对应的运动矢量。理论情况下,静态物体对应特征点的运动矢量为0,但由于匹配误差、投影误差等影响,存在静态特征点运动矢量不为0的现象,因而不能直接使用特征点运动矢量区分图像中的静态和动态特征点。

### 1.2 运动点确认

由文献[18]可知,当一个物体处于运动状态时,它所对应的运动矢量角度是趋于一致的,因而图像特征点运动矢量的角度服从高斯混合模型。

本文通过构建特征点运动矢量角度的高斯混合模型,并使用EM方法求解对应的高斯混合参数,其中均值最小的高斯模型为静态特征点运动矢量角度对应的分布模型。通过计算第 $k$ 帧图像特征点在运动物体对应高斯模型中的高斯值 $\varphi_p$ 及其在静态物体对应高斯模型中高斯值 $\varphi_s$ 。若满足 $\varphi_p > \varphi_s$ ,则将此特征点列入待选运动点集合 $P_T^{(k)}$ 中,并等待进一步筛选;否则列入静态点集合 $P_S^{(k)}$ 。

为了能充分利用第 $k-1$ 帧的运动点检测结果,将集合 $P_T^{(k)}$ 中的特征点与第 $k-1$ 帧的代选运动点集合 $P_T^{(k-1)}$ 中的特征点进行特征匹配,并以 $\delta_p$ 表示匹配结果,有:

$$\delta_p = \begin{cases} true, & \text{成功} \\ false, & \text{失败} \end{cases} \quad (6)$$

当匹配成功时,则将该点列入运动点集合 $P_M^{(k)}$ 中;否则列入状态未知点集合 $P_U^{(k)}$ 中。若以 $\phi_p$ 表示图像特征点检测结果,有:

$$\phi_p = \begin{cases} moving, & \varphi_p > \varphi_s \& \delta_p = true \\ unknown, & \varphi_p > \varphi_s \& \delta_p = false \\ static, & \text{其他} \end{cases} \quad (7)$$

### 1.3 帧间相机位姿优化

由于场景图像中可能存在运动物体,同时当相机移动过快,会存在运动矢量过大的特征点,均将导致求解的初始相机位姿具有较大的偏差。因而,为提高视觉里程计相机位姿估计准确度,同时保证算法计算效率,设置了相应的预处理,基于运动点检测结果,并设置阈值判断是否使用静态特征点对当前时刻相机位姿进行求解。

假设第 $k$ 帧图像在匹配筛选后特征点的个数为 $n_N$ ,其中待选运动点集合 $P_T^{(k)}$ 中特征点个数为 $n_T$ ,运动点集合 $P_M^{(k)}$ 中特征点个数为 $n_M$ 。以 $E_T$ 表示当前时刻相机位姿是否需要优化,有:

$$E_T = \begin{cases} false, & \frac{n_T}{n_N} < \varepsilon_{th1} \& \frac{n_M}{n_T} < \varepsilon_{th2} \\ true, & \text{其他} \end{cases} \quad (8)$$

式中:  $\varepsilon_{th1}$ 、 $\varepsilon_{th2}$  分别为设定的阈值。

如式(8)所示,当图像特征点中待选运动点占比较小,即存在较少的运动矢量过大的特征点,且运动点占比较小时,保持初始相机位姿不变;其余情况下,使用静态点集合  $P_s^{(k)}$  中的特征点再次求解相机位姿。

### 1.4 改进视觉里程计算法步骤

1) 通过深度相机获取场景的 RGB 和深度图像,并引入 ORB 算法提取 RGB 图像中的特征点;

2) 使用 FLANN 算法对第  $k-1$  帧图像和第  $k$  帧图像进行特征匹配,并通过 RANSAC 算法对初始匹配结果进行筛选,剔除其中的误匹配;

3) 使用 ICP 算法初步求解第  $k$  帧相机位姿,并放入 g2o 优化器进行优化,获取初始相机位姿  $(R, t)$ ;

4) 结合初始相机位姿,计算图像特征点的运动矢量,构建运动矢量角度的高斯混合模型,并引入 EM 方法求解对应的高斯混合模型参数;

5) 分别计算图像特征点在运动物体和静态物体对应高斯模型中的高斯值  $\varphi_p$ 、 $\varphi_s$ ,当满足  $\varphi_p > \varphi_s$ ,则将该点列入待选运动点集合  $P_T^{(k)}$  中;

