

室内环境仿人机器人快速视觉定位算法*

吴俊君^{1,2}, 胡国生¹

(1. 广东食品药品职业学院软件学院, 广东 广州 510520;
2. 华南理工大学机械与汽车工程学院, 广东 广州 510641)

摘要: 视觉定位是移动机器人视觉导航的关键问题之一。定位的实时性对视觉导航的性能有较大的影响。在确保定位有效性前提下, 提出一种快速的视觉定位方法: 采用 BRISK (Binary Robust Invariant Scalable Key Points) 特征作为局部不变性特征点, 基于集合理论将场景图像简化为物种种群集合, 采用种群相似系数索雷申系数测量场景的相似性, 避免繁杂的计算过程, 力求简洁快速有效地完成视觉定位。仿真和仿人机器人实验结果表明: 在非结构化室内环境下, 视觉定位有效性达到 99%, 场景相似性测量的平均时间 0.03 s (每秒 33 幅图像), 验证了该方法的有效性、鲁棒性和良好的实时性。

关键词: 视觉定位; 视觉导航; BRISK 特征; 相似场景; 索雷申系数

中图分类号: TP242 文献标志码: A 文章编号: 0529-6579 (2013) 04-0007-07

An Algorithm for Fast Visual Location with the Humanoid Robot in Indoor Environment

WU Junjun^{1,2}, HU Guosheng¹

(1. Software College, Guangdong Food and Drug Vocational College, Guangzhou 510520, China;
2. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China)

Abstract: Visual location is one of the most important issues in robot visual navigation. The locating efficiency has a large impact on the performance of the visual navigation. A fast visual location method is proposed; BRISK is employed as the local invariant features, the scene model based on aggregate is described and the similarity of the scenes is measured by Sorensen coefficient to avoid complex computation for efficiency. The method has been verified by simulation and experiment with a humanoid robot in indoor environment. The effectiveness is 99% and the average measuring time is 0.03 s (33 fps). The result demonstrates that the method has good performance in terms of effectiveness, robustness and real time.

Key words: visual location; visual navigation; BRISK feature; similar scene; sorensen coefficient

机器人技术是当今自然科学研究领域最活跃的分支之一。经过几十年的研究, 机器人在功能、构型及性能等方面都取得了长足的发展, 并在社会生产中产生了巨大的经济效益。机器人的应用领域正在从标准化的工业生产线逐步渗透至社会生活的各

个服务领域, 这一发展趋势对机器人的自主移动能力提出较高的要求。视觉传感器能够获取丰富的场景特征(颜色、纹理等), 其对环境的刻画能力优于其他传感器, 因此移动机器人视觉导航成为当前研究的热点。视觉导航是从仿生学的角度, 给机器

* 收稿日期: 2013-03-26

基金项目: 国家自然科学基金资助项目(50975089); 高等学校博士学科点专项科研基金资助项目(200805611091)

作者简介: 吴俊君(1981年生), 男; 通讯作者: 胡国生; E-mail: 86228034@qq.com

人配备一双能感知自身周边环境的眼晴, 帮助机器人完成避障, 路径规划和环境地图构建等任务。场景图像中存在的一些具有局部不变性特征的特征点对旋转、平移、尺度均保持不变性, 这有利于帮助机器人解决自然环境下视觉导航中各个重要环节的问题, 比如: 闭环探测和视觉返巢等问题^[1-2]。视觉定位是机器人视觉导航中的关键问题之一, 通过计算机器人观测到的场景相似度完成机器人所在位置的空间相关性估计。视觉定位的效果与闭环探测密切相关, 它将直接影响到: 视觉 SLAM (simultaneous localization and mapping) 的成败^[3], 视觉返巢中移动方向的判断以及转角的实时计算和环境地图节点中关键帧的选取^[4]等方面。与 GPS 定位和基于内部传感器的航迹推演定位等精确定位方法相比, 视觉定位更关注机器人当前所处的区域范围而非具体的坐标值。因为对于大范围环境下的视觉拓扑地图而言, 区域定位比精确定位更有意义。

