

DOI: 10.13973/j.cnki.robot.2017.0023

基于场景显著区域的改进闭环检测算法

李维鹏, 张国良, 姚二亮, 徐 君, 林志林

(火箭军工程大学, 陕西 西安 710025)

摘要: 针对闭环路径不一致、闭环视角变化较大的 VSLAM (视觉同时定位与地图创建) 闭环检测问题, 提出了一种基于场景显著区域的改进闭环检测算法. 首先设计了基于场景显著区域的闭环检测架构, 并构建了相应的闭环全概率模型. 其次, 提出了融合特征跟踪率与 RGB 直方图匹配的关键帧选取方法, 降低了闭环检测过程的信息冗余. 然后, 采用基于逆向索引的预匹配场景选取方法, 极大地提高了闭环检测的实时性. 最后, 根据本文的场景显著区域的几何特点给出了显著区域之间的观测几何匹配概率, 提高了闭环检测算法对歧义场景的分辨能力. 与 FAB-MAP 2.0 的对比实验表明, 在保证较高的闭环准确率的条件下, 本文算法能够有效提高对不重合路径闭环以及大视角变化的闭环的召回率, 并具有良好的实时性.

关键词: 同时定位与地图创建; 闭环检测; 逆向索引; 显著区域; 大视角变化

中图分类号: TP242

文献标识码: A

文章编号: 1002-0446(2017)-01-0023-13

An Improved Loop Closure Detection Algorithm Based on Scene Salient Regions

LI Weipeng, ZHANG Guoliang, YAO Erliang, XU Jun, LIN Zhilin

(Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: An improved loop closure detection algorithm based on scene salient regions is proposed for VSLAM (visual simultaneous localization and mapping) loop closures without consistent path and similar camera view. Firstly, a scene salient region based loop closure detection architecture is structured together with its probabilistic model. Then a key frame selecting method which combines feature tracking rate and RGB histogram matching is proposed to reduce the detection information redundancy. Next, a pre-matched scene selecting method based on inverted index is presented, which greatly improves the real-time performance of loop closure detection. Finally, according to the geometric property of the salient region in this paper, a geometric matching probability in observation is introduced to improve the algorithm's ability of distinguishing perceptual aliasing scenes. The comparative experiments show that the proposed algorithm gets a better recall ratio compared with FAB-MAP 2.0 in a condition of high precision and good real-time performance, when facing loop closures without consistent path and similar camera view.

Keywords: simultaneous localization and mapping; loop closure detection; inverted index; salient region; large view change

1 引言 (Introduction)

闭环检测 (loop closure detection) 作为 SLAM (同时定位与地图创建) 问题的关键环节和基础问题, 对消除机器人位姿估计的累积误差, 降低地图不确定性至关重要^[1,2]. 近年来, 随着视觉 SLAM 技术的不断发展, 基于视觉的闭环检测技术也有了长足的进步. 由于闭环检测的应用场景更加复杂, 环境规模越来越大, 其对图像信息压缩的要求也越来越高, 而能够满足此要求的基于图像特征相似性的闭环检测便在实际应用中占据了主导地位^[3].

由于检索效率的原因, 基于图像特征相似性的

闭环检测算法大多采用 BoVW (bag of visual word) 特征单词化结构^[4-8]. 其中 Angeli 等^[4]探讨了增量式视觉词典构建的相关问题, Cummins 等^[5-6]采用 Chow-Liu 树分析了 BoVW 中单词之间的相互关系并给出了基于 Chow-Liu 树继承关系的闭环观测概率计算方法, Labbé 等人^[7-8]通过设计分级内存管理机制提高了闭环检测的实时性. 总体而言, 上述文献的实验对象集中于对路径一致、观测视角相近的闭环进行检测, 而对于在实际情况中经常出现的路径不一致和大视角变化的闭环检测的研究却相对较少.

如图 1 所示, 出现路径不一致以及大视角变化

的闭环时, 因为视线遮挡和观测范围受限等因素, 容易导致观测图像之间的显著差异和闭环图像不连续. 此时, 采用传统的基于图像与图像匹配与贝叶斯滤波的闭环检测方式将难以检测到闭环的发生. 针对该问题, 本文将单幅图像分解为若干个显著区域, 并将连续 l 个关键帧的显著区域构成一个场景, 采用图像到场景的闭环检测策略, 从而克服闭环检测时因观测视角不同而带来的困难.

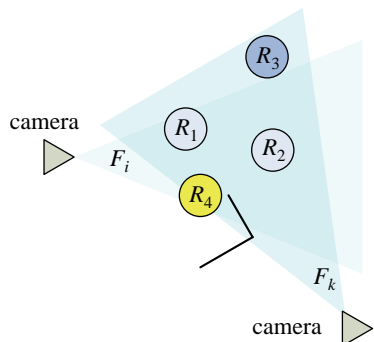


图1 大视角变化闭环下的观测显著差异

Fig.1 Distinct observation difference in loop closure with large view change

如图2所示, 本文定义的场景描述由4个级别构成: 基于BoVW的视觉单词、显著区域、关键帧以及场景. 其中显著区域是由特征较为密集的区域内的视觉单词组成的集合, 场景由若干个关键帧的显著区域组成. 其详细构建过程将在第3节予以介绍. 下面首先介绍本文采用的闭环概率模型.

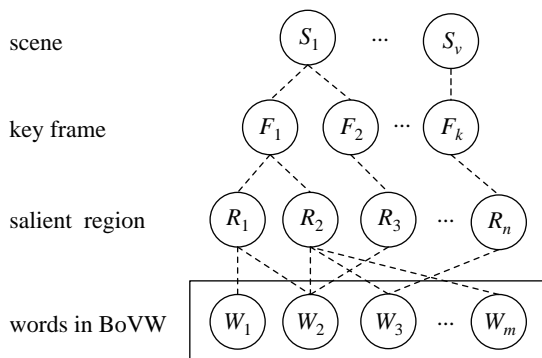


图2 本文定义的场景描述结构

Fig.2 The scene description structure defined in this article

2 闭环概率模型 (Loop closure probabilistic model)

本文的程序框架如图3所示. 首先对所采集的图像进行特征提取, 其次根据特征跟踪率与RGB统计直方图判断是否为关键帧, 随后分别进行图像特征单词化与显著区域选取, 之后更新从单词到场景显著区域的增量式逆向索引, 并根据逆向索引检

索预闭环场景, 最后通过贝叶斯更新计算各个预闭环场景的闭环概率. 要完成上述步骤, 首先需要构建基于场景显著区域的闭环概率模型.

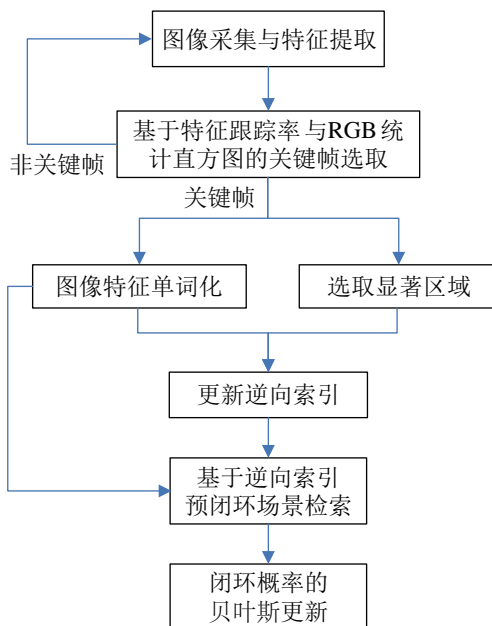


图3 基于场景显著区域的闭环检测流程

Fig.3 The procedure of loop closure detection based on scene salient regions

对于闭环概率模型问题, 文[8]采用贝叶斯滤波对闭环概率进行更新. 该方法较为简单, 闭环概率更新的计算量较小, 但文中并未说明归一化参数的选取问题. 文[6]则提供了闭环概率更新的全概率模型:

$$P(S_i|F^k) = \frac{P(F_k|S_i, F^{k-1})P(S_i|F^{k-1})}{P(F_k|F^{k-1})} \quad (1)$$

式(1)表示在前 k 个关键帧 F^k 下, k 时刻图像与历史场景 S_i 发生闭环的概率. 其中 F_k 为 k 时刻关键帧, $P(F_k|S_i, F^{k-1})$ 为闭环观测概率, $P(S_i|F^{k-1})$ 为闭环先验概率, $P(F_k|F^{k-1})$ 为与歧义场景相关的归一化参数.