6) 若集合  $P_T^{(k)}$  中存在特征点可以匹配前一帧  $P_T^{(k-1)}$  中的特征点时,则将该点列入运动点集合  $P_M^{(k)}$  中,否则列入状态未知点集合  $P_U^{(k)}$  中;

7) 通过设置阈值,当图像特征点中存在较多运动矢量过大的特征点,或运动点占比较大时,使用静态特征点通过 ICP 算法求解相机位姿,并进行优化;其他情况下,保持初始相机位姿不变;

其中,基于特征点运动矢量的运动点检测算法流程如图 3 所示。

## 2 改进闭环检测算法

由于系统噪声等因素影响,视觉里程计求解的帧间相机位姿,难免会存在误差,并且随着时间的推移,误差将一直累积,严重影响定位与建图的准确度。而 SLAM 系统闭环检测模块可以通过判断机器人是否再次运动到先前位置,即识别机器人运动轨迹是否为一个闭环,从而为后端非线性优化提供更多有效数据信息,以消除位姿估计中的累积误差。

针对场景中存在运动物体时,传统闭环检测算法准确率低、易失败的问题,本文基于运动点检测结果对闭环检测算法进行了改进。首先,为了提高算法的计算效率,同时避免运动点占比较大的图像对闭环检测的影响,改进算法通过设置图像预处理环节,减少了相似性比较的图像帧数;其次,通过使用分层  $K$  均值聚类算法构建视觉字典树,并使用剔除动态点后的图像特

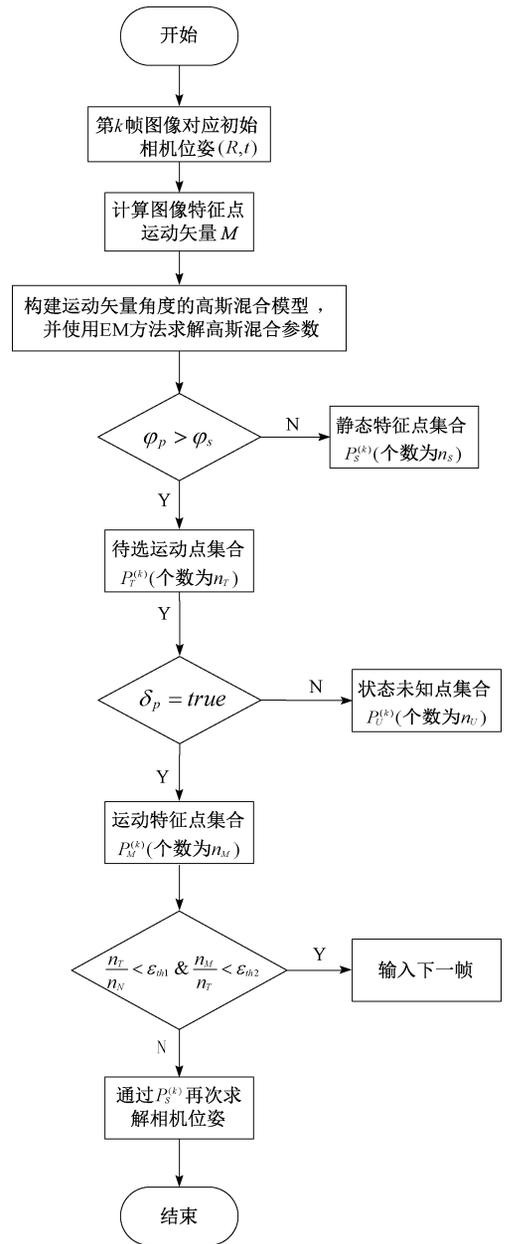


图 3 基于特征点运动矢量的运动点检测算法流程

Fig. 3 The flow chart of motion points detection algorithm based on feature point motion vector

征点向字典树投影,从而对场景图像进行描述;最后,为了提高相似性场景下闭环检测算法的鲁棒性,在文献[19]基础上改进了单个节点处两幅图像间的相似性计算函数。

### 2.1 图像预处理

由于相机获取的场景图像是连续的,因而通常相邻帧图像间相似性较高。通过舍弃与前一帧间相似度较高的图像,可以减少比较次数,提高闭环检测算法的计算效率。根据视觉字典树模型,定义相邻图像间相似度为:

$$sim(V_t, V_{t-1}) = \begin{cases} N_{pair}/N_{V_t}, N_{V_t} \geq N_{V_{t-1}} \\ N_{pair}/N_{V_{t-1}}, N_{V_t} < N_{V_{t-1}} \end{cases} \quad (9)$$