1 相关工作

目前视觉定位主要分为概率计算方法^[5-6]和图像匹配方法^[7-9]。概率计算方法将视觉定位归结为递归贝叶斯估计问题。首先采用图像建模方法描述机器人每一位置的场景图像, 估计已获取图像与对应位置的先验概率 A , 在当前时刻计算新场景图像与已访问位置匹配的后验概率 B , 若概率 B 大于预设阈值确定定位结果。图像匹配方法将视觉定位归结于序列图像匹配问题, 将当前时刻的图像与已获取的图像序列进行相似性匹配, 若相似度高于预设阈值, 则匹配图像被认为是对应了机器人所处的场景位置。

机器人在视觉导航过程中, 需要连续不断地定位, 即: 计算当前观测到的视觉信息与拓扑地图节点对应的关键帧的相似度, 但是较大的视觉信息量和不确定的环境因素制约了机器人视觉定位的实时性与鲁棒性。比如: 在视觉返巢中, 机器人依赖于密集的关键帧图像; 大范围内基于视觉的同时定位与地图构建 aVSLAM (Appearance Based Visual SLAM) 过程中, 机器人会生成信息量巨大的视觉地图节点。如何在保证定位效果的同时提高定位的实时性, 对移动机器人视觉导航的实际应用具有重要意义。

场景图像的相似度测量决定着视觉定位的效果。目前被用于视觉定位的场景相似度测量方法主要分为两类: ①采用图像整体特征的方法, 例如: 基于信息理论计算图像的信息熵确定图像之间的共

有信息量^[10]; ②采用局部不变性特征点的方法, 例如: 采用局部不变性特征点 (如: SIFT, PCA-SIFT, SURF 等) 作为图像中的稳定点^[11-13], 然后基于 BOW (Bag of Words) 模型对场景进行刻画, 再通过余弦距离计算模型的相似度^[14]。以上方法在有效性上的确取得了不错的效果, 但是计算过程繁琐 (比如: BOW 模型需要聚类过程和不断地更新视觉词汇字典), 实时性不佳, 实用性有待进一步提升。

本文研究工作在保证视觉定位有效性前提下, 提升定位的速度, 增强方法的实用性。方法创新点如下: ①采用 BRISK 局部特征作为视觉定位的特征点, 提高了场景相似度测量的实时性。②受物种群落相似性度量的启发, 将场景图像表征为由若干特征点组成的种群集合, 简化了图像场景模型, 提高了计算效率。③采用种群的索雷申相似系数, 度量场景图像的相似性, 计算过程简洁快速有效。

2 仿人机器人视觉定位算法设计

2.1 BRISK 特征点探测与描述

本文算法采用 BRISK 特征^[15]作为局部不变性特征点, 与 SIFT, PCA-SIFT, SURF 局部特征点相比, BRISK 采用二进制特征向量描述特征点, 利用汉明距离度量特征点的相似性, 匹配速度远高于其他特征点。

算法首先探测并描述 BRISK 特征, 具体过程主要包括 3 个步骤: ①基于尺度空间理论构建图像金字塔; ②特征点探测。在尺度空间探测局部极值点作为特征点, 如图 1 所示。探测结果如图 2 所示: 极值点所处的尺度 k , 位置 (x, y) 。③采用二进制向量描述特征点。

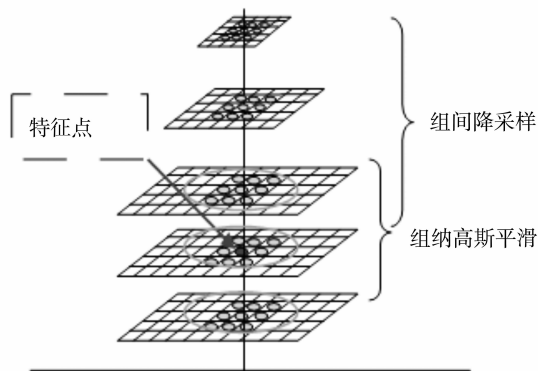


图 1 图像金字塔: 探测特征点

Fig. 1 Image pyramid: point detection



图 2 探测结果
Fig. 2 Detection result

特征点描述：用特征点 $p \{ (x, y), k \}$ 邻域的 60 个特征点之间的灰度大小关系来表征，生成二进制的特征向量^[15]。特征向量的相似性采用汉明距离来度量，如图 3 所示。

