

基于贪心策略的视觉 SLAM 闭环检测算法

王云峰, 翁秀玲, 吴 炜, 郭东辉

(厦门大学信息科学与技术学院, 厦门 361005)

摘 要: 动态环境与视觉混淆严重影响视觉闭环检测性能. 基于贪心策略, 提出了一种在线构建视觉词典的闭环检测算法. 算法优先处理 Surf 描述与已有单词 Surf 描述欧式距离最大的特征点, 改进特征点与单词 Surf 描述最近邻的约束条件, 生成了表征性能强、量化误差小的视觉词典, 算法具备实时性, 并在动态环境图像集与视觉混淆多发生的图像集上, 在确保 100% 准确率的条件下, 最大召回率分别提升了 5% 与 4%.

关键词: 即时定位与地图构建; 闭环检测; 视觉词典

中图分类号: TP391.4

文献标志码: A

文章编号: 0493-2137(2017)12-1262-09

Loop Closure Detection Algorithm Based on Greedy Strategy for Visual SLAM

Wang Yunfeng, Weng Xiuling, Wu Wei, Guo Donghui

(School of Information Science and Technology, Xiamen University, Xiamen 361005, China)

Abstract: The performance of loop closure detection is seriously affected by dynamic objects and perceptual aliasing in the environment. Based on greedy strategy, a real-time loop closure detection approach using online visual dictionary is proposed. The process of dictionary construction gives priority to dealing with Surf feature that has the maximum Euclidean distance from the closest vocabulary word. A more discriminative and representative visual vocabulary is produced through adding constraint condition to the nearest neighbor distance. This visual vocabulary guarantees a small quantization error. The proposed approach meets real-time constraints. Experiments based on datasets from dynamic environments and visually repetitive environments demonstrated that the largest recall rate increased by 5% and 4% respectively at 100% precision.

Keywords: simultaneous localization and mapping; loop closure detection; visual vocabulary

在未知环境中, 移动机器人可根据自身的传感器数据创建环境地图并进行精确定位, 这一行为被称为即时定位与地图构建(simultaneous localization and mapping, SLAM)^[1]. SLAM 是当前实现移动机器人自主导航的关键技术. 而闭环检测则是 SLAM 系统中的关键环节, 它是通过识别机器人当前位置是否位于已经访问过的区域, 作为是否进行后端优化的依据. 基于闭环的后端优化对消除机器人位姿估计的累积误差、实现地图的一致性至关重要. 近年来, 由于视觉传感器具有信息量大、性价比高、通用性好等特点, 基于视觉传感器的闭环检测引起了研究者的广

泛关注. 然而, 视觉传感器图像数据规模大, 所处环境往往为动态, 并且存在视觉混淆现象, 这些因素导致闭环检测成为 SLAM 系统研究中的难题.

闭环检测算法通常采用视觉词袋模型(bag of visual word, BoVW)^[2]对图像进行表示. 在此基础上, Cummins 等^[3-7]通过离线方式获得视觉词典, 训练数据时考虑视觉单词的相关性, 用 Chow-liu 树算法近似视觉单词的联合概率分布, 随后通过递归贝叶斯滤波获得闭环的后验概率. 后验概率法已经成为闭环检测的标准方法. 此外, 文献[8-13]通过在线方式构建视觉词典. Angeli 等^[8-9]使用基于局部颜色和

收稿日期: 2017-01-15; 修回日期: 2017-06-07.

作者简介: 王云峰(1977—), 男, 博士, 副教授.

通讯作者: 王云峰, yunfengwang@xmu.edu.cn.

基金项目: 国家自然科学基金资助项目(61274133).

Supported by the National Natural Science Foundation of China(No. 61274133).

SIFT 特征的在线视觉词典. Labbé 等^[10-11]提取 SURF 特征在线构建词典, 引入内存管理机制, 仅仅使用出现频率高的位置参与闭环检测, 实时性得以满足. 由于二进制特征描述具有运算和比较速度快优点, Garcia-Fidalgo 等^[12]采用关键点 FAST, 结合二进制描述子 BRIEF 在线构建词典. Khan 等^[13]基于连续图像匹配的二进制特征描述 BRISK 在线构建词典.

SLAM 系统对闭环检测的实时性和准确率要求较高, 但是在现有方法中, FAB-MAP^[6]离线视觉词典对特定环境没有很好的表征性能, 准确率较低. 文献[9, 12-13]中, 随着地图规模的增大, 当前帧需要进行比较的历史帧越来越多, 未达到实时性要求. RTAB-MAP^[11]虽然有较好的实时性和准确率, 但使用最近邻比例法更新视觉词典, 对特征点处理顺序敏感, 映射到视觉单词的特征量化误差大.

为了进一步提高闭环检测的准确率, 本文提出了一种在线构建视觉词典的闭环检测算法. 在线构建视觉词典过程中, 采用贪心策略, 优先处理图像中离上一时刻视觉单词最近邻距离最大的特征点, 通过改进最近邻欧氏距离比率法更新视觉词典, 生成了表征性能更强的视觉词典. 实验结果表明, 该方法有效地提高了闭环检测的准确率.