相比于文[8]采用的贝叶斯滤波方式, 式(1)较为完整地体现了闭环概率的迭代过程. 但由于本文应用环境为帧到场景的闭环, 在估计 k 时刻与场景 S_i 发生闭环 $L_k = i$ 的概率时, 历史关键帧集合 F^k 与历史场景集合 S^v 之间的关系都已知. 因此, 本文给出基于关键帧和场景的闭环概率模型更为完整的表述形式:

$$P(L_k = i|S^v, F^k) = \frac{P(F_k|L_k = i, S^v, F^{k-1})P(L_k = i|S^v, F^{k-1})}{P(F_k|S^v, F^{k-1})} \quad (2)$$

其中 F_k 为 k 时刻关键帧, S_i 为第 i 个历史场景, k 与 v 都从 0 开始标记. 为了计算闭环的观测概率, 假设各个关键帧的观测之间相互独立, 得到:

$$\begin{aligned} & P(F_k|L_k=i, S^v, F^{k-1}) \\ &= \frac{P(F_k|S^v, F^{k-1})P(L_k=i|F_k, S^v)}{P(L_k=i|S^v)} \end{aligned} \quad (3)$$

将式 (3) 代入式 (2) 得到:

$$\begin{aligned} & P(L_k=i|S^v, F^k) \\ &= \frac{P(L_k=i|F_k, S^v)P(L_k=i|S^v, F^{k-1})}{P(L_k=i|S^v)} \end{aligned} \quad (4)$$

其中 $P(L_k=i|F_k, S^v)$ 为闭环观测概率, $P(L_k=i|S^v, F^{k-1})$ 为闭环先验概率, $P(L_k=i|S^v)$ 表示拓扑闭环概率.

2.1 闭环观测概率

为了计算闭环观测概率, 首先分析本文闭环观测概率的具体含义:

$$P(L_k=i|F_k, S^v) \quad (5)$$

式 (5) 表示: 在已知当前观测 F_k 与历史场景集 S^v 时 F_k 与场景 S_i 发生闭环的概率. 由于本文中 S_i 与 F_k 由其包含的显著区域描述, 闭环观测概率完全依赖于 S_i 与 F_k 包含的显著区域的相似度:

$$\begin{aligned} & P(L_k=i|F_k, S^v) \\ &= P(R \in S_i | R \in F_k, F_k, S_i) \\ &= P(R' \in S_i | R' = R, R \in F_k, F_k, S_i) \cdot \\ & P(R' = R | R \in F_k, F_k, S_i) \end{aligned} \quad (6)$$

其中 $P(R' \in S_i | R' = R, R \in F_k, F_k, S_i)$ 表示与场景匹配的显著区域属于该场景的概率, 该概率与对显著区域观测的一致性相关, 本文将之解释为显著区域满足观测几何约束的概率. $P(R' = R | R \in F_k, F_k, S_i)$ 表示显著区域的匹配概率, 该概率采用显著区域所包含的视觉单词的相似性进行度量.

式 (6) 中的两个概率将在第 3 节中予以讨论, 下面介绍闭环先验概率的计算.

2.2 闭环先验概率

闭环先验概率表示当前时刻的闭环概率与历史闭环概率之间的关系, 该概率还与闭环场景 S_i 的邻接拓扑关系有关:

$$\begin{aligned} & P(L_k=i|S^v, F^{k-1}) \\ &= \sum_j P(L_k=i|L_q^j=j, S^v, F^{k-1})P(L_{q(j)}=j|S^v, F^q) \\ &= \sum_j P(L_k=i|L_{q(j)}=j, S^v)P(L_{q(j)}=j|S^v, F^q) \end{aligned} \quad (7)$$

其中 $P(L_k=i|L_{q(j)}=j, S^v)$ 表示在 S^v 所描述的拓扑关系下, 当场景 S_j 在 $q(j)$ 时刻发生闭环时, S_j 在 k 时刻发生闭环的条件概率, $P(L_{q(j)}=j|S^v, F^q)$ 表示历史场景 S_j 在上次出现闭环观测的 $q(j)$ 时刻的闭环概率.

文 [8] 与本文采用了类似的闭环先验概率计算方式, 所不同的是, 文 [8] 仅考虑了相邻时刻的闭环概率对当前闭环概率的影响, 而本文则考虑了所有场景上一次出现闭环的概率对当前闭环概率的影响. 由于对某个场景及其邻接场景的闭环观测不一定连续出现, 相比而言, 本文的闭环先验概率的计算方式对于非一致路径闭环, 尤其是对路径多段交叉的闭环, 有着更高的召回率.

引入 $L = -1$ 表示关键帧没有与任何场景发生闭环的情形, 则 $P(L_k=i|L_{q(j)}=j, S^v)$ 的计算分为 4 种情形. 下面分别予以介绍:

$$1) P(L_k=-1|L_{q(j)}=j, S^v)$$

表示在 $q(j)$ 时刻图像与 S_j 发生闭环的条件下, k 时刻图像未与任何历史场景发生闭环的概率. 该概率可以看作由 $q(j)$ 时刻到 k 时刻相机从已知范围 S^v 内离开的概率. 随着时间的推移, 一方面已探索范围 S^v 不断扩大, 另一方面与已知范围相接触的未知范围也不断扩大, 在该概率上两者作用相反. 鉴于当前场景不参与闭环检测, 为了简化问题, 本文认为该概率与单个历史场景的规模成反比, 该概率定义为

$$P(L_k=-1|L_{q(j)}=j, S^v) = \frac{1}{n_F} \quad (8)$$

其中 n_F 表示单个历史场景包含的关键帧个数.

$$2) P(L_k=i|L_{q(j)}=j, S^v), i \in \{0, 1, \dots, v-1\}$$

表示在 $q(j)$ 时刻图像与 S_j 发生闭环的条件下, k 时刻与 S_j 发生闭环的概率. 由于 $q(j)$ 到 k 时刻相机所能到达的场景与运动范围相关, 故该概率与 S_i 、 S_j 之间的拓扑距离有关, 本文定义其满足标准差为 $\sigma_i = 1.6\sqrt{\frac{k-q(j)}{n_F}}$ 的正态分布.

此外, $P(L_k=i|L_{q(j)}=j, S^v)$ 需要根据 $P(L_k=-1|L_{q(j)}=j, S^v)$ 进行归一化, 即:

$$\begin{aligned} & \sum_{i \neq -1} P(L_k=i|L_{q(j)}=j, S^v) \\ &= 1 - P(L_k=-1|L_{q(j)}=j, S^v) \\ &= 1 - \frac{1}{n_F} \end{aligned} \quad (9)$$

$$3) P(L_k=-1|L_q=-1, S^v)$$

表示在 q 时刻图像未发生闭环的条件下, k 时刻图像也未发生闭环的概率. 根据文 [4], 该概率直接赋值为

$$P(L_k = -1 | L_q = -1, S^v) = 0.9 \quad (10)$$

$$4) P(L_k = i | L_q = -1, S^v), i \in \{0, 1, \dots, v-1\}$$

由于

$$\sum_i P(L_k = i | L_q = j, S^v) = 1 \quad (11)$$

且闭环检测范围内共有 v 个场景, 因此

$$P(L_k = i | L_q = -1, S^v) = \frac{0.1}{v}, i \neq -1 \quad (12)$$

2.3 拓扑闭环概率

拓扑闭环概率表示因场景的邻接关系而出现闭环的概率, 该概率与场景的闭环转移概率和邻接拓扑结构有关:

$$P(L_k = i | S^v) = \sum_j P(L_k = i | L_{q(j)} = j, S^v) P(L_{q(j)} = j | S^v) \quad (13)$$

记 $a_{i,j} = P(L_k = i | L_{q(j)} = j, S^v)$, $p_i = P(L_k = i | S^v)$. 将 2.2 节所给出的闭环拓扑先验概率代入式 (13), 得到:

$$\begin{cases} p_i = \sum_{j \neq -1} a_{i,j} p_j + a_{i,-1} p_{-1} \\ \sum_i p_i = v \end{cases} \quad (14)$$

其中 v 为历史场景个数, 进而:

$$p_i = \sum_{j \neq -1} (a_{i,j} - a_{i,-1}) p_j + v a_{i,-1}, i = 0, 1, \dots, v-1 \quad (15)$$

记:

$$\mathbf{B} = \begin{bmatrix} a_{0,0} - a_{0,-1} & \cdots & a_{0,v-1} - a_{0,-1} \\ \vdots & \ddots & \vdots \\ a_{v-1,0} - a_{v-1,-1} & \cdots & a_{v-1,v-1} - a_{v-1,-1} \end{bmatrix},$$

$$\mathbf{c} = \begin{bmatrix} a_{0,-1} \\ \vdots \\ a_{v-1,-1} \end{bmatrix}, \mathbf{p} = \begin{bmatrix} p_0 \\ \vdots \\ p_{v-1} \end{bmatrix}$$

由式 (15) 得到:

$$\mathbf{p} = \mathbf{B}\mathbf{p} + \mathbf{c} \quad (16)$$

从而解得各个场景的拓扑闭环概率 \mathbf{p} . 由式 (13) 可以看出: 场景 S_i 的邻近节点越多, 邻近节点到 S_i 的闭环转移概率越大, 则 S_i 的拓扑闭环概率越高. 这与通常情况下多段路径交汇处的闭环概率较高的实际情况相一致.