式中:  $N_{V_t}$  和  $N_{V_{t-1}}$  分别为图像  $I_t$  和图像  $I_{t-1}$  对应视觉单词数目;  $N_{pair}$  为图像  $I_t$  与图像  $I_{t-1}$  之间共同的视觉单词对数目。若  $sim(V_t, V_{t-1})$  值越大, 即表示 2 幅图像之间相似性越高。当  $sim(V_t, V_{t-1})$  值大于设定的阈值时, 则可以舍弃当前帧; 反之, 则表示当前帧代表一个新的场景, 并且可用于闭环检测。

当场景图像中存在运动物体且图像特征点中运动点占比较大时, 原有的场景特征会被运动特征点代替, 对应图像间相似性将大大降低, 即使存在闭环也难以检测闭环的存在, 而由于采集的场景图像是连续的, 因而对存在较多运动特征点的图像予以舍弃。

假设第  $k$  帧图像在匹配筛选后特征点个数  $n_N$ , 其中检测到的动态特征点个数为  $n_M$ , 则若满足式 (10) 时, 将当前帧舍弃。

$$\frac{n_M}{n_N} > \varepsilon_{th3} \quad (10)$$

式中:  $\varepsilon_{th3}$  为设定的阈值。

### 2.2 基于视觉字典树的场景描述

本文采用分层  $K$  均值聚类算法构建视觉字典树, 并且考虑场景图像中存在运动物体的情况, 通过运动点检测结果, 使用剔除运动点后的图像特征点向视觉字典树进行投影, 获取图像在每个树节点的词频-逆文件频率 (TF-IDF) 熵作为该节点的得分权重, 并以图像在视觉字典树的得分向量描述场景图像。其中,  $L$  层  $k$  分支的视觉字典树可以表示为图 4 所示。

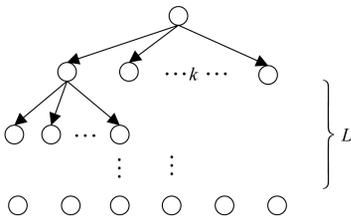


图 4 视觉字典树

Fig. 4 The visual vocabulary tree

定义图像在单个树节点处的 TF-IDF 熵作为该节点处的得分权重, 有:

$$\omega_i^l(X) = \frac{n_i}{n} \log \frac{N}{N_i} \quad (11)$$

式中:  $l \in \{0, 1, \dots, L\}$ ;  $i \in \{1, 2, \dots, k^l\}$ ;  $N$  表示待处理图像总数;  $n_i$  表示图像  $X$  投影到节点  $i$  的特征个数;  $N_i$  表示至少有 1 个特征投影到节点  $i$  的图像数。

则图像  $X$  的场景描述向量可以表示为:

$$W(X) = (W^1(X), W^2(X), \dots, W^L(X)) \quad (12)$$

式中:  $W^l(X) = (\omega_1^l(X), \omega_2^l(X), \dots, \omega_k^l(X))$  表示图像  $X$  在视觉字典树中第  $l$  层的得分向量。

### 2.3 改进的相似性得分计算方法

闭环检测算法的难点之一在于相似性场景所引起的感知歧义, 感知歧义将引入误正闭环, 从而破坏所构建地图的一致性。而针对上述问题, 本文在文献 [19] 视觉字典树 TF-IDF 得分匹配方法的基础上, 对 2 幅图像在视觉字典树中单个节点处的相似性得分函数进行改进, 有效减少了由相似性场景所引发的感知歧义。

假设图像  $X$  和图像  $Y$  在节点  $O_i^l$  处的 TF-IDF 熵分别为  $\omega_i^l(X), \omega_i^l(Y)$ , 文献 [19] 通过最小值函数计算该节点处图像间相似性得分, 即:

$$S_i^l(X, Y) = \min \{ \omega_i^l(X), \omega_i^l(Y) \} \quad (13)$$

然而使用最小值函数求取 2 幅图像在单个节点处的相似性得分, 具有一定的不合理性。对于图像  $X, Y, Z$ , 在节点  $O_i^l$  处有 TF-IDF 熵  $\omega_i^l(X), \omega_i^l(Y), \omega_i^l(Z)$ , 且满足  $\omega_i^l(Z) > \omega_i^l(Y) > \omega_i^l(X)$ 。根据式 (13) 可得,  $S_i^l(X, Z) = S_i^l(X, Y)$ , 即表示图像  $X$  与图像  $Y, Z$  在节点  $O_i^l$  处相似性相同。然而, 根据人们对相似性的理解, 在节点  $O_i^l$  处图像  $X$  与图像  $Y$  之间的相似性理应更高, 因而在式 (13) 中存在对相似场景的感知歧义。因而文献 [20] 对相似性得分函数进行了改进, 通过使用反比例函数计算单个节点处图像间的相似性得分, 有:

$$S_i^l(X, Y) = \frac{1}{|\omega_i^l(X) - \omega_i^l(Y)| + 1} \quad (14)$$

通过式 (14) 求解的图像间相似性得分与对应图像在指定节点的 TF-IDF 熵之差成反比, 更符合人们对相似性的理解习惯。

而对于式 (14), 若满足  $\omega_i^l(Z) > \omega_i^l(Y) > \omega_i^l(X)$  和  $\omega_i^l(Z) - \omega_i^l(Y) = \omega_i^l(Y) - \omega_i^l(X)$ , 则  $S_i^l(X, Y) = S_i^l(Y, Z)$ 。而根据 TF-IDF 的定义思想, TF-IDF 越大, 表明单词在图像中出现的频率高, 并且在其他图像中较少出现, 具有很好的区别分类性, 因此在节点  $O_i^l$  处图像  $Y$  与图像  $Z$  之间的相似性理应更高。针对上述问题, 本文改进了图像  $X$  与图像  $Y$  在节点  $O_i^l$  处的相似性得分:

$$S_i^l(X, Y) = \min \{ \omega_i^l(X), \omega_i^l(Y) \} + e^{-|\omega_i^l(X) - \omega_i^l(Y)|} \quad (15)$$

式 (15) 中最小值函数项是相似性判断的主要依据, 而第 2 项的负指数函数求解的相似性得分与图像  $X$  和图像  $Y$  在节点  $O_i^l$  处的 TF-IDF 熵之差成反比, 即表示当 2 幅图像在节点处的 TF-IDF 熵越接近, 图像间对应相似性也就越高, 从而克服了对于个别感知歧义不适用的问题。

基于式 (15), 图像  $X$  与图像  $Y$  在第  $l$  层的相似性得分可以定义为:

$$S^l(X, Y) = \sum_{i=1}^{k^l} S_i^l(X, Y) =$$

$$\sum_{i=1}^{k^l} \min \{ \omega_i^l(X), \omega_i^l(Y) \} + \sum_{i=1}^{k^l} e^{-|\omega_i^l(X) - \omega_i^l(Y)|} \quad (16)$$

使用自上而下计算图像间相似性增量的方法可以有效避免重复累计相似性,因而定义图像  $X$  与图像  $Y$  在第  $l$  层的相似性得分增量  $\Delta S^l(X, Y)$  为:

$$\Delta S^l(X, Y) = \begin{cases} S^l(X, Y) - S^{l+1}(X, Y), 1 \leq l < d \\ S^d(X, Y), l = d \end{cases} \quad (17)$$

定义金字塔匹配核为:

$$K(X, Y) = \sum_{l=1}^L \eta_l \Delta S^l(X, Y) \quad (18)$$

式中:  $\eta_l$  为视觉字典树在第  $l$  层的匹配强度系数,  $l$  越小则对特征空间的划分越粗,不同图像特征投影到同一视觉单词的约束越小,因而更加容易获得较高的匹配得分。根据文献[19]选取  $\eta_l = 1/k^{l-1}$  抑制不同层数的匹配差异,则式(18)可改写为:

$$K(X, Y) = K(W(X), W(Y)) = S^L(X, Y) + \sum_{l=1}^{L-1} \frac{1}{k^{l-1}} (S^l(X, Y) - S^{l+1}(X, Y)) \quad (19)$$

## 2.4 闭环确认

若图像间相似性得分大于给定全局相似性阈值,则提取当前图像与其匹配图像为候选闭环。然而候选闭环中往往存在误正闭环,不仅不能修正位姿估计的累积误差,还会破坏已构建地图的一致性,因此剔除误正闭环对于闭环检测环节尤为重要。本文主要从时间连续性和对极几何约束两个方面剔除误正闭环。

时间连续性是 SLAM 采集图像的一个重要特征,可作为约束条件剔除误正闭环。由于相机采集到的场景图像具有时间连续性,即相邻帧图像对应着场景的连续变化,因而在当前帧发生闭环时,闭环也应该连续出现一定帧数。若闭环的发生不满足时间连续性,则可认为该闭环为误正闭环,并予以剔除。