1 基于视觉词典的闭环检测

在视觉 SLAM 闭环检测中, 首先需采用词袋模型对图像进行表示; 然后进行图像预处理剔除连续相似度高的图像; 最后运用贝叶斯滤波不断更新当前图像与历史图像的后验概率来获取闭环.

1.1 基于视觉词典的图像表示

为了快速实现字典的构造和闭环检测, 目前通常利用图像特征点进行视觉词典的构造. 假设 $t-1$ 时刻视觉词典为

$$V_{t-1} = \{w_1, w_2, \dots, w_a, \dots, w_{n-1}, w_n\} \quad (1)$$

式中: a 为单词的标号; w_a 为第 a 个单词, 是由 j 时刻的图像的第 b 个特征点转换而来. 设单词的 64 维 Surf 描述向量为 $\mathbf{Des}(P_{jb})$, 则单词可表示为

$$w_a = (a, \mathbf{Des}(P_{jb})) \quad (2)$$

在 t 时刻获取图像 I_t , 利用 Surf 算法^[14]对图像 I_t 提取 m 个特征点 $P_t = \{P_{t1}, P_{t2}, \dots, P_{tm}\}$, 计算其 64 维描述向量 $\mathbf{Des}_t = \{\mathbf{Des}(P_{t1}), \mathbf{Des}(P_{t2}), \dots, \mathbf{Des}(P_{tm})\}$. 从已有视觉词典 V_{t-1} 中寻找任一特征点 P_{tk} ($1 \leq k \leq m$) 的欧氏距

离最近邻和次近邻, 如果最近邻距离和次近邻距离之比大于阈值, 将此特征点的描述向量看作新单词 $w_{n+1} = (n+1, \mathbf{Des}(P_{tk}))$, 形成新的视觉词典 V_t 如式 (3) 所示.

$$V_t = \{w_1, w_2, \dots, w_n, w_{n+1}\} \quad (3)$$

P_{tk} 由新单词表示; 否则, P_{tk} 由与其距离最近的视觉单词表示, 视觉词典不变. 即每个特征点有唯一的视觉单词与之对应.

所谓词袋模型就是利用视觉词典把图像转化为一个低维的向量表示. 由于图像 I_t 的任一特征点均有唯一的视觉单词与之相对应, 所以基于视觉词典 V_t , 图像 I_t 可以被表示成式 (4) 所示的词袋模型.

$$Z_t = \{z_{t1}, z_{t2}, \dots, z_{tm}\} \quad (4)$$

式中 $z_{tk} = a$.

1.2 图像预处理

移动机器人在采集场景图像过程中, 通常相邻图像具有相似性, 如果每一帧图像都进行闭环检测, 系统需要处理的数据规模大, 可通过式 (5) 基于词袋模型的单词计算相邻图像间的相似度 $\text{sim}(Z_t, Z_{t-1})$, 当相似度大于一个阈值 T_{sim} 时, 删除图像 I_t . N_{pair} 为图像 I_t 和图像 I_{t-1} 之间共同的视觉单词对数目, N_{Z_t} 和 $N_{Z_{t-1}}$ 分别为图像 I_t 和 I_{t-1} 的词袋模型单词数目.

$$\text{sim}(Z_t, Z_{t-1}) = \begin{cases} N_{\text{pair}} / N_{Z_t} & N_{Z_t} \geq N_{Z_{t-1}} \\ N_{\text{pair}} / N_{Z_{t-1}} & N_{Z_t} < N_{Z_{t-1}} \end{cases} \quad (5)$$

1.3 贝叶斯滤波

贝叶斯滤波被广泛运用在闭环检测当中, 文献[3-12]均使用贝叶斯滤波进行闭环检测, 本文也采用贝叶斯滤波获取闭环. 随机变量 S_t 表示 t 时刻闭环存在, $S_t = i, i \in [0, t-1]$ 表示随机事件“图像 I_t 与过去的图像 I_i 发生闭环”, $S_t = -1$ 表示随机事件“ t 时刻没有发生闭环”. $I' = I_0, I_1, \dots, I_t$ 表示到 t 时刻为止所有的图像序列. 在贝叶斯概率框架中, t 时刻闭环假设的后验概率分布函数如式 (6) 所示.

$$p(S_t | I') = \eta p(I_t | S_t) p(S_t | I'^{-1}) \quad (6)$$

式中: $p(I_t | S_t)$ 为观测模型; $p(S_t | I'^{-1})$ 为状态预测, 根据 $t-1$ 时刻为止的图像信息预测 t 时刻闭环状态 S_t ; η 为归一化因子, 如式 (7) 所示.

$$\eta = \frac{1}{\sum_{i=-1}^{t-1} p(I_t | S_t = i) p(S_t = i | I'^{-1})} \quad (7)$$

观测模型 $p(I_t | S_t)$ 使用似然函数进行评估^[11], 若图像 I_t 发生闭环, 首先通过式 (5) 得到图像 I_t 与图像 $I_i, i \in [0, t-1]$ 的相似度 sim_i , 统计相似度之间的标准差 σ 和均值 μ , 然后使用式 (8) 计算 $p(I_t | S_t)$.

$$p(I_t | S_t = i) = \begin{cases} \frac{\text{sim}_i - \sigma}{\mu} & \text{sim}_i \geq \sigma + \mu \\ 1 & \text{其他} \end{cases} \quad (8)$$