3 显著区域选取与场景描述 (Salient region extraction and scene description)

出现大视角的闭环时, 不同视角的观测因其视线遮挡和背景不同, 会造成单幅图像观测范围和观测结果的显著差异. 因此, 需要融合连续图像的信息构成场景, 并采用图像到场景的闭环检测方式代替传统的基于图像间匹配的闭环检测.

对于本文应用背景而言, 图像融合有两种常用方法. 其一是直接融合图像特征. 该方法相对较为简单, 但由于使用了所有的图像特征, 产生了信息冗余, 且容易在图像融合过程中对同一特征重复度量, 因此其对场景的描述易出现偏差. 第二种方案是将分布较为密集的特征聚集为一个显著区域表示. 该方法仅使用部分特征, 所使用的信息较少, 且能根据显著区域观测时所应满足的对极几何关系提高对歧义场景的分辨能力. 尽管显著区域法有着多种优势, 但该方法存在对显著区域的选取策略问题, 其实现相对较为繁琐. 下面首先讨论本文的显著区域选取策略.

3.1 显著区域的选取与场景描述

本文显著区域的选取主要依赖 SURF (加速鲁棒特征) 的分布状况. 其选取过程如图 4 所示.

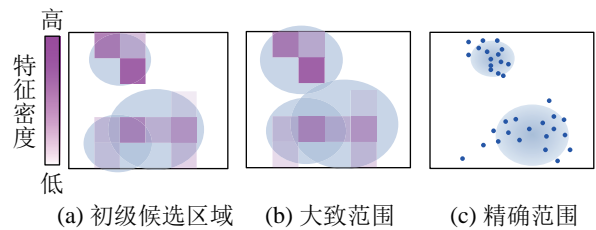


图 4 基于特征分布的显著区域选取过程

Fig.4 Process of salient region extraction based on feature distribution

首先, 选取不同模板尺度下, 特征分布密度均值局部极大的区域为初级候选区域. 随后, 根据初级候选区域包含的网格特征分布密度, 计算大致的显著区域中心坐标:

$$\mathbf{p}_R^{[F]} = \frac{\sum_i \rho_{g,i} \mathbf{p}_{g,i}^{[F]}}{\sum_i \rho_{g,i}} \quad (17)$$

其中 $\mathbf{p}_R^{[F]}$ 表示显著区域 R 的中心在图像坐标系 F 下的坐标, $\mathbf{p}_{g,i}^{[F]}$ 、 $\rho_{g,i}$ 分别为初级候选区域所包含的第 i 个网格的中心坐标及其 SURF 特征分布密度.

在获得大致的显著区域中心坐标后, 需获得显著区域的大致半径. 根据显著区域表示特征较为集

中的区域的实际情况, 本文采用半径增长法确定其大致半径. 由式 (17) 得到的 $\mathbf{p}_R^{[F]}$ 为中心逐渐增大显著区域半径, 直到其半径边界上的网格特征分布密度都低于区域的平均特征分布密度, 或半径达到预设的最大值.

在获得显著区域的大致范围后, 需进一步计算其精确值. 显著区域的精确范围采用改进的均值漂移算法求解, 具体流程如图 5 所示.

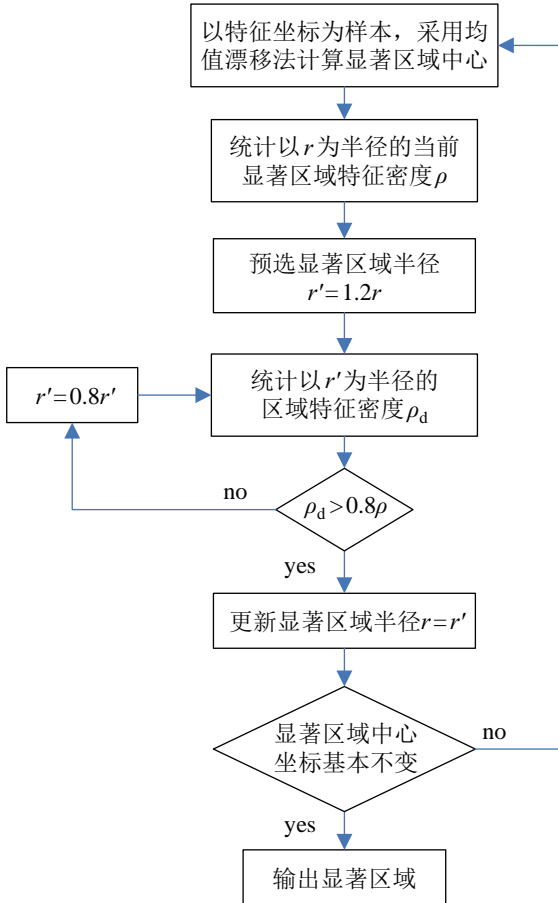


图 5 求解显著区域精确范围的均值漂移算法流程图
Fig.5 Flowsheet of the mean-shift algorithm for solving the accurate salient regions

将本文显著区域选取方法分别与分水岭算法、基于 DCT (离散余弦变换) 的图像分割^[9] 进行比较, 结果如图 6 所示.

由图 6 可以看出, 从分割效果来看, 基于 DCT 变换的图像分割效果最好, 但该算法对噪声敏感, 并且需要预先设定提取目标的尺寸, 导致分割结果关于观测距离变化的鲁棒性较差. 此外, 其对像素为 240×320 的图像处理时间长达 0.2 s, 无法满足实时性的需求. 分水岭算法速度最快, 但并未按照目标物体分割区域, 且需要采用其他算法预先分析种子点, 效果较差. 相比而言, 本文基于特征分布

的显著区域选取算法成功地从图像中分割出了富含特征的稳定区域, 且用时较短, 达到了快速提取图像显著区域的目的.

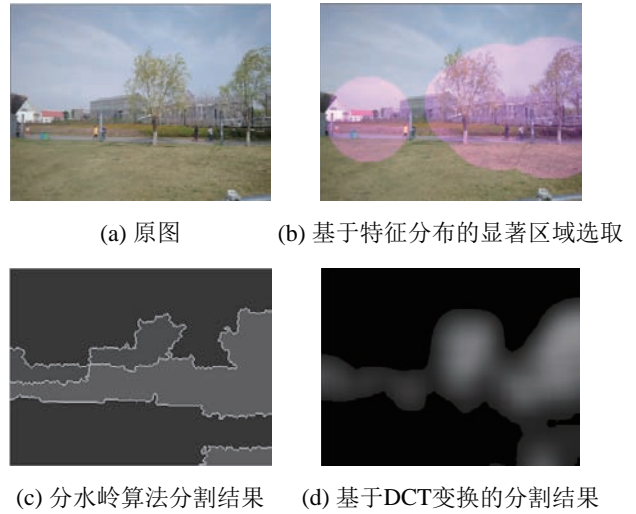


图 6 图像分割算法效果对比图

Fig.6 Results of different image segmentation algorithms

在获得显著区域后, 采用 BoVW 将显著区域范围内的 SURF 特征描述向量单词化, 采用视觉单词描述显著区域, 进而由若干关键帧的显著区描述场景.

下面介绍本文关键帧的选取方法.

