

doi: 10.7690/bgzd.2019.09.019

基于特征地图的改进回环检测算法

徐彬彬, 刘鹏远, 张峻宁

(陆军工程大学石家庄校区, 石家庄 050003)

摘要: 为提高移动机器人回环检测模块的准确率和实时性, 提出一种基于特征地图的改进回环检测算法。在传统模型的基础上, 对关键帧的特征点进行筛选, 选取高效特征点构建特征地图, 利用视觉字典树对特征地图和关键帧进行场景描述; 对词袋模型进行改进, 将场景分割用在图像信息的提取和特征聚类上; 建立基于分层 K++ 均值的视觉字典树, 得到改进的基于分层金字塔 TF-IDF (term frequency-inverse document frequency) 的匹配方法。实验结果证明: 相比 FAB-MAP (fast appearance-based mapping) 和 RGB-D SLAM v2, 改进算法在特征点规模、实时性、召回率方面性能更优。

关键词: 回环检测; 特征地图; 场景分割; 视觉字典树; TF-IDF

中图分类号: TP391.41 **文献标志码:** A

Improved Loop Detection Algorithm Based on Feature Map

Xu Binbin, Liu Pengyuan, Zhang Junning

(Shijiazhuang Campus of PLA University of Army Engineering, Shijiazhuang 050003, China)

Abstract: In order to improve the accuracy and real-time of the loop detection module of mobile robot, an improved loop detection algorithm based on feature map is proposed. On the basis of traditional model, selecting the feature points of key frames and efficient feature points to build feature maps, using visual dictionary tree to description scene of feature map and key frame. Secondly, to improve the bag of words model, scene segmentation is applied to image information extraction and feature clustering. Finally, a visual dictionary tree based on hierarchical K++ means is established, and an improved method of score matching based on hierarchical pyramid TF-IDF (term frequency-inverse document frequency) is obtained. The test results show that compared with FAB-MAP (fast appearance-based mapping) and RGB-D SLAM v2, the improved algorithm has better performance in feature point size, real-time performance and recall rate.

Keywords: loop detection; feature map; scene segmentation; visual dictionary tree; TF-IDF

0 引言

同时定位与地图构建 (simultaneous localization and mapping, SLAM) 是实现移动机器人自主导航的关键技术^[1]。然而, 在位姿估计的过程中, 由于算法本身的局限性, 不可避免地会产生累积误差。回环检测通过外观信息来检测是否访问先前的位置, 从而消除累积误差, 是实现视觉 SLAM 系统全局一致性的重要组成部分, 在同步定位与构建, 以及基于外观的机器人重定位等方面得到广泛应用。

在视觉 SLAM 中, 人们将计算机视觉领域的词袋模型 (bag-of-words BoW)^[2] 应用在回环检测模块中, 并在此基础上, 提出一系列的改进回环检测算法。2008 年, Adrien Angeli 等^[3] 提出基于视觉词袋模型的快速回环检测方法, 将形状和颜色信息增加词袋模型的条件, 高效率地解决感知混淆问题; 2012 年, Dorian 和 Juan 等^[4] 提出一种基于二进制特征的树加速词袋模型算法, 较传统词袋模型快一个数量

级; 2013 年, Mathieu Labbe^[5] 提出一种在线长时间处理大规模回环检测的方法, 通过限制回环检测的频率, 使计算时间保持在实时约束下; 2015 年, Sheraz Khan 等^[6] 提出一种基于关键帧图像的特征跟踪并结合姿态不变性的在线回环检测。

现有的回环检测算法仍然面临大规模场景变换引起的算法不稳定问题, 从而导致错误回环和消耗大量计算资源选取候选关键帧。针对以上问题, 笔者在基于帧到模型的基础上^[7], 对特征点进行筛选, 选取高效特征点构建特征地图, 并对特征区域进行正常视野和扩展视野分块, 维护特征地图; 在词袋模型方面进行改进, 将场景分割应用到特征信息提取, 将图像特征分块处理, 含重叠区域进行特征提取并采用聚类效果更好的 K++ 均值算法生成视觉单词, 最后采取李博等^[8] 提出基于分层的得分匹配算法计算相似性, 在 TUM 数据集下与 RTAB-MAP 算法^[2] 进行对特征规模、实时性、准确率-召回率对比。