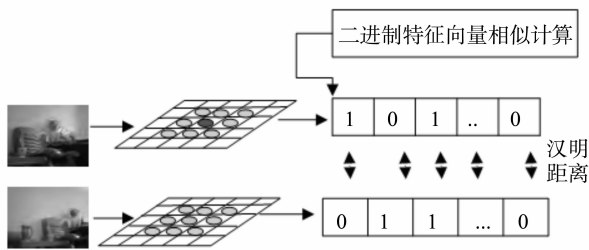


图 3 二进制描述符匹配
Fig. 3 Binary descriptor match

2.2 场景建模与相似性测量

由于不同的图像自匹配峰值差别较大，因此两幅场景图像的相似性无法直接通过两者包含的特征点匹配数目来度量（如图 6 所示：峰值为图像之间的特征点匹配数）。因此，需要利用这些离散的特征点来合理地描述一幅场景，即：场景建模。具体描述如下：

假设一幅图像 I 包含若干个 BRISK 特征点，任意一个特征点 P_i 及其描述符 P_d 如公式 (1)，公式 (2) 所示

$$P_i = \{x, y, k, \varphi\} \quad (1)$$

其中 x, y, k, φ 分别表示特征点的坐标，尺度和主方向。

$$P_d = [b_i] \quad (2)$$

其中 b_i 是当前特征点描述符的二进制比特位。

则该场景图像 I 的描述，如公式 (3) 所示

$$I = \{P_d^{[x]} \mid x = n\} \quad (3)$$

其中 n 是 BRISK 特征点数目。

假设图像 I 里的 BRISK 特征点被看成若干类不同的物种，则一幅场景图像可等价为一个物种群落，如公式 (4) 所示

$$Q = \{w^{[x]} \mid w^{[x]} = P_d^{[x]}, x = m\} \quad (4)$$

其中 m 是群落 Q 的物种种类数量。

生物群落的相似性计算本质上是基于集合理论。无论是一个群落 Q 还是一幅包含若干特征点的图像 I ，能否被视为一个集合，必须满足“逻辑性”这个充分必要条件，即：每一个对象都能确定是不是某一个集合的元素。将图像的特征点进行自匹配可知（如图 4 所示）：不仅可以确定一个特征点是否属于某幅图像，而且图像特征点之间也满足互异性（即：特征点一一对应）属性条件，因此图像被简化为一个特征点集合是合理的。如图 5 所示：图像的相似性度量转化为求两个种群集合的相似性。

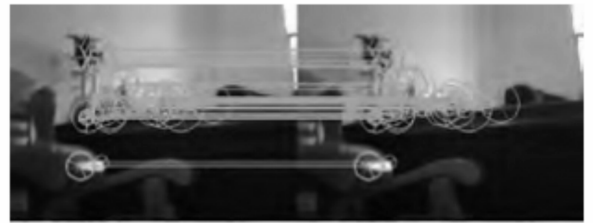


图 4 自匹配
Fig. 4 Self-match

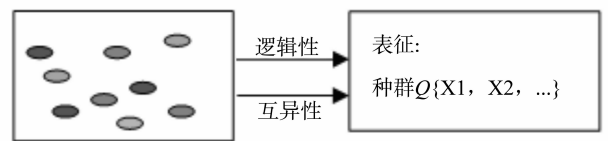


图 5 场景模型
Fig. 5 Scene model

在生物学领域，为了研究生物种群群落的生长状况受环境因素的影响，需要计算两个种群的相似性。种群里的物种总数和共有种数的综合性指标（即：群落相似性系数）被用来表征种群组成的相似度，相似系数包括：索雷申系数，杰卡特系数和芒福德系数。本文通过计算两个种群的相似性来度量两幅场景图像的相似性，实验采用的相似系数为索雷申系数 SC ，如公式 (5)，两幅图像的相似距离 DSC ，如公式 (6) 所示

$$SC(Q_i, Q_j) = \frac{2 * Q_i \cap Q_j}{Q_i \cup Q_j} \quad (5)$$

$$DSC(Q_i, Q_j) = \frac{Q_i \cup Q_j - 2 * Q_i \cap Q_j}{Q_i \cup Q_j} \quad (6)$$