3.2 关键帧的选取

与传统的基于位姿变化和简单采样的关键帧选取不同, 本文的关键帧选取依赖于当前图像与上一关键帧之间的相似度. 为了提高相似度计算结果的可靠性, 本文采用基于特征跟踪率与 RGB 统计直方图的双重参考策略. 其相似度函数定义为

$$\gamma = \alpha_{\text{color}} |h_k - h_c| + \alpha_{\text{SURF}} \cdot \left[\left(1 - \frac{m_M}{\max(m_k, m_c)} \right) + \frac{1}{\max(m_k, m_c)} \sum_{i=1}^{m_M} \| \mathbf{x}_i^k - \mathbf{x}_i^c \| \right] \quad (18)$$

其中 α_{SURF} 为 SURF 特征匹配的得分权重, α_{color} 为图片色彩直方图匹配的得分权重, 且 $\alpha_{\text{SURF}} + \alpha_{\text{color}} = 1$. m_k 、 m_c 分别为第 k 个关键帧与当前图片的 SURF 特征数目, m_M 表示两幅图片匹配的 SURF 特征个数. \mathbf{x}_i^k 与 \mathbf{x}_i^c 为匹配的 SURF 特征描述向量, h_k 、 h_c 表示两幅图像的 RGB 统计直方图. 所采用的 SURF 特征跟踪方法参考文 [10] 的 ORB (oriented FAST and rotated BRIEF) 特征跟踪方法.

在式 (18) 中, α_{SURF} 越大, SURF 特征的差异对于关键帧选取的影响越大, RGB 统计直方图的差异对关键帧选取的影响越小. 鉴于 SURF 特征数目较

少时, 特征匹配过程中的随机因素容易影响关键帧选取的合理性, 应当根据 SURF 特征数目适当调整 α_{SURF} , 因此本文选取:

$$\alpha_{\text{SURF}} = \frac{1}{2} - \frac{1}{2} \exp\left(-\frac{\min(m_k, m_c)}{m_e}\right) \quad (19)$$

其中 m_e 表示期望的 SURF 特征提取个数.

根据实验效果, 本文的关键帧选取条件为 $\gamma > 0.3$.

4 基于逆向索引的预匹配场景选取 (Pre-matched scene selection based on inverted index)

为了计算闭环概率, 首先需要选取与当前图像 F_k 可能发生闭环的历史场景. 由于历史场景由若干显著区域组成, 要获得可能的闭环场景, 首先需要搜索与当前图像中各个显著区域匹配的历史显著区域.

当前基于视觉的闭环检测算法主要采用 BoVW 模型, 该方法由 Baeza-Yates 等人借鉴文本检索领域的词袋 (BoW) 模型而提出. 尽管 BoVW 采用单词描述图像, 降低了图像匹配的计算量, 但仍未解决从巨量历史图像中快速搜索相似图像的问题.

本文借鉴由 BoVW 扩展的逆向索引技术^[6]实现预匹配历史场景的快速检索, 如图 7 所示. 逆向索引的方法为: 在 BoVW 中标记每一个单词出现过的显著区域及场景, 当需要对当前时刻图像搜索可能的闭环场景时, 只需根据其显著区域包含的单词检索在 BoVW 的某一层级中相同单词个数满足一定比例的相似显著区域, 进而将各个相似显著区域所出现的场景作为预匹配历史场景.

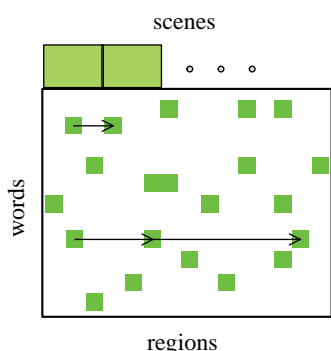


图 7 基于逆向索引的预匹配场景检索

Fig.7 Pre-matched scene selection based on inverted index

根据图 7 方式获得预匹配场景后, 还需要剔除掉包含的预匹配显著区域明显少于当前图像显著区域个数的场景, 余下的场景继续参与闭环概率的计算.

5 显著区域匹配概率 (Matching probability of salient regions)

要获得闭环观测概率, 需要计算发生闭环时各个显著区域的匹配概率. 根据 2.1 节式 (6), 将显著区域的匹配概率分为两部分: 基于 BoVW 得分的匹配概率与观测几何匹配概率. 由于两种概率计算所依赖的信息相互独立, 显著区域 R 与 R' 之间的匹配概率为

$$P(L_k = i | F_k, S^v) = P(R' = R | R \in F_k, F_k, S_i) \cdot P(R' \in S_i | R' = R, R \in F_k, F_k, S_i) \quad (20)$$

其中, $P(R' = R | R \in F_k, F_k, S_i)$ 为基于 BoVW 单词的显著区域匹配概率; $P(R' \in S_i | R' = R, R \in F_k, F_k, S_i)$ 为显著区域满足观测几何约束的概率, 本文将其称为观测几何匹配概率. 下面分别探讨两种概率.

5.1 基于 BoVW 单词的显著区域匹配概率

要获得基于 BoVW 单词的显著区域匹配概率, 首先需要给出基于 BoVW 单词的显著区域匹配评分函数. 根据文 [11] 提出的分层 TF-IDF (term frequency-inverse document frequency) 熵金字塔得分匹配, 显著区域 R 与 R' 之间的匹配得分满足:

$$\begin{cases} \phi(R, R') = \varphi^L(R, R') + \sum_{l=1}^{L-1} \frac{1}{k^{L-l}} (\varphi^l(R, R') - \varphi^{l+1}(R, R')) \\ \varphi^l(R, R') = \sum_{i=1}^k \min\{w_i^l(R), w_i^l(R')\} \\ w_i^l(X) = \frac{n_i}{n} \log \frac{N}{N_i} \end{cases} \quad (21)$$

其中 $\phi(R, R')$ 为 TF-IDF 匹配得分, $\varphi^l(R, R')$ 为显著区域在第 l 层的匹配得分. n_i 表示显著区域 X 中第 i 个单词的出现个数, N_i 为该单词出现过的显著区域个数; n 为某个显著区域 X 包含的单词个数, N 表示显著区域的总数.

由于该评分函数为非归一化函数, 且易受感知歧义干扰, 由此直接计算闭环概率较为困难^[12], 因而本文采用该函数的归一化形式:

$$\begin{cases} \phi(R, R') = \varphi^L(R, R') + \sum_{l=1}^{L-1} \frac{1}{k^{L-l}} (\varphi^l(R, R') - \varphi^{l+1}(R, R')) \\ \varphi^l(R, R') = \frac{(\mathbf{W}^l(R))^T \mathbf{W}^l(R')}{\|\mathbf{W}^l(R)\| \|\mathbf{W}^l(R')\|} \\ \mathbf{W}^l(X) = (w_i^l) \\ w_i^l(X) = \frac{n_i}{n} \log \frac{N}{N_i} \end{cases} \quad (22)$$

其中 $\mathbf{W}^l(X) = (w_i^l)$ 表示显著区域 X 在 BoVW 第 l 层的视觉描述向量, $|\mathbf{W}^l(R)|$ 表示视觉描述向量 $\mathbf{W}^l(X)$ 的模. 经过归一化处理, 分层 TF-IDF 熵金字塔得分在 $[0, 1]$ 之间, 且得分越大时显著区域越相似. 根据式 (22) 的匹配得分结果, 基于 BoVW 匹配得分的显著区域匹配概率为

$$P(R' = R | R \in F_k, F_k, S_i) = \begin{cases} \frac{\phi(R, R') - \sigma_\phi}{\mu_\phi}, & \phi(R, R') \geq \sigma_\phi + \mu_\phi \\ 1, & \text{else} \end{cases} \quad (23)$$

其中 μ_ϕ 、 σ_ϕ 为与 R 相关的所有匹配得分的均值与方差.

5.2 观测几何匹配概率

根据 3.1 节, 场景 S_i 中保存了在每一个关键帧中出现的显著区域的归一化观测坐标 $\mathbf{p}_R^{[O]}$, 若当前图像 F_k 与历史场景 S_i 相匹配, 且较多显著区域包含于 S_i 中某一个关键帧 $F_h \subset S_i$, 则两幅图像中相匹配的显著区域 $R_i \in F_k$, $R'_i \in F_h$ 必须满足几何约束:

$$\begin{cases} \mathbf{p}_h^{[C]} = \mathbf{R}_{h,k} \mathbf{p}_k^{[C]} + \mathbf{t}_{h,k} \\ \frac{r_h}{d_h} = \frac{r_k}{d_k} \end{cases} \quad (24)$$

其中 $\mathbf{p}_h^{[C]}$ 、 $\mathbf{p}_k^{[C]}$ 分别为 h 、 k 时刻相机坐标系下显著区域中心的空间坐标, $\mathbf{R}_{h,k}$ 、 $\mathbf{t}_{h,k}$ 分别为相机位姿的旋转与平移变换, r_h 、 r_k 分别表示两个显著区域观测半径, d_h 、 d_k 分别表示两个显著区域中心到相机平面的距离.