收稿日期: 2019-04-25; 修回日期: 2019-06-30

作者简介: 徐彬彬 (1993—), 男, 山东人, 硕士, 从事视觉 SLAM 研究。E-mail: 983736124@qq.com。

1 基于改进帧到模型的特征地图构建

理想条件下，视觉 SLAM 系统智能机器人在移动过程中获得场景图像并从中选取关键帧，由于连续关键帧之间移动机器人的视角变化较小；因此，连续关键帧表示的场景相似性较高^[9]。笔者利用 Kinect 传感器提出一种基于帧到模型配准模型的视觉里程计方法，可以将连续关键帧的特征点融合到一个模型集中，从而提高了帧间配准的效率。在此基础上，增加特征点的筛选和扩张特征视野，并将关键帧特征点保留形成特征地图。

1.1 构建特征地图

假设 Kinect 相机在第 m 时刻采集的数据集为 I_m ，特征点 j 时刻的像素位置为 $[u_j, v_j]$ ，深度信息为 d ，假设关键帧特征点在相机模型下集合为 $L_m = l_j (j=1, \dots, n)$ ，其中特征点信息包括位置 P_j^L 和角点相应强度 R_j^L ， $R_j^L = [x, y, z]$ 依次为第 j 个特征点在相机坐标系下的位置向量，见式(1)。 R_j^L 是通过 opencv 的 SIFT^[10]库检测特征点的角点响应强度，即特征点是否高效的判断条件。以角点响应强度作为特征点的排序条件，根据实验经验，保留 300 左右的特征足以满足帧间匹配要求。

$$P_j^L = [x, y, z] = \begin{cases} x = \frac{u - c_x}{f_x} z \\ y = \frac{v - c_y}{f_y} z \\ z = d \end{cases} \quad (1)$$

1.2 特征地图的维护

在不考虑严重遮挡和光照变化的情况下，关键帧的特征点是冗余，如何剔除冗余特征点是帧间匹配的关键问题。为了维护特征地图的规模，使其正常运行，首先对关键帧图像进行预处理，以提出不在视野内的特征点。假设特征点 p_j 的像素坐标为 $[u_j, v_j]$ ， T_j 为第 j 帧相对于世界坐标系的变换矩阵，当前帧的相对帧间运动为 $T_{j,j+1}$ ，前一帧的相对帧间运动为 $T_{j-1,j}$ ，假设帧间运动相等，可通过式(2)位姿变换矩阵，删除不在视野内的特征点：

$$\left. \begin{aligned} T_{j,j+1} &= T_j^{-1} T_{j+1} \\ T_{j+1} &= T_{j-1} T_j \end{aligned} \right\} \quad (2)$$

通过式(3)计算特征点的像素坐标 $[u_j, v_j]$ ：

$$\begin{bmatrix} u_j \\ v_j \\ 1 \end{bmatrix} = \frac{KT_{j+1}P_j}{d} \quad (3)$$

该算法的伪代码如下：

```
if(u > 0 && v > 0) then
    if(u < r && v < c) 保存特征点
    else (u < r+a || v < c+b) 备份特征点
end if
end if
```

上述算法中的 r 、 c 为正常视野的长和宽。为了防止出现特征点匹配不足的问题，将视野扩大至 $r+a$ 和 $c+b$ ，在特征点不足的情况下，增加特征点，以提高帧间匹配率。通过算法将图像分为特征图像的正常视野和扩张视野 2 部分，并将扩张视野的特征点(外部 8 块)和正常视野内的特征点(内部 4 块)存入局部地图，特征图像区域划分如图 1。