视觉 SLAM 中匹配的闭环图像是对同一场景的不同视角成像,因而满足对极几何约束。通过使用 RANSAC 算法对闭环图像的特征匹配点进行估计,获取 2 幅图像间的基础矩阵,当满足该基础矩阵的内点数比例低于阈值时可认为此次闭环为误正闭环,并从候选闭环中剔除。

## 3 实验与分析

本文运行程序电脑配置为 CPU 为 i5 处理器,4 GB 内存,主频为 2.5 GHz,系统为 Ubuntu16.04,实验所用数据集均来自 TUM 标准数据集。

### 3.1 运动点检测效果评价

为了评价基于运动矢量的运动点检测算法的检测效

果,本文从 TUM 数据集中选取带有运动对象的数据集 freiburg3\_walking\_xyz 进行运动点检测实验,实验结果如图 5 所示,其中图 5(a) 为输入图像;图 5(b) 为 RANSAC 算法筛选匹配后的图像,其中圈出部分表示匹配筛选后的图像特征点;图 5(c) 为运动点检测算法对特征点进行筛选的图像,其中加深部分表示检测到的运动点。

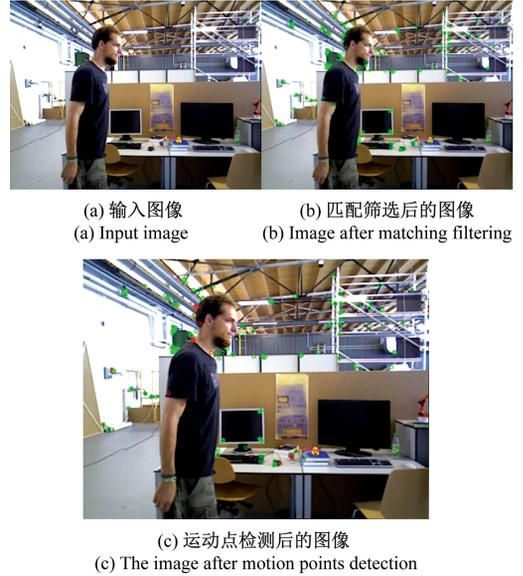


图 5 基于运动矢量的运动点检测结果

Fig. 5 The motion points result based on motion vector

由图 5 可知,当场景图像存在运动物体时,本文中的运动点检测算法可以有效检测出图像中对应的运动特征点,且具有较高的准确度。

### 3.2 算法位姿估计精度评估

为了检验改进视觉 SLAM 算法的位姿估计精度,本文从 TUM 数据集中选取了一组标准数据集进行实验,数据集中均提供了相机移动的真实轨迹。图 6 所示为数据集下本文算法的实验结果,其中黑线表示真实轨迹,蓝线表示估计轨迹,红线表示估计估计与真实轨迹之间的误差。

由于数据集中包含了相机移动的真实轨迹,因而可以通过求解对应的均方根误差 (root mean squared error, RMSE) 表示相机移动的真实轨迹与 SLAM 算法估计轨迹之间的误差,从而评判算法位姿估计的精确度。本文算法与文献[13]中 ORB-SLAM2 算法在不同数据集下的均方根误差如表 1 所示。

如表 1 所示,相比 ORB-SLAM2 算法,本文所提改进算法估计轨迹与真实轨迹之间均方根误差值相对较小,并且针对存在运动物体的实验数据集,本文所提算法位姿估计精度相对较高。上述实验表明了在不同的数据集下,本文所提算法具有较好的鲁棒性,并且视觉里程计求