取归一化投影坐标为 $\mathbf{p}^{[O]} = \frac{1}{d} \mathbf{p}^{[C]}$, 得到:

$$\begin{bmatrix} p_{hx}^{[O]} \\ p_{hy}^{[O]} \\ 1 \end{bmatrix} = \frac{r_k}{r_h} \begin{bmatrix} \mathbf{R}_{h,k}(1) \\ \mathbf{R}_{h,k}(2) \\ \mathbf{R}_{h,k}(3) \end{bmatrix} \mathbf{p}_k^{[O]} + \frac{1}{d_h} \begin{bmatrix} t_{h,kx} \\ t_{h,ky} \\ t_{h,kz} \end{bmatrix} \quad (25)$$

其中 $p_{hx}^{[O]}$ 、 $p_{hy}^{[O]}$ 为 $\mathbf{p}_h^{[O]}$ 在 x 、 y 方向上的分量, $t_{h,kx}$ 、 $t_{h,ky}$ 、 $t_{h,kz}$ 分别为 $\mathbf{t}_{h,k}$ 在 x 、 y 、 z 方向上的分量, $\mathbf{R}_{h,k}(i)$ 为旋转矩阵 $\mathbf{R}_{h,k}$ 的第 i 行. 由式 (25) 第 3 行得到:

$$\frac{1}{d_h} = \frac{1}{t_{h,kz}} - \frac{r_k}{t_{h,kz} r_h} \mathbf{R}_{h,k}(3) \mathbf{p}_k^{[O]} \quad (26)$$

将式 (26) 代入式 (25), 得到:

$$\begin{bmatrix} p_{hx}^{[O]} \\ p_{hy}^{[O]} \end{bmatrix} = \frac{r_k}{r_h} \begin{bmatrix} \mathbf{R}_{h,k}(1) - \frac{t_{h,kx}}{t_{h,kz}} \mathbf{R}_{h,k}(3) \\ \mathbf{R}_{h,k}(2) - \frac{t_{h,ky}}{t_{h,kz}} \mathbf{R}_{h,k}(3) \end{bmatrix} \mathbf{p}_k^{[O]} + \begin{bmatrix} \frac{t_{h,kx}}{t_{h,kz}} \\ \frac{t_{h,ky}}{t_{h,kz}} \end{bmatrix} \quad (27)$$

将式 (27) 表示为矩阵向量形式:

$$\begin{cases} \mathbf{p}_h^{[O]} = \frac{r_k}{r_h} \boldsymbol{\Psi}_{h,k} \mathbf{p}_k^{[O]} + \boldsymbol{\tau}_h \\ \boldsymbol{\Psi}_{h,k} = \begin{bmatrix} \mathbf{R}_{h,k}(1) - \frac{t_{h,kx}}{t_{h,kz}} \mathbf{R}_{h,k}(3) \\ \mathbf{R}_{h,k}(2) - \frac{t_{h,ky}}{t_{h,kz}} \mathbf{R}_{h,k}(3) \end{bmatrix} \\ \boldsymbol{\tau}_h = \begin{bmatrix} \frac{t_{h,kx}}{t_{h,kz}} & \frac{t_{h,ky}}{t_{h,kz}} \end{bmatrix}^T \end{cases} \quad (28)$$

从而将图像中各个显著区域的归一化重投影误差均值作为观测几何目标函数:

$$\min_{\boldsymbol{\Psi}_{h,k}, \boldsymbol{\tau}_h} f = \frac{1}{n_R^F} \sum_{i=1}^{n_R^F} \left| \frac{r_k}{r_h} \boldsymbol{\Psi}_{h,k} \mathbf{p}_i^{[O]} + \boldsymbol{\tau}_h - \mathbf{p}_i^{[O]} \right|^2 \quad (29)$$

其中 n_R^F 为历史关键帧 F_h 与当前关键帧 F_k 中匹配的显著区域个数. 式 (29) 至少需要求解 3 对匹配显著区域. 若匹配显著区域低于 3 对, 则取 $f = 0.4$.

为了求得目标函数 (29) 的最小值, 将旋转矩阵 $\mathbf{R}_{h,k}$ 表示为

$$\mathbf{R}_{h,k} = \exp([\mathbf{w}]_\times) = \mathbf{I} + \frac{\sin |\mathbf{w}|}{|\mathbf{w}|} [\mathbf{w}]_\times + \frac{1 - \cos |\mathbf{w}|}{|\mathbf{w}|^2} [\mathbf{w}]_\times^2 \quad (30)$$

其中 \mathbf{w} 为旋转向量, $[\mathbf{w}]_\times$ 表示 \mathbf{w} 的反对称矩阵.

最终得到观测几何匹配概率:

$$P(R' \in S_i | R' = R, R \in F_k, F_k, S_i) = \exp\left(-\frac{f^2}{2\sigma_f^2}\right) \quad (31)$$

其中 f 为式 (29) 所述的目标函数值, $\sigma_f = \frac{0.1}{\sqrt{n_R^F}}$.

6 实验与分析 (Experiment and analysis)

本文实验所用传感器为 Logitech Webcam C930e, 运行计算机配置: CPU 为 i3 处理器, 主频 2.5 GHz, 内存 4G, 不使用 GPU 加速, 系统为 Windows 7. 为了客观地评价算法性能, 分别使用公共的 VSLAM 数据集和实际场景数据进行实验. 实验所用 BoVW 为 5 层 4 分支树, 所用特征为 SURF^[13], 对比算法采用 FAB-MAP 2.0.

6.1 VSLAM 数据集实验

实验所用数据集为 KITTI Visual Odometry^[14]/SLAM Evaluation 2012, 属于城市及郊区公路数据集, 包含较多的相似场景.

6.1.1 城市公路数据集实验

城市公路数据集实验采用 KITTI Visual Odometry 的 0 号数据集. 算法获得的高概率闭环分布如图 8 所示, 图中蓝色表示移动平台的运动轨迹, 绿

色表示实际闭环路段，红色和黄色为闭环概率大于90%的路段，且路段的闭环概率由黄到红逐渐增大，其最大概率为1。从图中可以看到，绝大多数高概率闭环都集中于实际闭环路段。

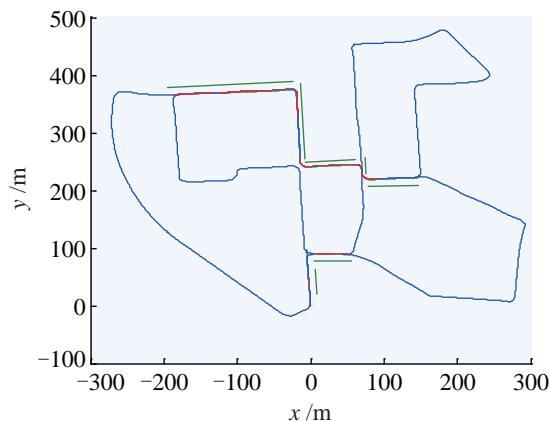


图8 城市公路数据集实验的高概率闭环分布

Fig.8 Distribution of high-probability loop closures in urban dataset experiment

图9是本文的闭环检测算法在城市公路数据集的闭环实例图像，图中绿色圆圈表示检测到的SURF特征，蓝色线段为各个SURF特征相对于其参考关键帧的跟踪轨迹（不是闭环的几何验证基线），紫色区域为根据3.1节算法获得的场景显著区域。从图9可以看出，实时图像与闭环的历史关键帧的显著区域之间存在着较好的对应关系。