若图像 I_t 没有发生闭环, 则 $p(I_t | S_t = -1)$ 为

$$p(I_t | S_t = -1) = \frac{\mu}{\sigma} + 1 \quad (9)$$

状态预测 $p(S_t | I^{t-1})$ 根据全概率公式展开, 如式 (10) 所示.

$$p(S_t | I^{t-1}) = \sum_{j=-1}^{t-1} p(S_t | S_{t-1} = j) p(S_{t-1} = j | I^{t-1}) \quad (10)$$

式中: $p(S_{t-1} = j | I^{t-1}), j \in [0, t-2]$ 为上一时刻图像闭环检测的后验概率, 当 $j = t-1$ 时, $p(S_{t-1} = j | I^{t-1})$ 赋值为 0; $p(S_t | S_{t-1} = j)$ 为时间演变模型.

时间演变模型 $p(S_t | S_{t-1} = j)$ 根据机器人在 $t-1$ 和 t 时刻的运动模型, 预测 S_t 的分布. 根据经验, 有以下 4 种取值方式^[11].

$$\begin{cases} p(S_t = -1 | S_{t-1} = -1) = 0.9 \\ p(S_t = -1 | S_{t-1} = j) = 0.1 & j \in [0, t-1] \\ p(S_t = i | S_{t-1} = -1) = 0.1/M & i \in [0, t-1] \\ p(S_t = i | S_{t-1} = j) = \frac{0.9f(i)}{\sum_{i=j}^{j+16} f(i)} & i, j \in [0, t-1] \end{cases} \quad (11)$$

式中: M 为 $t-1$ 时刻为止经预处理后剩余图像总数; $f(i)$ 为高斯分布, 如式 (12) 所示, 其中 $\sigma = 1.6, \mu = j$.

$$f(i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(i-\mu)^2}{2\sigma^2}} \quad (12)$$

通过贝叶斯滤波得到闭环检测的后验概率 $p(S_t | I^t)$. 如果图像 I_t 没有发生闭环的概率 $p(S_t = -1 | I^t)$ 低于一个阈值 T_{loop} , 则认为后验概率最高的 $p(S_t = i | I^t)$ 发生闭环, 即图像 I_t 和图像 I_i 发生闭环.

2 基于贪心策略在线构建视觉词典

视觉词典直接影响闭环检测的准确性, 其建立方

式可分为离线构建和在线构建. 离线构建以聚类算法^[15]作为主要方式, 但需要提前训练数据集, 而且不适合表示训练数据集中没有的场景图像. 因此本文将基于贪心策略在线构建视觉词典.

2.1 最近邻欧氏比率法

在线视觉词典构建^[10-12]应用最多的就是最近邻欧氏距离比率法. 该方法首先利用式 (13) 计算图像 I_t 的特征点 P_{tk} 与视觉词典中所有单词的 64 维 Surf 算法描述的欧式距离, 选出最近邻 (nearest distance, ND) 和次近邻 (next nearest distance, NND), 如算法 1 所示. 可通过随机 kd 树算法^[16]对词典 V_{t-1} 中的视觉单词创建索引树, 加速最近邻搜索过程. 如果 ND 与 NND 二者满足式 (14), 则该特征点作为新单词进入词典, 并由新单词代表; 否则视觉词典不变, 该特征点由最近邻表示. 如算法 2 所示.

$$\text{Dist}(P_{tk}, w_a) = \left[\sum_{i=1}^{64} (\text{Des}_i(P_{tk}) - \text{Des}_i(P_{jb}))^2 \right]^{1/2} \quad (13)$$

$$\text{ND}/\text{NND} > T_{\text{NND}} \quad (14)$$

算法 1 findneighbor(P_{tk}, V)

输入: 特征点 P_{tk}

视觉单词集合 V

输出: 最近邻距离 ND

次近邻距离 NND

最近邻单词编号 nd

开始

(1) ND = NND = 100.0;

(2) for w_a in V

(3) $D = \text{Dist}(P_{tk}, w_a)$;

(4) if ($D < \text{ND}$)

(5) NND = ND;

(6) ND = D ;

(7) nd = a ;

(8) else $D < \text{NND}$

(9) NND = D ;

(10) end if

(11) end for

(12) return (ND, NND, nd);

算法 2 nearestupdate(V_{t-1}, P_t)

输入: $t-1$ 时刻视觉词典 V_{t-1}

t 时刻特征点集 P_t

输出: t 时刻视觉词典 V_t

t 时刻图像表示 Z_t

开始

- (1) $v = 0$;
- (2) $V_t = V_{t-1}$;
- (3) for P_{tk} in P_t
- (4) badDist = false;
- (5) $(ND, NND, nd) = \text{findneighbor}(P_{tk}, V_t)$;
- (6) if $(ND/NND > T_{\text{NNDR}})$
- (7) badDist = true;
- (8) if badDist
- (9) $v = v + 1$;
- (10) $w_{n+v} = (n + v, \text{Des}(P_{tk}))$;
- (11) newWords. push_back(w_{n+v});
- (12) V_t . push_back(w_{n+v});
- (13) $z_{tk} = n + v$;
- (14) Z_t . push_back(z_{tk});
- (15) else
- (16) $z_{tk} = nd$;
- (17) end if
- (18) end for
- (19) return (V_t, Z_t);