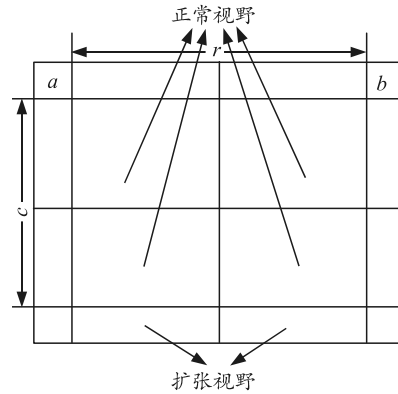


图 1 视野区域划分

2 基于图像分割的词袋模型

目前，基于计算机视觉的 BoW 分类方法是视觉 SLAM 系统回环检测的主流方法，包括以精确度高的 SIFT^[10]或 SURF^[11]特征构建词袋模型，以实时性高的二进制 ORB 或 BRISK^[12]特征构建词袋模型。二者分别在于高精度、高计算复杂度和高实时性、高效内存分配方面进行取舍。文中的词袋模型在精确度和实时性 2 方面进行折中处理，选择二进制 ORB 特征克服实时性差的问题，并采用 K-mean++ 算法进行聚类，构建基于图像分割的改进词袋模型，以提高回环检测的精度。

2.1 基于图像分割的特征提取征地图的维护

机器人视觉场景的描述依靠摄像机采集建立的视觉模型，可以直接影响视觉 SLAM 系统的性能。传统的基于平面匹配的词袋模型忽略了关键点空间信息，从而使特征信息高度聚类制约视觉单词的

表征能力，并占据大量计算空间，容易产生误回环和实时性下降。笔者在传统词袋模型的基础上增加空间信息，提取图像的 ORB 特征并采用 FAST (features from accelerated segment test) 算法^[13]进行关键点检测，场景分割如图 2 所示，具体步骤如下：

1) 首先对关键帧图像进行分块处理，设图像的长和宽为 x, y ，每一区域为 $(a+c) \times (b+c)$ ，共计 16 块等面积区域，重复区域为 $((b+c) \times c)$ 和 $((a+c) \times c)$ ，各变量之间的关系如下式，其中 $c = a/4$ 。

$$\left. \begin{aligned} 4a+c &= y \\ 4b+c &= x \\ a-c &= e \\ b-c &= d \end{aligned} \right\} \quad (4)$$

2) 每一图像区域与其他区域有重叠区域，如图 2 所示， m, n, r 为各自区域的独立区域，而 a, b, c, d, e 为重叠区域，分为 3 种情况如下式：

$$\left. \begin{aligned} S_1 &= m+a+b+c \\ S_2 &= n+2a+2c+d \\ S_3 &= r+4c+2d+2e \end{aligned} \right\} \quad (5)$$

3) 按照分组标记分别对每一块区域提取 ORB 特征，并采用 FAST 算法检测关键点，每一块区域用 $R_i (i=1,2,3,\dots,16)$ ， $R_i = \sum_{j=1}^{128} V_j^i$ ，其中 V_j^i 为 128 向量。

4) 通过 K-mean++ 聚类算法，生成视觉字典。

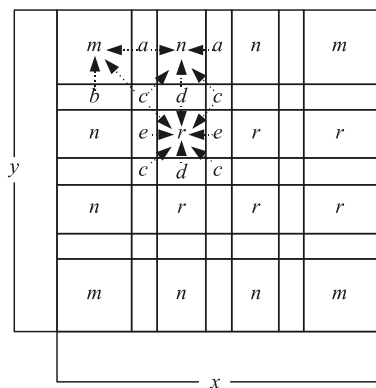


图 2 图像特征分割原则

2.2 基于分层 K-mean++ 聚类视觉字典树

为了提高特征信息聚类的表征性能和计算效率，在李博等^[8]基于分层 K-均值聚类的基础上，采用 K-mean++ 进行聚类，生成 L 层 K 分支的视觉词典如图 3 所示，其视觉单词的数量级远远高于传统

视觉字典。K-mean++ 聚类图聚类具体步骤如下：

1) 从特征区域数据中随机选取样本，进行聚类初始化；

2) 计算每个样本与初始样本 c_1 的欧式距离 $D(x)$ ，计算每个样本成为下一个聚类中心的概率 $\frac{D(x)}{\sum_{x=1}^{x=N} D(x)^2}$ ，通过轮盘法选取其余聚类中心；