(a) 实时图像



(b) 闭环的历史关键帧

图9 基于显著区域匹配的闭环图像

Fig.9 Loop-closed images based on salient region matching

图10为该数据集实验中各个闭环检测算法的准确率—召回率曲线，其中准确率表示所检测到的

闭环中的正确闭环比例，召回率表示所检测到的正确闭环占路径真实闭环个数的比例。图中蓝色、红色、紫色曲线分别对应本文基于显著区域的闭环检测算法、基本FAB-MAP 2.0与包含闭环几何验证的FAB-MAP 2.0。从图中可以看出，在准确率—召回率方面，本文闭环检测算法（蓝色曲线）总体上优于基本FAB-MAP 2.0，但与包含闭环几何验证的FAB-MAP 2.0在高准确率方面存在一定差距，这主要是因为闭环几何验证（尽管较为耗时）剔除了FAB-MAP 2.0中出现的拥有高闭环概率的误闭环，从而保证了其较高的闭环准确率。

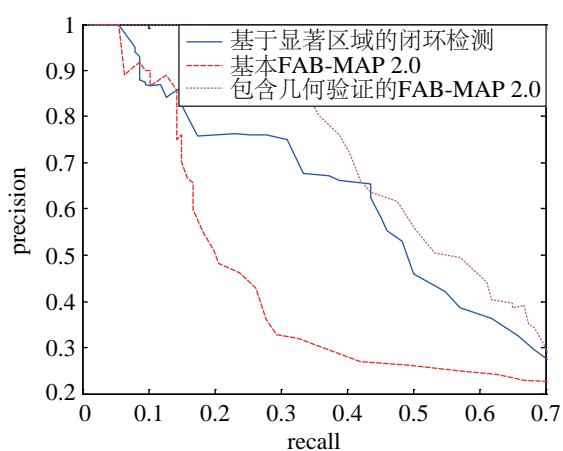


图10 城市公路数据集实验准确率—召回率对比

Fig.10 Contrast of precision-recall ratios in urban dataset experiment

图11为该数据集实验中各个算法的闭环检测用时。可以看到，本文闭环检测用时大于基本FAB-MAP 2.0，但远小于包含闭环几何验证的FAB-MAP 2.0（图11(b)、(c)的用时叠加）。主要因为FAB-MAP 2.0所处理的图像个数小于本文所处理的显著区域个数，因而基本FAB-MAP 2.0对闭环概率的计算较快；但FAB-MAP 2.0在闭环几何验证环节需要对所有拥有高闭环概率的图像中的特征进行对极几何约束检验，因而相对于本文算法其实时性较差。

图12为该数据集实验过程中特征提取及本文闭环检测算法各环节的用时，其中特征提取用时不包括在闭环检测用时之内。由图12可以看出，闭环检测各环节中平均用时最长的是闭环概率的贝叶斯更新，这是由于本文算法需要不断地更新闭环先验概率和拓扑闭环概率，以此满足对闭环观测不连续的非重合路径闭环进行有效检测的目的。将图12(b)~(d)与图12(a)对比，可见本文闭环检测算法各个环节的用时远小于SURF特征提取用时，满足实时性的要求。

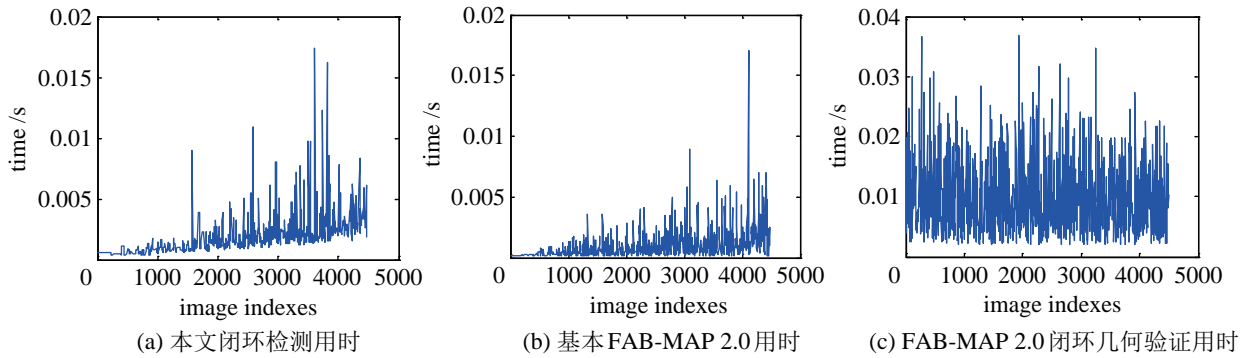


图 11 城市公路数据集实验实时性对比

Fig.11 Contrast of real-time performance in urban dataset experiment

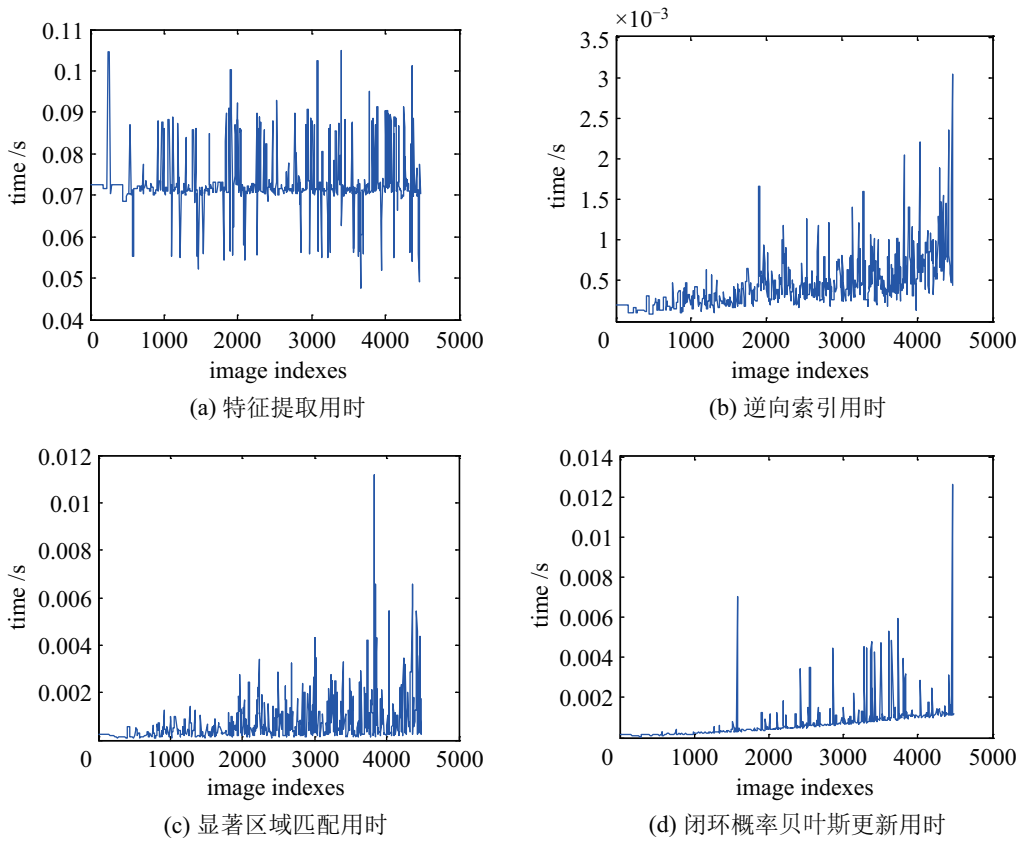


图 12 城市公路数据集实验本文算法各环节用时

Fig.12 Time used for each procedure of the proposed algorithm in urban dataset experiment

6.1.2 郊区公路数据集实验

郊区公路数据集实验采用 KITTI Visual Odometry 的 2 号数据集. 算法获得的高概率闭环分布如图 13 所示, 图中蓝色表示移动平台的运动轨迹, 绿色表示实际闭环路段, 红色和黄色为闭环概率大于 85% 的路段, 且路段的闭环概率由黄到红逐渐增大. 从图中可以看到, 绝大多数高概率闭环都集中于实际闭环路段.

图 14 为该数据集实验中各个闭环检测算法的准确率-召回率曲线. 图中蓝色、红色、紫色曲线

分别对应本文基于显著区域的闭环检测算法、基本 FAB-MAP 2.0 与包含闭环几何验证的 FAB-MAP 2.0. 从图中可以看出, 在准确率-召回率上, 本文闭环检测算法 (蓝色曲线) 在高闭环准确率的部分其召回率优于基本 FAB-MAP 2.0, 但准确率方面总体低于包含闭环几何验证的 FAB-MAP 2.0, 这主要是因为其闭环几何验证 (尽管较为耗时) 剔除了 FAB-MAP 2.0 中出现的拥有高闭环概率的误闭环, 从而保证了其较高的闭环准确率.