其中 Q_i, Q_j 为群落, 且 $SC(Q_i, Q_j) \in [0, 1]$ 和 $DSC(Q_i, Q_j) \in [0, 1]$ 。

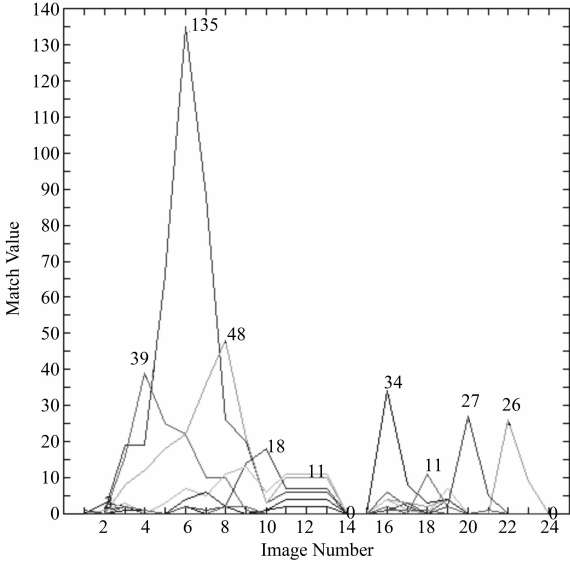


图 6 图像匹配结果
Fig. 6 Match result

视觉定位算法设计, 如图 7 所示。算法的时间复杂性分析如下: 整个算法的执行主要包括四个方面: BRISK 特征点的探测, BRISK 二进制描述符的生成, 描述符的匹配和图像相似度计算。其中特征点描述符的匹配环节是影响算法时间复杂度的关键因素。假设当前待比较的两幅图像包括的特征点数分别为: $N, M (N < M)$, 那么匹配环节的双重循环次数为 $N * M$, 则算法的时间复杂度为 $O(N * M)$ 。

3 仿真实验

3.1 场景相似阈值

实验采用室内随机捕获的一组场景图像作为数据集 F , 数据集 F 包含 1 002 幅 $640 * 480$ 大小介于 $60 \sim 200$ K 之间的图像。数据集 F 被分为四类场景图像: 同一场景集合 A, 相似场景集合 B, 非相似场景集合 C, 完全不同的场景集合 D。采用本文的算法对数据集里的图像做成对的相似性计算: 以 10 幅图像为例, 部分测量结果分别如图 8 和图 9 所示, 每条曲线反映了当前某个场景与其他场景的相似度。