3) 计算所有样本 c_i 到 K 个聚类中心的距离并分配到距离最小的聚类中心；

4) 对于选取的聚类中心重新计算质心 $c_i = \sum_{x=1}^{x=i} x \frac{1}{|c_i|}$ ；

5) 循环 3) 和 4)，一直到聚类中心满足阈值要求。

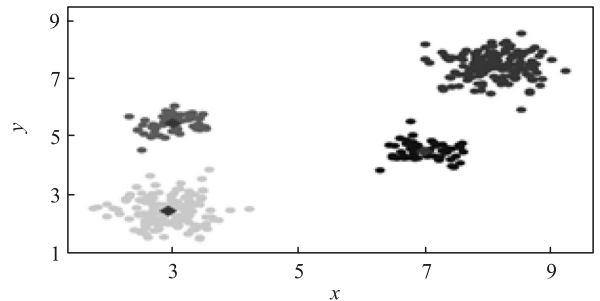


图 3 K-mean++ 聚类

2.3 基于分层 TF-IDF 熵的金字塔得分匹配方法

在对图像进行场景分割的基础上，提取图像 X 的 n 个 128 维特征， $X = \{x_1, x_2, \dots, x_n, x_j \in R^{128}\}$ ，按照金字塔结构映射到视觉字典树，构建图像-单词的逆向索引，计算图像在每个树节点的 TF-IDF 熵作为图像在该视觉单词的得分权重，用 $\omega_j^l(X)$ 表示图像 X 在视觉字典树的第 l 层， $l \in \{0, 1, \dots, L\}$ 的第 i 个节点 O_j^l 出的映射得分，定义 TF-IDF 熵为

$$\omega_j^l(X) = \frac{n_j}{n} \log \frac{N}{N_j} \quad (6)$$

其中： N 是图像的数量； N_j 是特征映射到节点 j 的数量。

图像 X 在整个视觉字典树中的得分向量为：

$$W(X) = (W^1(X), W^2(X), \dots, W^L(X)) \quad (7)$$

其中， $W^L(X) = (\omega^1(X), \omega^2(X), \dots, \omega^L(X))$ 代表第 l 层的得分向量，由于第 1 层为根节点，没有表征能力，因此，从第 1 层计算。

传统的视觉字典树是基于平面特征进行匹配，没有考虑不同层视觉单词表征能力不同的问题^[14]，

按照统一权重进行得分匹配；因此，生成的视觉字典树鲁棒性不强。笔者考虑到树分层的结构问题，对每一层分开计算相似度，设计权重函数，最后整合形成得分匹配计算公式。

首先定义图像 X 和图像 Y 在同一节点 O_j^l 的相似度函数为

$$S_j^l(X, Y) = \min\{\omega_j^l(X), \omega_j^l(Y)\}。 \quad (8)$$

图像在第 1 层的相似度函数为

$$K(X, Y) = \sum_{l=1}^L \beta \Delta S^l(X, Y)。 \quad (9)$$

将式(6)代入得：

$$S_j^l(X, Y) = \sum_{j=1}^{k^l} \min\{\omega_j^l(X), \omega_j^l(Y)\}。 \quad (10)$$

视觉字典树按照金子塔结构由上而下不断地细化，生成的视觉单词各自区分性越好，越能大大减少感知混淆现象^[15]；因此，相似度计算得分减少。为了减少计算资源的浪费，自下而上进行相似性计算，定义第 1 层的相似度得分增量

$$\Delta S^l(X, Y) = \begin{cases} S^l(X, Y), & l=L \\ S^l(X, Y) - S^{l+1}(X, Y), & 1 \leq l \leq L-1 \end{cases}。 \quad (11)$$