最近邻欧氏距离比率法对特征点处理顺序敏感。在线构建词典过程中,最近邻欧氏距离比率法随机选择当前图像 I_t 要处理的特征点,先处理的特征点的最近邻和次近邻距离比值若大于阈值,则作为新单词,后面处理的特征点 P_{tk} ($1 \leq k \leq m$) 从当前图像新生成的单词和视觉词典 V_{t-1} 中寻找最近邻和次近邻,该特征点的最近邻和次近邻受到当前图像 I_t 新生成的单词影响并且决定了该特征点是否进入词典。而处理 P_{tk} 时当前图像 I_t 新生成的单词受特征点 $\{P_{t1}, P_{t2}, \dots, P_{t(k-1)}\}$ 处理顺序的影响,因此特征点的处理顺序对该特征点是否进入词典造成了影响。Surf 算法对图像 I_t 提取特征点 $P_t = \{P_{t1}, P_{t2}, \dots, P_{tm}\}$, 如图 1(a) 所示,其中圆点代表提取出来的特征点。通过最近邻欧氏比率法更新视觉单词,如果特征点处理顺序为 $\{P_{t1}, P_{t2}, \dots, P_{tm}\}$, 结果如图 1(b) 所示,其中“+”号表示用来建立新单词的特征点位置。如果特征点处理顺序为 $\{P_{tm}, P_{t(m-1)}, \dots, P_{t1}\}$, 结果如图 1(c) 所示。两种处理顺序产生的视觉词典有很大的差别。

最近邻欧氏距离比率法在线构建词典存在特征量化误差大,会产生表征性弱的单词两个缺陷。该方法根据最近邻距离 ND、次近邻距离 NND 比率 NNDR 判断是否更新单词,存在 4 种情况,如表 1 所示。当发生第 2 种情况时,特征点由与其距离最近的单词代表,ND 比较大,产生的量化误差大,降低了

图像描述的准确性;当发生第 3 种情况时,特征点作为新单词进入词典,但 ND 太小时,产生的新单词可能就是最近邻单词的匹配点,词典表征能力弱。



(a) 特征点



(b) 顺序处理特征点结果



(c) 逆序处理特征点结果

图 1 最近邻比率法对特征点处理顺序的敏感性分析

Fig.1 Sensitivity analysis of feature point processing order using nearest-neighbor distance ratio method

表 1 最近邻比率法更新词典情况

Tab.1 Updating vocabulary condition of nearest-neighbor distance ratio method

ND	NND	NNDR	结果
大	微大于 ND	$T_{\text{NNDR}} < \text{NNDR}$	新单词
大	远大于 ND	$\text{NNDR} \leq T_{\text{NNDR}}$	最近单词表示
小	微大于 ND	$T_{\text{NNDR}} < \text{NNDR}$	新单词
小	远大于 ND	$\text{NNDR} \leq T_{\text{NNDR}}$	最近单词表示

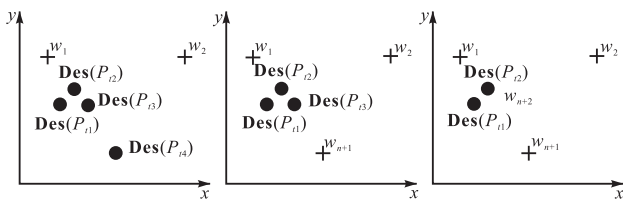
2.2 贪心策略构建视觉词典

为了保证闭环检测的实时性与准确性,不仅要求在线词典规模小而且单词表征能力强。本质就是当

特征点与已有所有单词差别都很大时,即 ND 很大时,该特征点应作为新单词进入词典;当 ND 很小时,该特征点不作为新单词进入词典,而是由 Surf 描述最近欧式距离的单词表示;当 ND 适中时,应根据 ND、NND 的比值判断. 已公开发表的最近邻欧氏距离比率法就是因为没有单独考虑 ND 很大与很小的情况才存在特征量化误差大、表征能力弱的单词的缺点. 此外最近邻欧氏距离比率法在线构造词典对特征点处理顺序敏感,已公开发表的文献鲜见此方面的研究. 由于为了准确性,ND 很大的特征点必须进入词典,因此基于贪心策略,优先处理 ND 最大的特征点可以先建立视觉词典轮廓,减小特征量化误差,生成表征性能强的词典.

假设特征的描述向量空间为二维,图 2(a)中“+”号为视觉词典 V_{t-1} 中单词 w_1 、 w_2 在向量空间中的位置,圆点为图像 I_t 的特征点描述向量 $\text{Des}(P_{i1})$ 、 $\text{Des}(P_{i2})$ 、 $\text{Des}(P_{i3})$ 、 $\text{Des}(P_{i4})$ 在向量空间中的位置. 根据最近邻比率法更新单词,最后处理 $\text{Des}(P_{i4})$ 的结果如图 2(b)所示, $\text{Des}(P_{i1})$ 、 $\text{Des}(P_{i2})$ 、 $\text{Des}(P_{i3})$ 均使用单词 w_1 表示, $\text{Des}(P_{i4})$ 看作新单词 w_{n+1} . 基于贪心策略从最近邻距离最远的 $\text{Des}(P_{i4})$ 开始处理,则将 $\text{Des}(P_{i4})$ 、 $\text{Des}(P_{i3})$ 看作新单词 w_{n+1} 、 w_{n+2} ,随后 $\text{Des}(P_{i1})$ 、 $\text{Des}(P_{i2})$ 通过单词 w_{n+2} 表示,结果如图 2(c)所示. 图 2(c)的平均量化误差明显小于图 2(b),如式(15)所示.