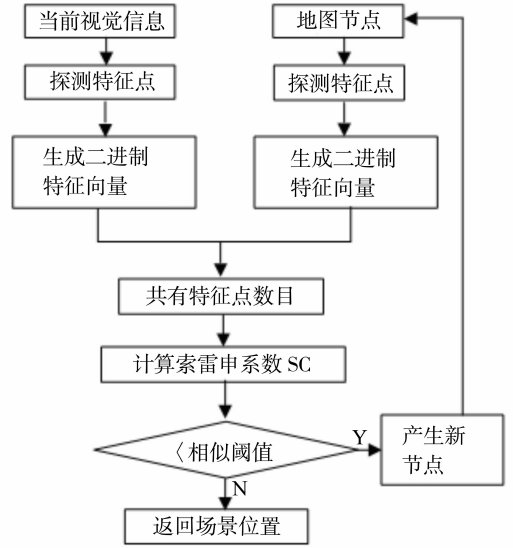


图 7 算法流程图

Fig. 7 Algorithm flow chart

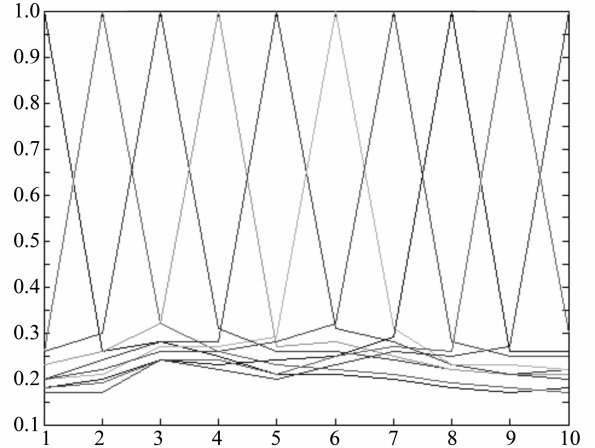


图 8 相似场景: SC 系数

Fig. 8 Similar scene: SC coefficient

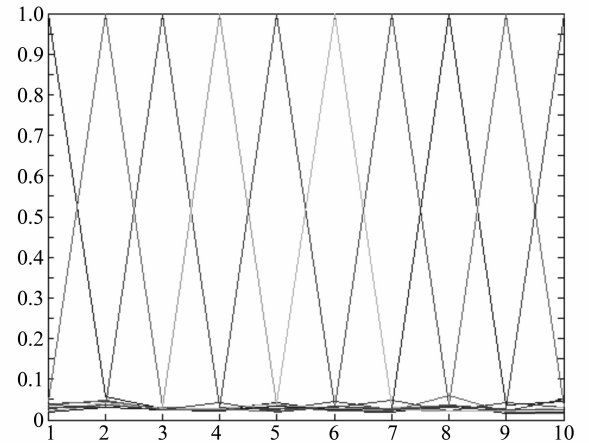


图 9 不同场景: SC 系数

Fig. 9 Different scene: SC coefficient

由仿真实验可知：相同或相似场景的 SC 系数均处于 0.15 以上，故算法将相似阈值设为 0.15。为了便于计算，在实际的验证实验中将 SC 系数放大 100 倍，相似阈值设为 15。

3.2 算法验证

本小节对算法的有效性、鲁棒性和实时性，与现有方法进行对比验证。

有效性：在数据集 F 包含的场景区域范围内，随机拍摄一组图像 C 作为机器人当前的位置。根据设定的阈值 15，从实验数据集估算场景位置，成功率为 99%。各个场景集合图像的相似性测量结果如表 1。

表 1 算法的有效性

Table 1 The effectiveness of algorithm

场景集合	相似系数	系数放大 100 倍
相同场景 A	1	100
相似场景 B	0.15 ~ 0.95	15 ~ 95
非相似场景 C	0.025 ~ 0.062	2.5 ~ 6.2
完全不同场景 D	0 ~ 0.014	0 ~ 1.4

此外，实验将本文方法 FVL (Fast Visual Location) 与其他场景相似性测量方法 (SIFT-BoW^[14], SIFT-BoRF^[16]) 的定位效果进行了对比。实验的具体方法如下：将机器人连续观测到的 40 幅场景图像作为测试数据集 T ，随后任意选取一幅场景图像 n 作为机器人在当前位置的观测值，然后分别采用 FVL, SIFT-BoW 和 SIFT-BoRF 方法对机器人的实际位置概率分布进行估计。以图像 $n=15$ 为例：三种方法分别计算了图像 15 与数据集 T 中其他图像的空间相似度，然后进一步将三组相似系数分别归一化到 $[0.4, 0.6]$, $[0.6, 0.8]$, $[0.8, 1.0]$ 三个区间，三条曲线分别反映了机器人当前所处位置的概率分布，曲线的顶点峰值对应的场景作为视觉定位结果。如图 10 所示：三条曲线中，在图像 15 附近的场景被估算为实际位置的概率明显高于其他位置。这表明：在视觉定位效果上，三种方法均准确有效地完成视觉定位，而且本文方法 FVL 和 SIFT-BoRF 的效果相当，SIFT-BoW 方法次之。

鲁棒性：为了验证方法的鲁棒性，实验进一步将相似场景细分为在不同角度、不同距离、不同光照三种情况。算法分别对其进行了相似性测量。如表 2：同一场景在左右平移 2 m，前后直行 1.5 m，白天室内环境开灯和关灯情况下，计算得到的相似系数均在阈值 15 以上，算法具有良好的适应性。

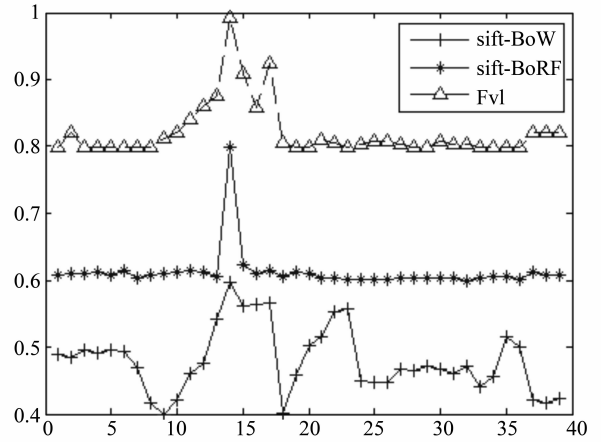


图 10 三种方法定位效果对比

Fig. 10 The locating efficiency of three methods

表 2 算法的鲁棒性

Table 2 The robustness of algorithm

相同场景	测试方法	相似系数
不同距离	平移 2 m	17.5 ~ 26.3
不同角度	直行 1.5 m	20.5 ~ 32.0
不同光照	白天灯光 (开和关)	16.2 ~ 40.2