定义金子塔匹配核为

$$K(X, Y) = \sum_{l=1}^L \beta \Delta S^l(X, Y)。 \quad (12)$$

其中， β 为视觉字典树第 1 层的权重系数，层数 1 与匹配强度成反比，通过每层的权重系数更好地计算对应的相似度得分，抑制不同层次的差异。

2.4 回环后验处理

当 2 张关键帧图像的相似度计算大于给定阈值时，即 $S(X, Y) > T_y$ ，选取当前关键帧为候选回环。

但是，由于聚类算法的局限性将不同特征映射到同一聚类中心或真实场景特征高度相似，导致产生误回环。

针对这一现象，需要时间连续性进行后验处理。机器人在运动过程中不断采集关键帧图像，关键帧之间相互连续。如图 3 所示，当识别到先前走过的位置产生回环时，回环检测的位置应该依次连续出现，即回环位置所处的时间在前一对之后，也就是说回环对为 3-10、4-11、5-12 3 对符合时间连续性。而 4-13 回环对中，位置 4 在位置 5 的时间之前，且位置 13 在位置 12 之前；因此，本对回环为错误回环，予以剔除。

3 实验与分析

文中实验均在 Ubuntu14.04 的环境下完成。计算机运行内存 8 GB、CPU2.6 GHz、Intel i7-6700，实验数据集来自于 TUM RGB-D Benchmark。实验主要分为 2 部分：1) 与 RGB-D SLAM 在前端特征点规模和匹配估计时间进行对比；2) 改进算法与 FAB-MAP 算法在回环检测所用时间、准确率-召回率 2 方面进行对比。

3.1 特征点规模和匹配估计时间对比

文中算法的前端是基于改进帧到模型匹配模型，构建基于特征地图匹配的前端，并与基于帧到模型的 RGB-D SLAM 算法在前端的特征点规模进行对比。实验结果如表 1，表明对前端高效特征点的筛选构建特征地图，有效降低了特征点的规模，有利于特征地图的维护。

表 1 前端特征点规模对比

| 算法名称 | 数据集 desk | 数据集 desk2 | 数据集 Long-office-household |
|---------------|-------------------|-------------------|---------------------------|
| 文中改进算法 | 4.3×10^4 | 4.6×10^4 | 1.4×10^5 |
| RGB-D SLAM V2 | 4.7×10^4 | 6.3×10^4 | 2.3×10^5 |

前端匹配估计时间对比如表 2 所示。从表可以看出，筛选高特征点一定程度上减少了特征点冗余。场景规模越大，对前端时间的优化效果越显著。

表 2 前端匹配估计时间对比 ms

| 算法名称 | 数据集 desk | 数据集 desk2 | 数据集 Long-office-household |
|---------------|----------|-----------|---------------------------|
| 文中改进算法 | 44.72 | 47.73 | 194.800 |
| RGB-D SLAM V2 | 46.40 | 56.40 | 194.804 |

3.2 改进回环检测模型对比实验

本部分实验将改进的回环检测模型在 TUM 数据 fr3/nostructure-withloop，对回环检测模型的实时性进行验证，与 RTAB-MAP 的回环检测模型进行对比。实验结果如图 4 和 5。