$$\frac{\sum_{k=1}^2 \text{Dist}(P_{ik}, w_{n+2})}{2} < \frac{\sum_{k=1}^3 \text{Dist}(P_{ik}, w_1)}{3} \quad (15)$$



(a) 已有单词和特征点描述向量 (b) 最近邻比率法更新新词典 (c) 贪心策略更新词典

图 2 基于贪心策略构建词典对比

Fig.2 Comparison of building vocabulary based on greedy strategy

为了更进一步改善最近邻欧氏距离比率法,使其构建的在线词典量化误差更小,表征能力更强,本文单独考虑了 ND 很大与很小的情况,改进后的最近邻约束,更新单词情况如表 2 所示. $\text{ND} > T_{\text{ND}2}$ 时,该特征点作为新单词进入词典,避免了原最近邻欧氏距离比率法可能存在的量化误差大的缺点. 当 ND 小于

阈值 $T_{\text{ND}1}$ 时,该特征点直接由最近邻单词表示,剔除了最近邻欧氏距离比率法可能存在表征性能弱的单词,减小了视觉词典规模,提高了建立索引树效率.

基于贪心策略与改进后的最近邻约束的在线字典构造过程如算法 3 所示. 对图像 I_t 的特征点集合 P_t 中的 m 个特征点,根据式(13)计算与视觉词典 V_{t-1} 的所有单词的欧氏距离,得到 m 个最近邻距离,选择最近邻距离最大的特征点 P_{ic} 优先进行处理. 从视觉词典 V_t 中得到特征 P_{ic} 的最近邻、次近邻距离,根据表 2 更新词典,并将图像 I_t 用词袋模型表示.

表 2 改进最近邻比率法更新词典情况

Tab.2 Updating vocabulary condition of improved nearest-neighbor distance ratio method

ND	NND	NNDR	结果
$\text{ND} > T_{\text{ND}2}$			新单词
$T_{\text{ND}1} \leq \text{ND} \leq T_{\text{ND}2}$	微大于 ND	$T_{\text{NNDR}} < \text{NNDR}$	新单词
$T_{\text{ND}1} \leq \text{ND} \leq T_{\text{ND}2}$	远大于 ND	$\text{NNDR} \leq T_{\text{NNDR}}$	最近单词表示
$\text{ND} < T_{\text{ND}1}$			最近单词表示

算法 3 greedyupdate(V_{t-1}, P_t)

输入: $t-1$ 时刻视觉词典 V_{t-1}

t 时刻特征点集 P_t

输出: t 时刻视觉词典 V_t

t 时刻图像表示 Z_t

开始

- (1) for P_{ik} in P_t
- (2) $(\text{ND}, \text{NND}, \text{nd}) = \text{findneighbor}(P_{ik}, V_{t-1})$;
- (3) end for
- (4) $v = 0$;
- (5) $V_t = V_{t-1}$;
- (6) $P'_t = \text{Sort } P_t \text{ according to ND}$
- (7) for P_{ic} in P'_t
- (8) $(\text{ND}, \text{NND}, \text{nd}) = \text{findneighbor}(P_{ic}, \text{newWords})$;
- (9) if $(\text{ND} > T_{\text{ND}2})$
- (10) $\text{badDist} = \text{true}$;
- (11) else if $(\text{ND} < T_{\text{ND}1})$
- (12) $\text{badDist} = \text{false}$;
- (13) else if $(\text{ND}/\text{NND} > T_{\text{NNDR}})$
- (14) $\text{badDist} = \text{true}$;
- (15) else
- (16) $\text{badDist} = \text{false}$;
- (17) end if
- (18) if (badDist)


```

(19)  $v = v + 1$ ;
(20)  $w_{n+v} = (n + v, \text{Des}(P_{ic}))$ ;
(21) newWords.push_back( $w_{n+v}$ );
(22)  $V_t$ .push_back( $w_{n+v}$ );
(23)  $z_{ic} = n + v$ ;
(24)  $Z_t$ .push_back( $z_{ic}$ );
(25) else
(26)  $z_{ic} = \text{nd}$ ;
(27) end if
(28) end for
(29) return( $V_t, Z_t$ );

```

3 闭环检测

基于贪心策略与改进后的最近邻约束, 本文提出的闭环检测算法如图 3 所示, 首先提取特征点, 通过贪心策略在线更新视觉单词, 基于词袋模型对图像进行表示, 使用贝叶斯滤波融合当前观测和历史闭环信息估计闭环检测后验概率, 获取闭环. 本文采用了 RTAB-MAP^[11] 相同的内存管理策略, 将内存分为短期内存、工作内存和长期内存, 只有工作内存的图像参与闭环检测, 有较好的实时性.