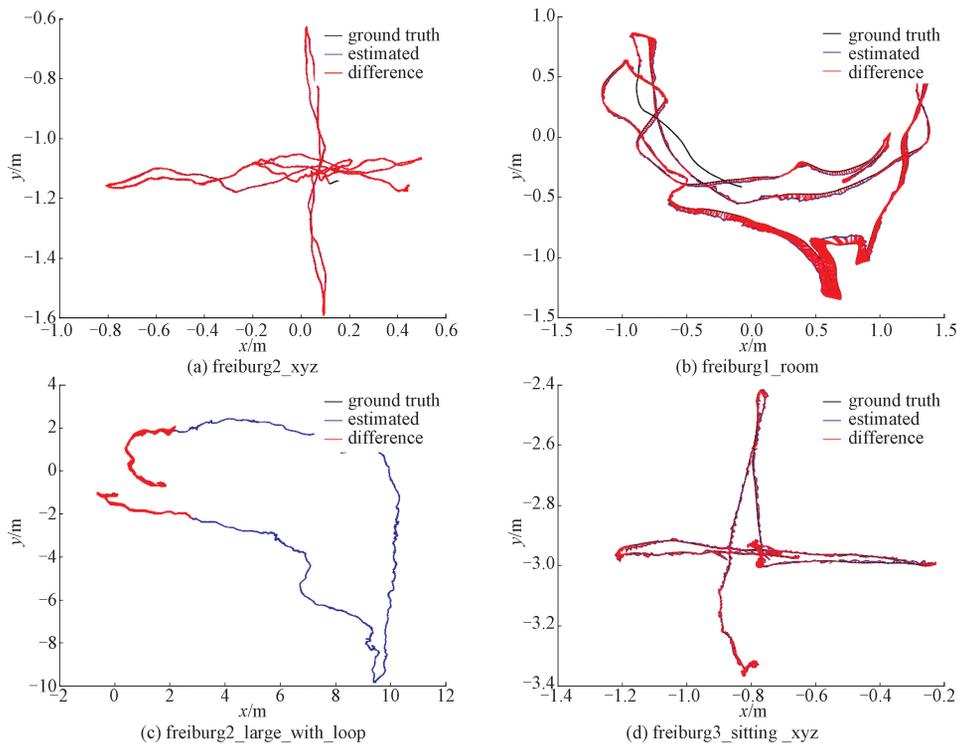


图6 绝对轨迹误差

Fig. 6 The chart of absolute trajectory error

解的相机位姿具有较高的精度。

表1 本文算法与 ORB-SLAM2 算法 RMSE 对比表

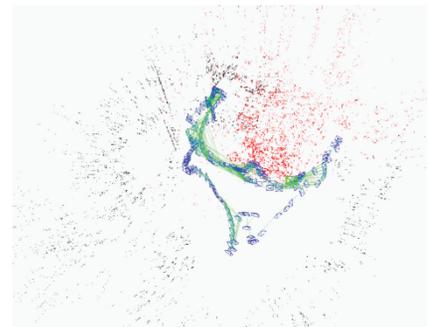
Table 1 Comparison of RMSE between the algorithm in this paper and ORB-SLAM2

数据集	ORB-SLAM2	本文算法
freiburg1_xyz	0.011 063	<b>0.009 405</b>
freiburg1_room`	<b>0.059 376</b>	0.060 164
freiburg2_xzy	0.004 269	<b>0.003 598</b>
freiburg2_large_with_loop	0.132 603	0.115 513
freiburg3_sitting_xyz	<b>0.009 921</b>	0.010 112
freiburg3_sitting_halfsphere	0.020 282	<b>0.018 633</b>
freiburg3_walking_xyz	0.684 541	<b>0.332 561</b>

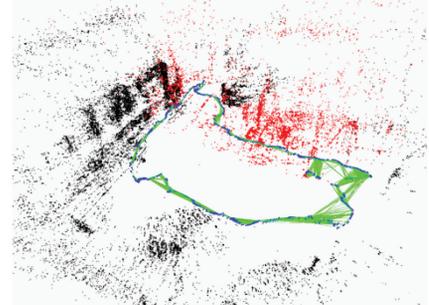
### 3.3 闭环检测与建图效果评估

为了验证改进视觉 SLAM 算法的闭环检测效果,选取了 2 种不同规模的带有闭环的 TUM 数据集进行实验。如图 7 所示,图 7(a)为本文算法在数据集 freiburg1\_room 下生成的闭环轨迹;图 7(b)为算法在数据集 freiburg2\_large\_with\_loop 下生成的闭环轨迹。

由图 7 可知,对于不同规模的数据集,本文算法均可以有效检测出闭环的存在,并且具有一定的鲁棒性。为了进一步验证本文所提算法的性能,通过选取数据集 freiburg1\_desk2 和 freiburg1\_360 进行三维重建实验,其生成的轨迹与点云如图 8 与 9 所示,图 8 表示在 freiburg1\_desk2 数据集下生成的结果,图 9 表示在 freiburg1\_360 数



(a) freiburg1\_room 下闭环轨迹  
(a) Loop closure trajectory in freiburg1\_room



(b) freiburg2\_large\_with\_loop 下闭环轨迹  
(b) Loop closure trajectory in freiburg2\_large\_with\_loop

图7 不同数据集下的闭环轨迹

Fig. 7 Loop closure trajectory under different datasets  
数据集下生成的结果,其中图 8(a)、9(a)和图 8(b)、9(b)