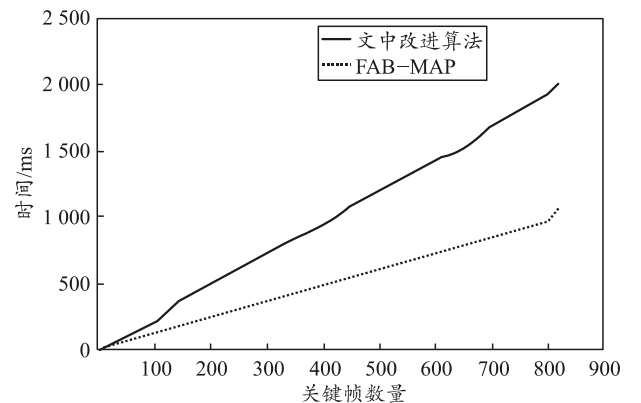


图 4 回环检测时间对比

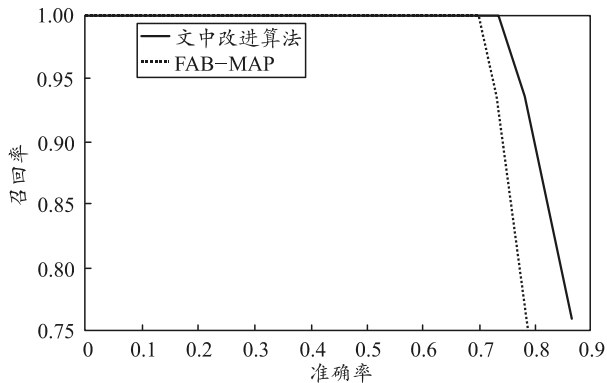


图 5 准确率-召回率对比曲线

从图 4 可以看出：回环检测需要的时间随着历史帧数的增加而变长，但是从数据集的对比实验可以看出，实时性由于传统的回环检测模型是基于所有历史信息计算，回环时间较文中算法长。从图 5 可以看出：回环性能在小规模场景数据集回环检测中略优，但是在大规模数据集的回环检测中，回环性能进一步体现优越性。

4 结论

回环检测模块是 SLAM 框架的重要组成部分之一，是消除累积误差和保持地图一致性的关键技术。笔者从前端的帧间匹配模型和视觉字典树 2 方面进行改进，提出基于特征地图的回环检测算法。首先对前端特征点进行筛选，构建特征地图并稳定维护其规模；其次将场景分割应用到图像信息的提取，对图像进行分块处理，对不同区域部分重叠提取特征来减少感知混淆现象；最后采用分层的金字塔 TF-IDF 得分匹配方法，提高相似度计算的有效性。通过 TUM 数据集对比实验可知，改进算法在特征点规模、实时性、准确率方面均有提高。

参考文献：

- [1] SMITH R, SELF M, CHEESEMAN P. Estimating Uncertain Spatial Relationships in Robotics[J]. Machine Intelligence & Pattern Recognition, 2013, 5(5): 435-461.
- [2] CUMMINS M, NEWMAN P. Accelerating FAB-MAP

- With Concentration Inequalities[J]. Robotics IEEE Transactions on, 2010, 26(6): 1042-1050.
- [3] ADRIEN A, DAVID F, STEPHANE D, et al. Fast and incremental method for loop-closure detection using bags of visual words[J]. IEEE Transactions on Robotics, 2008, 24(5): 1027-1037.
- [4] MATHIEU L, FRANCOIS M. Appearance-based loop closure detection for online large-scale and long-term operation[J]. IEEE Transactions on Robotics, 2013, 29(3): 734-745.
- [5] DORIAN G, JUAN D T. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.
- [6] KHAN S, WOLLHERR D. IBuILD: Incremental bag of Binary words for appearance based loop closure detection[C]//IEEE International Conference on Robotics and Automation. IEEE, 2015: 5441-5447.
- [7] 王旒军, 陈家斌, 余欢, 等. RGB-D SLAM 综述[J]. 导航定位与授时, 2017, 4(6): 11-18.
- [8] 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): 666-673.
- [9] 李戈星, 刘士荣, 仲朝亮, 等. 基于改进关键帧选择的 RGB-D SLAM 算法[J]. 大连理工大学学报, 2017, 57(4): 412-417.
- [10] DRYANOVSKI I, VALENTI R G, XIAO J. Fast visual odometry and mapping from RGB-D data[C]//IEEE International Conference on Robotics and Automation. IEEE, 2013: 2305-2310.
- [11] LOWE D G. Distinctive image features form scale-invariant keypoints international Journal of Computer Vision[J]. Robotic, 2014, 4(2): 91-110.
- [12] BOAL J, SANCHEZ-MIRALLES A, et al. Topological simultaneous localization and mapping: a survey[J]. Robotical, 2014, 32(5): 803-821.
- [13] 刘国忠, 胡钊政. 基于 SURF 和 ORB 全局特征的快速闭环检测[J]. 机器人, 2017, 39(1): 37-45.
- [14] 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): 664-672.
- [15] 梁志伟, 陈燕燕, 朱松豪, 等. 基于视觉词典的单目视觉闭环检测算法[J]. 模式识别与人工智能, 2013, 26(6): 562-570.