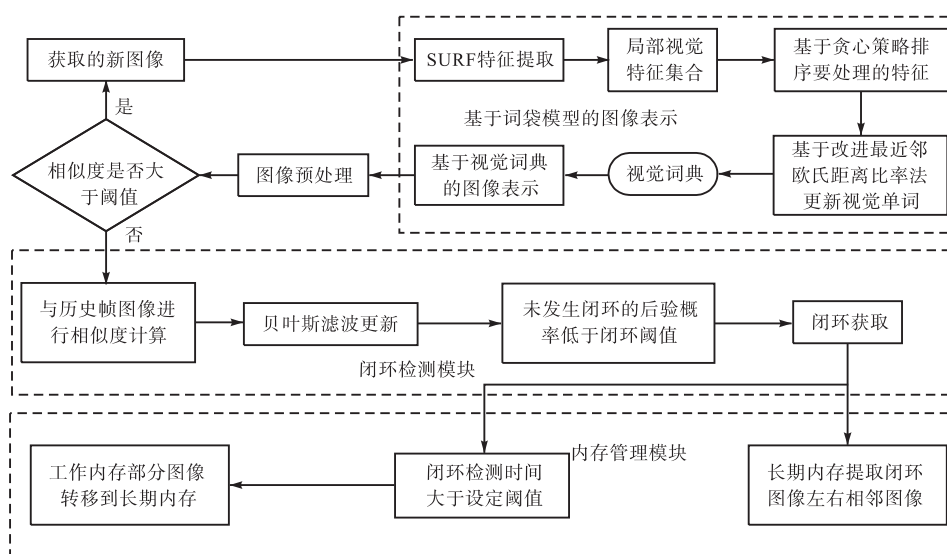


图 3 视觉闭环检测算法的框架

Fig.3 Framework of visual loop-closure detection algorithm

64 维欧氏距离的运算是闭环检测最耗时的运算. 设从图像 I_t 检测到 m 个特征点 $P_t = \{P_{t1}, P_{t2}, \dots, P_{tm}\}$, 视觉词典 $V_{t-1} = \{w_1, w_2, \dots, w_n\}$, 使用最近邻欧氏距离比率法更新视觉单词时, 时间复杂度最优的情况是每个特征点都使用与其距离最近的单词表示, 则共需 mn 次欧式距离运算; 时间复杂度最坏的情况是每个特征点均进入词典, 则共需 $mn + m(m-1)/2$ 次欧式距离运算. 基于贪心策略更新视觉单词时, 在处理前, 先运行 mn 次欧式距离运算, 进行特征点处理排序, 处理特征点时无需再进行相同的欧式距离计算, 因此时间复杂度最优的情况也是每个特征点都使用与其距离最近的单词表示, 共需 mn 次欧式距离运算; 时间复杂度最坏的情况是每个特征点均进入词典, 共需 $mn + m(m-1)/2$ 次欧式距离运算. 二者相比, 唯一的差别在于贪心策略构造词典增加了 m 个最近邻距离排序, 相对于欧氏距离计算可以忽略不计, 因此基于

贪心策略的时间复杂度与最近邻比率法相当.

在标准图像集, 即 FAB-MAP^[6] 中的 New College (NC)、City Centre (CiC) 和 IAB-MAP^[9] 中的 Lip6 Outdoor (L6O)、Lip6 Indoor (L6I) 上进行了实验. 数据集信息如表 3 所示. 实验中各个阈值的选取影响闭环检测的性能, 相关参数设置如表 4 所示, 阈值 T_{sim} 、 T_{loop} 、 T_{NNDR} 与 RTAB-MAP^[11] 取相同值, 阈值

表 3 数据集的信息

Tab.3 Details about datasets

数据集	描述	频率/Hz	图像大小	图像/幅
NC	室外, 轻微动态, 易视觉混淆	0.5	1 280 像素 × 480 像素	1 073
CiC	室外, 城市, 动态	0.5	1 280 像素 × 480 像素	1 237
L6O	室外, 轻微动态	0.5	240 像素 × 192 像素	531
L6I	室内, 静态	1.0	240 像素 × 192 像素	388

表 4 实验参数设置

Tab.4 Parameters used in experiments

T_{sim}	T_{loop}	T_{NDDR}	T_{ND1}	T_{ND2}
0.2	0.11	0.8	0.001	0.08

T_{ND1} 、 T_{ND2} 由统计大量 Surf 算法匹配结果得到. 为了满足实时性, 闭环检测时间阈值 T_{TIME} 根据图像采集频率确定, 频率为 0.5 Hz, 设为 1.4 s; 频率为 1.0 Hz, 设为 0.7 s.

基于贪心策略在线构造的视觉词典与最近邻比率法在线构建的视觉词典相比, 词典规模接近, 如图 4 所示. 原因在于虽然阈值 T_{ND2} 的判断使词典规模变大, 但同时阈值 T_{ND1} 判断剔除了表征性能低的单词.