实时性：从数据集 F 随机取一个样本依次与场景图像序列 (500 幅) 进行相似计算，记录下该样本与图像序列相似性测量所耗费的时间。反复取若干样本重复这个计算过程，取时间的平均值。如图 11 所示：每次计算的时间基本处于 0.05 s 以下，平均耗时约为 0.03 s。计算时间主要受制于两个因素：特征提取的时间耗费，场景模型的复杂度。实验将本文的方法 FVL 与 SIFT-BoW^[14], SIFT-BoRF^[16] 在计算时间上做了对比分析，结果发现：SIFT-BoW 方法提取 SIFT 特征后，需要进一步聚类分析构建单词本，故相似计算的总耗费时间远大于 SIFT 特征提取的时间；SIFT-BoRF 方法测量时间与 SIFT 特征提取时间相近，如表 3 所示。该方法在计算速度上要明显高于另外两种方法，耗费时间仅为前两种方法的 1%。

表 3 计算时间对比

Table 3 The comparison of computing time

测量方法	SIFT-BoW	SIFT-BoRF	BRISK-FVL
时间范围/s	≥ 2.95	≈ 2.95	≈ 0.03

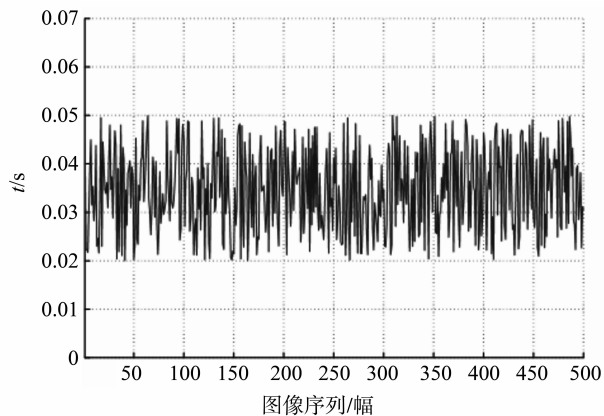


图 11 时间耗费

Fig. 11 Time cost

4 实验验证

本实验采用法国 Aldebaran Robotics 公司研发的小型仿人机器人 NAO 作为研究平台, 该机器人具有典型的双足机器人行走特点。机器人配有两个上下排列的针孔摄像机, 上方的摄像机视野向前观测周围环境, 下方的摄像机视野向斜下方观测地面。

机器人 NAO 利用前向摄像机在办公室环境中构建拓扑地图: 机器人的移动轨迹由直线速度 v 、转角 α 和转角速度 ω 决定。实验中机器人的直线速度设为固定, 转角 $\alpha = 0.3$ 。每隔 2 s 取一幅关键帧, 在室内构建 56 个拓扑节点, 如图 12 和图 13 所示。每个地图节点用一幅场景图像进行刻画。机器人在室内漫游, 实时进行视觉定位: 机器人首先计算出所有与自己位置场景相似系数大于阈值 15 的地图节点, 然后选出最大的相似节点 MN (Max - similar node), 此节点 MN 被认为是机器人