图 15 为该数据集实验中各个算法的闭环检测

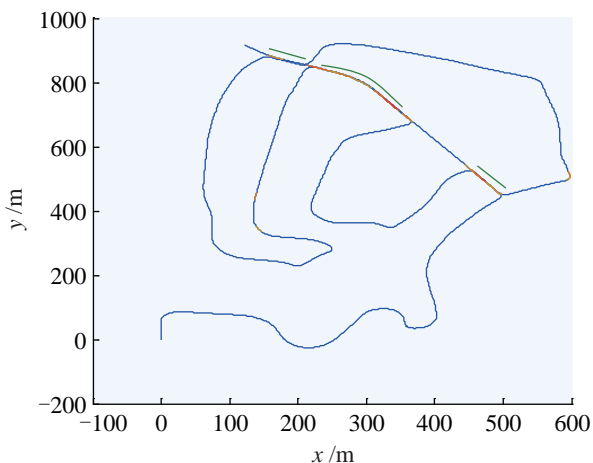


图 13 郊区公路数据集实验的高概率闭环分布

Fig.13 Distribution of high-probability loop closures in suburban dataset experiment

用时. 可以看到, 本文闭环检测用时大于基本 FAB-MAP 2.0, 但小于包含闭环几何验证的 FAB-MAP 2.0 (图 15(b)、(c) 的用时叠加). 主要因为 FAB-MAP 2.0 所处理的图像个数小于本文所处理的显著

区域个数, 因而基本 FAB-MAP 2.0 对闭环概率的计算较快; 但 FAB-MAP 2.0 在闭环几何验证环节需要对所有拥有高闭环概率的图像中的特征进行对极几何约束检验, 因而相对于本文算法其实时性较差.

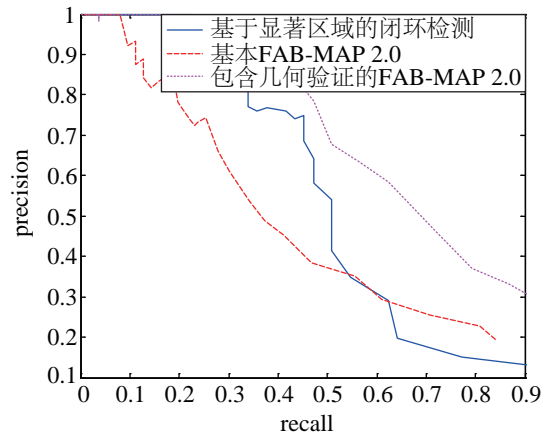


图 14 郊区公路数据集实验准确率—召回率对比

Fig.14 Contrast of precision-recall ratios in suburban dataset experiment

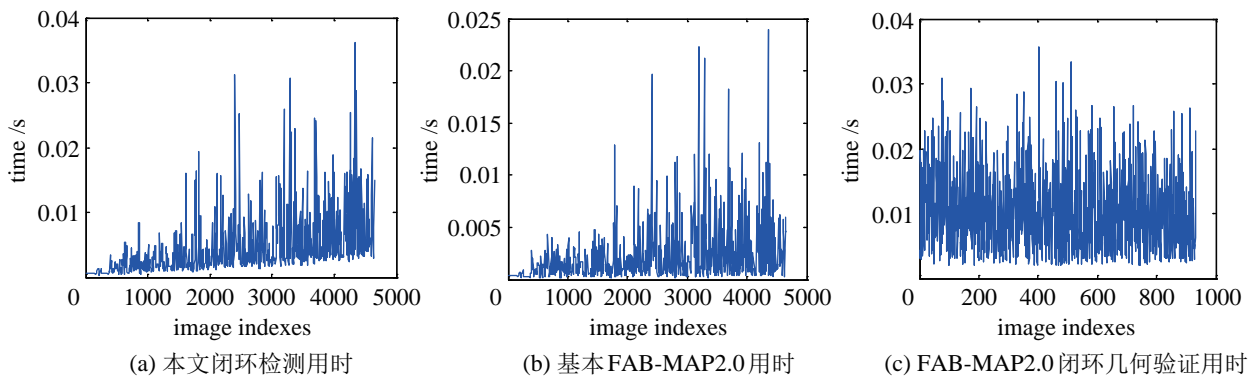


图 15 郊区公路数据集实验本文算法各环节用时

Fig.15 Time used for each procedure of the proposed algorithm in suburban dataset experiment

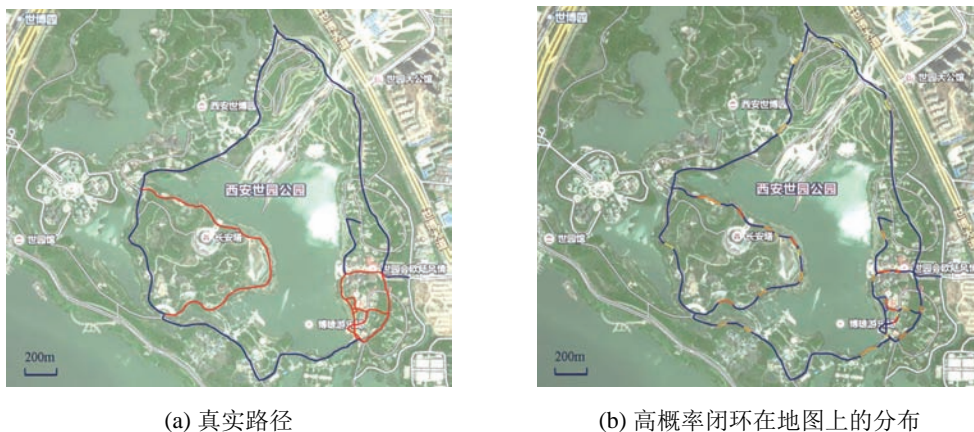
6.2 实际场景实验

本文选取西安世园会作为实际场景实验区域, 如图 16(a) 所示, 其中蓝色曲线为真实未闭环路径, 红色曲线为真实闭环路径. 由于实验环境的路况具有多样性, 因此采集数据时既采用了车载摄像头, 也在部分路段使用手持摄像头. 实验环境既包含城市街道场景, 也包含园林场景, 且部分不同地点的园林场景相似度较高 (见图 17). 总的来讲, 该实验环境对 VSLAM 闭环检测具有一定的挑战性.

本文算法获得的高概率闭环在地图上的分布如图 16(b) 所示, 图中蓝色轨迹表示未检测到高概率闭环的路段, 红色和黄色轨迹为闭环概率大于 80% 的路段, 且路段的闭环概率由黄到红逐渐增大, 其

最大概率为 1. 对比图 16(a)、(b) 可以看到, 在林木丰富的场景中存在着少数较高概率的误闭环, 这主要是由于所用视觉词典的规模较小, 不足以区分相似园林. 其次, 正确的高概率闭环大多集中于建筑密集区域, 这与该实验环境中的建筑可区分性大于园林场景可区分性的实际情况相一致.

本文的闭环检测算法对大视角变化闭环的检测效果如图 18 所示, 其中图 18(a) 为实时图像, 图 (b) 和 (c) 为发生闭环的历史图像. 图 (a) 中的左、右两个显著区域分别与图 (b) 右侧的显著区域和图 (c) 右侧的显著区域相匹配. 可以看到, 尽管观测视角和距离发生了较大的变化, 本文的闭环检测算法依然能够有效地检测出闭环的发生.