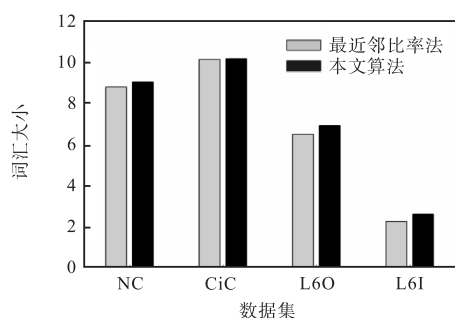


图 4 视觉词典大小对比

Fig.4 Comparison of visual vocabulary size

图 5 为本文提出的闭环检测算法的准确率-召回率曲线. 准确率^[17]指的是检测出来的闭环里含正确闭环的比例. 召回率是指检测出来的正确闭环占真实存在的闭环比例. 闭环检测的性能通常由正确率为 100% 的情况下的最大召回率来衡量. 如图 5 所示, NC 与 CiC 两个图像集的最大召回率最小, 这是因为视觉混淆与动态环境是影响闭环检测的主要因素. 基于贪心策略的闭环检测与同样具有实时性的 RTAB-MAP^[11]相比, 在视觉混淆多发的 NC 图像集上最大召回率提升了 4%, 在动态环境的 CiC 图像集上最大召回率提升了 5%.

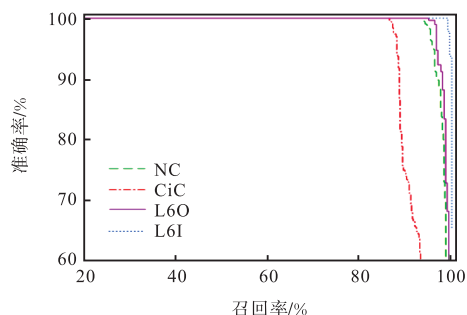


图 5 准确率-召回率曲线

Fig.5 Precision-recall curves

以动态图像集 CiC 进行实验为例, 图像序列词

袋模型的平均量化误差如图 6 所示, RTAB-MAP^[11]生成的视觉词典产生的平均量化误差大于基于贪心策略得到的图像词袋模型, 表明基于贪心策略构建的视觉词典提升了图像描述的准确性, 因而检测到了更多的闭环. 当前帧图像和历史帧图像发生闭环的后验概率最大值如图 7 所示, 图 7(a)为基于贪心策略闭环检测结果, 图 7(b)为 RTAB-MAP^[11]检测结果. 本文算法从第 690 幅开始有超过 90% 的后验, 可以检测出闭环了, 而 RTAB-MAP^[11]则需要从第 730 幅左右才开始检测出闭环. 图 8 为发生闭环的场景图像, 可见本文算法对动态环境具有一定的鲁棒性. 在实时性上, 闭环检测时间阈值限制了工作内存

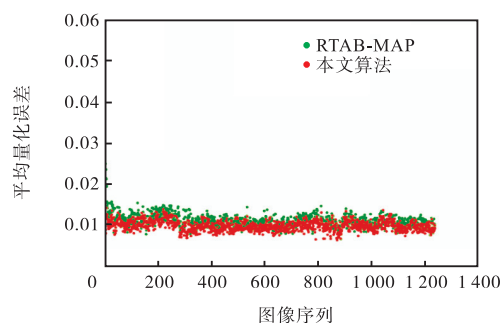
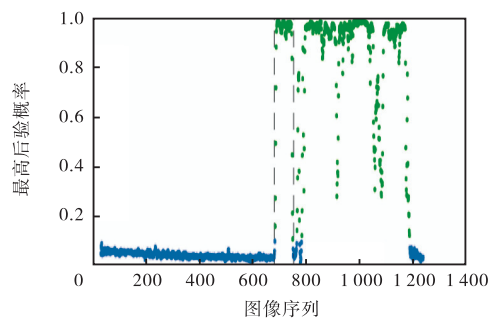
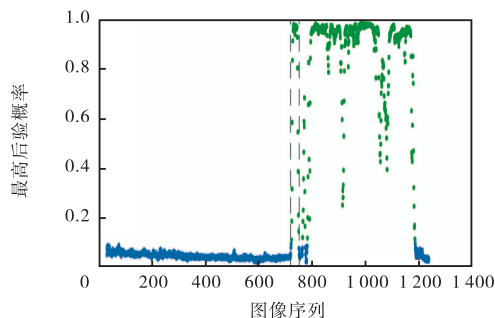


图 6 CiC 动态图像集词袋模型的平均量化误差

Fig.6 Mean quantization error using bag of word under dynamic environment dataset CiC



(a) 贪心策略闭环检测



(b) RTAB-MAP

图 7 CiC 动态图像集闭环检测后验概率对比

Fig.7 Comparison of loop closure detection posterior probability under dynamic environment dataset CiC



(a) 图像序号为 807

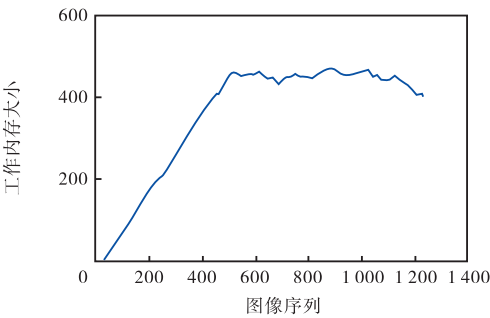


(b) 图像序号为 269

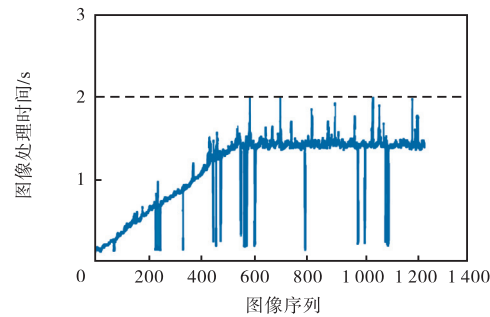
图 8 动态数据集发生闭环的图像

Fig.8 Detected loop closure image under dynamic environment dataset

中图像数目,工作内存大小如图 9(a)所示. 由于当前图像仅和工作内存中图像进行闭环检测,因此绝大部分图像处理时间小于图像采集时间 2 s,如图 9(b)所示,满足实时性要求.