分别为生成的轨迹和对应的点云图像。

由图 8 和 9 可知,通过在 freiburg1\_desk2 和 freiburg1\_360 数据集下进行实验,得到了对应的轨迹图像和点云图像。其中,从生成的点云图像可以看出本文算法点云拼接效果良好,与真实环境基本相符,从而再一次验证了改进算法位姿估计具有较高的精确度,且具有较好的建图效果。

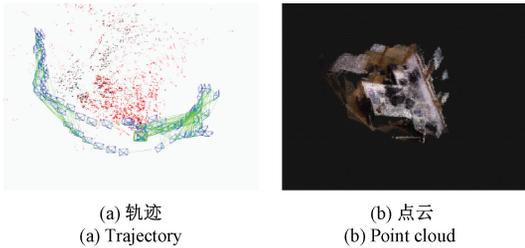


图 8 freiburg1\_desk2 下对应结果图像

Fig. 8 Corresponding result under freiburg1\_desk2

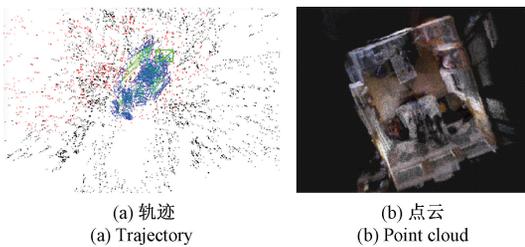


图 9 freiburg1\_360 下对应结果

Fig. 9 Corresponding result under freiburg1\_360

## 4 结 论

针对场景图像中存在运动物体时,给 SLAM 系统带来的位姿估计错误与地图构建误差等问题,本文提出了一种基于特征点运动矢量的改进视觉 SLAM 算法。一方面,通过为视觉里程计增加运动点检测算法,优化了求解得到的初始相机位姿,提高了在室内动态环境中视觉里程计位姿估计的精度;另一方面,基于运动点检测结果对闭环检测输入图像进行预处理,有效提高了闭环检测的计算效率,同时利用剔除动态点后的图像特征点对场景图像进行描述,并改进了单个节点处图像间相似性得分计算函数,经闭环确认后检测出闭环的存在。本文通过多种数据集进行实验,验证了本文所提算法的有效性,且具有较高的位姿估计精度和较好的建图效果。后续计划融合惯性测量单元(IMU)数据进行算法改进,使算法在低纹理环境下也能具有较高的精度和较好的鲁棒性。

## 参考文献

- [1] 姚二亮,张合新,张国良,等. 基于 Vision-IMU 的机器人同时定位与地图创建算法[J]. 仪器仪表学报, 2018, 39(4): 230-238.  
YAO ER L, ZHANG H X, ZHANG G L, et al. Robot simultaneous localization and mapping algorithm based on vision and IMU [J]. Chinese Journal of Scientific Instrument, 2018, 39(4): 230-238.
- [2] 张震,郑宏,周璇,等. 一种结合 ORB 特征和视觉词典的 RGB-D SLAM 算法[J]. 计算机工程与应用, 2018, 54(12): 170-176.  
ZHANG ZH, ZHENG H, ZHOU X, et al. RGB-D SLAM algorithm based on ORB features and visual dictionary [J]. Computer Engineering and Applications, 2018, 54(12): 170-176.
- [3] 朱朔凌,毛建旭,王耀南,等. 基于惯性导航角度补偿的室内激光 SLAM 方法[J]. 电子测量与仪器学报, 2019, 33(3): 1-7.  
ZHU SH L, MAO J X, WANG Y N, et al. Indoor laser SLAM method based on inertial navigation angle compensation [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(3): 1-7.
- [4] CADENA C, CARLONE L, CARRILLO H, et al. Past, present, and future of simultaneous localization and mapping: towards the robust-perception age [J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
- [5] GEORAGES Y, DANIEL A, ELIE S, et al. Keyframe-based monocular SLAM: design, survey, and future directions [J]. Robotics and Autonomous Systems, 2017, 98: 67-88.
- [6] 唐宏伟,孙炜,杨凯,等. 基于 SURF 特征的多机器人栅格地图拼接方法[J]. 电子测量与仪器学报, 2017, 31(6): 859-868.  
TANG H W, SUN W, YANG K, et al. Grid map merging approach of multi-robot based on SURF feature [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(6): 859-868.
- [7] 朱代先,王晓华. 基于精确稀疏扩展信息滤波的粒子滤波 SLAM 算法研究[J]. 计算机工程与科学, 2012, 34(7): 140-145.  
ZHU D X, WANG X H. Research on the particle filter SLAM algorithm based on exactly sparse extended information filter [J]. Computer Engineering & Science, 2012, 34(7): 140-145.
- [8] 付星星,王满意,卜雄洙,等. 基于多信道衰减尺度的无源被动粒子滤波定位算法研究[J]. 国外电子测量技术, 2018, 37(2): 58-62.  
FU X X, WANG M Y, BU X ZH, et al. Research on