(a) 真实路径 (b) 高概率闭环在地图上的分布

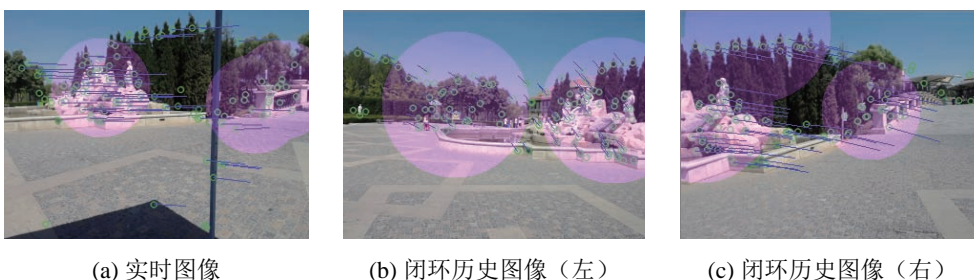
图 16 实际场景实验的真实路径与高概率闭环在地图上的分布

Fig.16 Distribution of high probability loop closures and real path in real scene experiment



图 17 不同地点园林的相似场景

Fig.17 Similar scenes at different locations



(a) 实时图像 (b) 闭环历史图像 (左) (c) 闭环历史图像 (右)

图 18 大视角变化闭环的检测结果

Fig.18 Detection results of loop closures with great view change

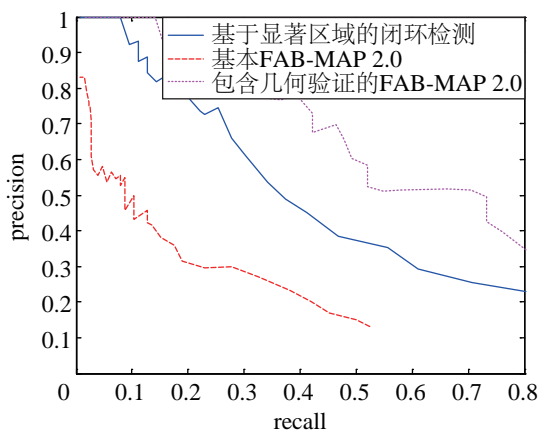


图 19 实际场景实验准确率-召回率对比

Fig.19 Contrast of precision-recall ratios in real scene experiment

实际场景实验的准确率-召回率对比曲线如图 19 所示. 图中蓝色、红色、紫色曲线分别对应本文基于显著区域的闭环检测算法、基本 FAB-MAP 2.0 与包含闭环几何验证的 FAB-MAP 2.0. 从图中可以看出, 在准确率-召回率方面, 本文闭环检测算法优于基本 FAB-MAP 2.0, 但与包含闭环几何验证的 FAB-MAP 2.0 存在相当差距. 本文算法在准确率-召回率方面优于基本 FAB-MAP 2.0 的原因有两个: 其一是基于显著区域的匹配对图像视角一致性的要求较低, 从而极大地提高了闭环检测的召回率; 其二是显著区域的观测几何匹配在一定程度上提高了闭环检测准确率. 相比于包含闭环几何验证的 FAB-MAP 2.0, 本文算法的差距主要由于 FAB-MAP 2.0 闭环几何验证环节剔除了大量的类似

于图 17 所示的歧义场景，通过降低其实时性，换取了相当的准确率。

图 20 反映了实际场景实验过程中本文算法的各环节用时情况，从图中可以看到，本文算法耗时

最长的环节是闭环概率的贝叶斯更新，这是由于为了有效地检测到闭环观测不连续的闭环场景，算法需要不断更新闭环先验概率和拓扑闭环概率，而两者的计算量随着拓扑节点数量呈线性增长。

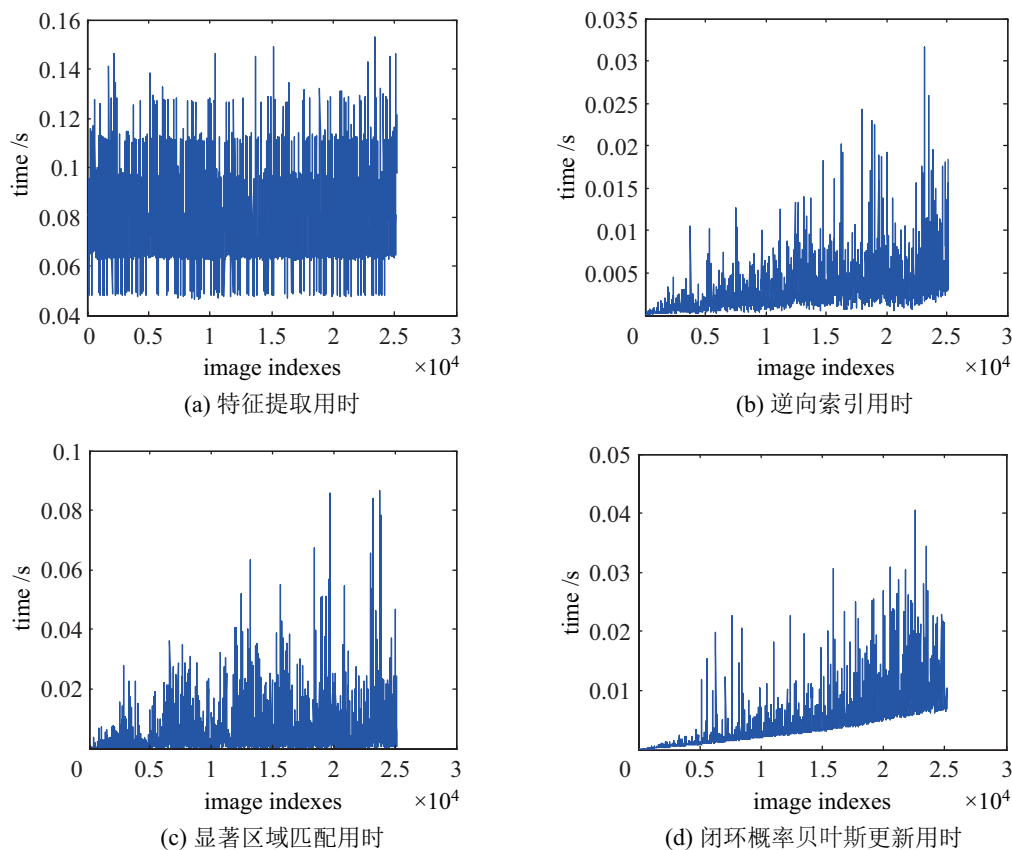


图 20 实际场景实验的各环节用时

Fig.20 Time used for each procedure of the proposed algorithm in real scene experiment

7 总结 (Conclusion)

本文提出了一种基于场景显著区域的闭环检测算法，旨在增强 VSLAM 闭环检测对路径不一致和大视角变化闭环的识别能力。首先设计了基于场景显著区域的闭环检测架构，构建了基于显著区域的闭环全概率模型，并提出了针对闭环观测不连续情形的闭环先验概率和拓扑闭环概率的计算方法。其次，介绍了本文的关键帧与显著区域选取方法，以及基于逆向索引的预匹配场景搜索策略。然后，针对显著区域的匹配问题分别讨论了基于 BoVW 单词的匹配概率和观测几何匹配概率的计算，并将两者结合，提高了闭环检测算法对不同场景的分辨能力。数据集实验和实际场景实验表明，本文闭环检测算法能够对路径不一致闭环和大视角变化的闭环进行有效检测，在保证较高准确率的条件下，提高了闭环检测的召回率。

下一步将主要在本文工作的基础之上，探索具

有更好的稳定性和复现性的显著区域选取策略，提高显著区域提取和匹配的鲁棒性，使算法对复杂环境具有更强的适应能力。

参考文献 (References)

- [1] Bazeille S, Filliat D. Combining odometry and visual loop-closure detection for consistent topo-metrical mapping[J]. *RAIRO-Operations Research*, 2010, 44(4): 365-377.
- [2] Ho K L, Newman P. Loop closure detection in SLAM by combining visual and spatial appearance[J]. *Robotics and Autonomous Systems*, 2006, 54(9): 740-749.
- [3] Williams B, Cummins M, Neira J, et al. A comparison of loop closing techniques in monocular SLAM[J]. *Robotics and Autonomous Systems*, 2009, 57(12): 1188-1197.
- [4] 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.
- [5] Cummins M, Newman P. Probabilistic appearance based navigation and loop closing[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2007: 2042-2048.

- [6] Cummins M, Newman P. Highly scalable appearance-only SLAM – FAB-MAP 2.0[C]//Robotics: Science and Systems. 2009. doi: 10.15607/RSS.2009.V.039.
- [7] Labbe M, Michaud F. Memory management for real-time appearance-based loop closure detection[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2011: 1271-1276.
- [8] Labbe 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.
- [9] Hou X D, Harel J, Koch C. Image signature: Highlighting sparse salient regions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 194-201.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [14] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231-1237.

作者简介:

李维鹏 (1992 –), 男, 硕士生. 研究领域: 移动机器人 VSLAM.

张国良 (1970 –), 男, 博士, 教授. 研究领域: 智能机器人, 先进控制理论与应用, 组合导航.

姚二亮 (1991 –), 男, 博士生. 研究领域: 移动机器人 SLAM.