(a) 工作内存



(b) 图像处理时间

图 9 动态数据集实时性分析

Fig.9 Real-time analysis under dynamic environment dataset

本文所提出基于贪心策略的闭环检测算法与其他一些闭环检测算法性能分析如表 5 所示,本文算法结果在所有数据集上正确率 100%的最大召回率均比 FAB-MAP、IAB-MAP、IBuILD 高,而 Garcia-Fidalgo

在数据集 CiC 上与本文算法相当,但是在数据集 NC 上远不如本文算法,且不具备实时性. 与同样具备实时性的 RTAB-MAP 相比,本文算法在数据集 NC 有 4% 提高,数据集 CiC 有 5% 提高,而另外两个数据集结果相当.

表 5 百分百准确率时召回率

Tab.5 Recall rate at 100% precision

检测算法	实时性	召回率 %			
		NC	CiC	L6I	L6O
FAB-MAP 算法	不具备	46	37		
IAB-MAP 算法	不具备			80	71
Garcia-Fidalgo 算法	不具备	53	88		
IBuILD 算法	不具备		39	91	77
RTAB-MAP 算法	具备	89	81	98	95
本文算法	具备	93	86	98	94

4 结 语

闭环检测作为 SLAM 的重要环节,对减小移动机器人的累积误差、实现地图的一致性至关重要. 本文提出一种在线构建词典的闭环检测算法,基于贪心策略优先处理 Surf 描述与已有单词差别最大的特征点,改进最近邻约束更新视觉单词,生成了表征能力强、量化误差小的视觉词典,提升了在动态图像集与视觉混淆多发生的图像集上的闭环检测最大召回率.

参考文献:

[1] 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.

[2] Botterill T, Mills S, Green R. Bag-of-words-driven, single-camera simultaneous localization and mapping [J]. *Journal of Field Robotics*, 2011, 28(2): 204-226.

[3] Cummins M, Newman P. Probabilistic appearance based navigation and loop closing[C]// *IEEE International Conference on Robotics and Automation*. Roma, Italy, 2007: 2042-2048.

[4] Cummins M, Newman P. FAB-MAP: Probabilistic localization and mapping in the space of appearance[J]. *International Journal of Robotics Research*, 2008, 27(6): 647-665.

[5] Cummins M, Newman P. FAB-MAP: Appearance-based place recognition and mapping using a learned visual vocabulary model[C]// *International Conference on Machine Learning*. Haifa, Israel, 2010: 3-10.

- [6] Cummins M, Newman P. Accelerating FAB-MAP with concentration inequalities[J]. *IEEE Transactions on Robotics*, 2010, 26(6): 1042-1050.
- [7] Cummins M, Newman P. Appearance-only SLAM at large scale with FAB-MAP 2.0[J]. *International Journal of Robotics Research*, 2011, 30(9): 1100-1123.
- [8] Angeli A, Doncieux S, Meyer J A, et al. Real-time visual loop-closure detection [C]// *IEEE International Conference on Robotics and Automation*. Pasadena, USA, 2008: 1842-1847.
- [9] Angeli A, Filliat D, Doncieux S, et al. A fast and incremental method for loop-closure detection using bags of visual words[J]. *IEEE Transactions on Robotics*, 2008, 24(5): 1027-1037.
- [10] Labbé M, Michaud F. Memory management for real-time appearance-based loop closure detection [C]// *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco, USA, 2011: 1271-1276.
- [11] Labbé M, Michaud F. Appearance-based loop closure detection for online large-scale and long-term operation[J]. *IEEE Transactions on Robotics*, 2013, 29(3): 734-745.
- [12] Garcia-Fidalgo E, Ortiz A. On the use of binary feature descriptors for loop closure detection[C]// *IEEE Emerging Technology and Factory Automation*. Barcelona, Spain, 2014: 1-8.
- [13] Khan S, Wollherr D. IBuILD: Incremental bag of binary words for appearance based loop closure detection[C]// *IEEE International Conference on Robotics and Automation*. Washington, USA, 2015: 5441-5447.
- [14] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features(SURF) [J]. *Computer Vision & Image Understanding*, 2008, 110(3): 346-359.
- [15] Luxburg U V. A tutorial on spectral clustering[J]. *Statistics and Computing*, 2007, 17(4): 395-416.
- [16] Muja M, Lowe D G. Scalable nearest neighbor algorithms for high dimensional data[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 36(11): 2227-2240.
- [17] Cummins M. Probabilistic Localization and Mapping in Appearance Space[D]. Oxford: Balliol College, University of Oxford, 2009.

(责任编辑: 王晓燕)