- particle filtering for device-free localization based on multi-channel and fade level [J]. *Foreign Electronic Measurement Technology*, 2018, 37(2): 58-92.
- [9] MAURO C, TAIHU P, FEDERICO B. DS-PTAM: Distributed stereo parallel tracking and mapping SLAM system [J]. *Journal of Intelligent & Robotic Systems*, 2019, 95(2):365-377.
- [10] HENRY P, KRAININ M, HERBST E, et al. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments [C]. *12th International Symposium on Experimental Robotics*, 2010:22-25.
- [11] ENGLEHARD N, ENDRES F, HESS J, et al. Real-time 3D visual SLAM with a hand-held RGB-D camera [C]. *Proceedings of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, 2011:1-15.
- [12] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: A versatile and accurate monocular SLAM system [J]. *Robotics IEEE Transactions on Robotics*, 2015, 31(5): 1147-1163.
- [13] MUR-ARTAL R, TARDOS 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.
- [14] 莫邵文, 邓新蒲, 王帅, 等. 基于改进视觉背景提取的运动目标检测算法 [J]. *光学学报*, 2016, 36(6): 196-205.
- MO SH W, DENG X P, WANG SH, et al. Moving object detection algorithm based on improved visual background extractor [J]. *Acta Optica Sinica*, 2016, 36(6): 196-205.
- [15] BIBBY C, REID I D. Simultaneous localization and mapping in dynamic environments (SLAMIDE) with reversible data association [C]. *Proceedings of Robotics Science and Systems*, 2007: 105-112.
- [16] ALCANTARILLA P F, YEBES J J, ALMAZAN J, et al. On combining visual SLAM and dense scene flow to increase the robustness of localization and mapping in dynamic environments [C]. *IEEE International Conference on Robotics & Automation*, 2012: 1290-1297.
- [17] SHIMAMURAL J, MORIMOTO M, KOIKE H. Robust vSLAM for dynamic scenes [C]. *MVA 2011 IAPR Conference on Machine Vision Applications*, 2011: 344-347.
- [18] 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):195-203.
- LIN ZH L, ZHANG G L, YAO ER L, et al. Stereo visual odometry based on motion object detection in dynamic scene [J]. *Acta Optica Sinica*, 2017, 37(11): 195-203.
- [19] 李博, 杨丹, 邓林. 移动机器人闭环检测的视觉字典树金字塔 TF-IDF 得分匹配方法 [J]. *自动化学报*, 2011, 37(6):665-673.
- LI B, YANG D, DENG L. Visual vocabulary tree with pyramid TF-IDF scoring match scheme for loop closure detection [J]. *Acta Automatica Sinica*, 2011, 37(6): 665-673.
- [20] 李永锋, 张国良, 王峰, 等. 一种基于历史模型集的改进闭环检测算法 [J]. *机器人*, 2015, 37(6): 663-673.
- LI Y F, ZHANG G L, WANG F, et al. An improved loop closure detection algorithm based on historical model set [J]. *Robot*, 2015, 37(6): 663-673.

### 作者简介



孙龙龙, 2017 年于安徽工程大学获得学士学位, 现为安徽工程大学硕士研究生, 主要研究方向为移动机器人 vSLAM。

E-mail: slonlong@163.com

**Sun Longlong** received his B. Sc. degree from Anhui Polytechnic University in 2017.

Now he is a M. Sc. candidate at Anhui Polytechnic University. His main research interests include mobile robot vSLAM.



江明(通信作者), 1993 年于上海工业大学(现上海大学)获得硕士学位, 现为安徽工程大学教授、硕士生导师, 主要研究方向为机器人智能控制系统和先进检测技术。

E-mail: kjjm@ahpu.edu.cn

**Jiang Ming** (Corresponding author) received his M. Sc. degree from Shanghai Technology University (now Shanghai University) in 1993. He is a professor and M. Sc. supervisor at Anhui Polytechnic University. His main research interests include robotic intelligent control system and advanced detection technology.