图 12 视觉定位场景

Fig. 12 Visual location scene

当前所处的位置, 定位结果如图 14 所示。若相似系数均小于 15 则认为机器人处在一个新位置, 此场景被作为一个新的节点加入到环境地图中。

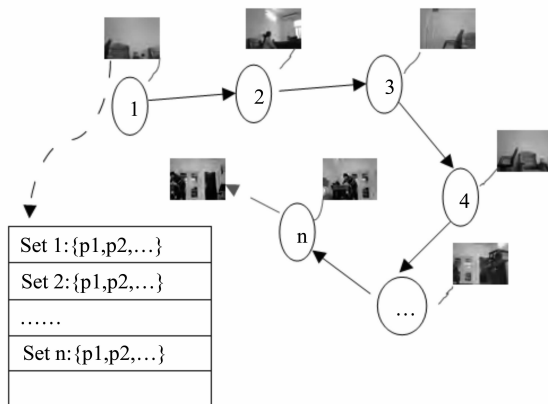


图 13 室内拓扑地图构建

Fig. 13 Topological map in indoor environment



图 14 定位结果

Fig. 14 Location result

视觉定位分析: 机器人在室内漫游, 利用本文算法进行视觉定位, 对于同一场景从不同视觉角度的观测值, 均被有效地测量出来, 如图 14 所示。整个定位过程中对于已经到过的拓扑地图节点位置识别率达到 99%, 每对图像的相似系数平均计算时间: 0.03 s, 每秒可处理 33 帧图像。此外, 实验过程也发现如下现象: 个别死角或非常昏暗的位置由于无法提取到有效的特征点而导致相似度测量

失败,如图15所示。此类“迷失问题”需要进一步研究有效的视觉地图节点生成方法,比如:对无效的关键帧进行剔除,或将模糊的图像清晰化。该问题将在下一步研究工作中展开。



图15 定位迷失(黑暗视角或模糊图像)

Fig. 15 Lost location (dark perspective or blurred image)

5 结论

仿人机器人主要以室内工作环境为主,视觉定位方法能快速确定机器人所处的场景区域,对机器人进行视觉导航及完成后续任务具有重要的现实意义。非结构化室内环境存在较多的不确定因素影响视觉定位效果,尤其是定位的实时性这一重要指标有待提升。本文受生物群落相似度测量的启发,提出一种快速的视觉定位算法,并利用仿真实验和仿人机器人平台对算法的实时性、有效性和鲁棒性进行了验证,结果表明:和现有的方法相比,本文方法FVL不仅取得了良好定位效果,而且实时性具有较大的提升。一次场景相似性计算平均耗时0.03 s (33 fps),定位有效性为99%,对具有一定差异的相似场景,相似度计算具有良好的鲁棒性。算法的整个计算过程比较简洁实用,对提升视觉导航性能,尤其是实时性方面具有积极意义。

后期的研究工作将继续致力于提高室内环境下仿人机器人视觉导航的性能,并专注细节问题的研究:例如考虑如何选取有效的拓扑地图节点,研究动态环境下视觉定位的鲁棒性和大范围环境中移动机器人视觉闭环探测的实时性等问题。

参考文献:

- [1] JAUREGI E, IRIGOIEN I, SIERRA B, et al. Loop-closing: a typicality approach [J]. *Robotics and Autonomous Systems*, 2010, 59(3): 218 - 227.
- [2] MOLLER R, KRZYKAWSKI M, GERSTMAYR L. Three 2D-warping schemes for visual robot navigation [J]. *Autonomous Robot*, 2010, 29(11): 253 - 291.
- [3] MORIOKA H, HASEGAWA O. Vision-based mobile robot's SLAM and navigation in crowded environments [C]//*IEEE International Conference on Intelligent Robots and Systems*, San Francisco, CA, 2011: 3998 - 4005.
- [4] ZHANG H, LI B, YANG D. Key frame detection for appearance-based visual slam [C]//*IEEE International Conference on Intelligent Robots and Systems*, Taipei, 2010: 2071 - 2076.
- [5] RANGANATHAN A, DELLAERT F. Online probabilistic topological mapping [J]. *International Journal of Robotics Research*, 2011, 30(6): 755 - 771.
- [6] CUMMINS M, NEWMAN P. Probabilistic appearance based navigation and loop closing [C]//*IEEE International Conference on Robotics and Automation*, Rome, 2007: 2042 - 2048.
- [7] 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.
- [8] CALLMER J, GRANSTROM K, NIETO J, et al. Tree of words for visual loop closure detection in Urban SLAM [C]//*the Australasian Conference on Robotics and Automation*, 2008: 1 - 8.
- [9] HO K L, NEWMAN P. Detecting Loop closure with scene sequences [J]. *International Journal of Computer Vision*, 2007, 74(3): 261 - 286.
- [10] DAME A, MARCHAND E. A new information theoretic approach for appearance-based navigation of non-holonomic vehicle [C]//*IEEE International Conference on Robotics and Automation*, Shanghai, 2011: 2459 - 2464.
- [11] LOWE D G. Distinctive image features from scale-invariant key points [J]. *International Journal of Computer Vision*, 2004, 60(2): 91 - 119.
- [12] BAY H, ESS A, TUYTELAARS T, et al. Surf: speeded up robust features [C]//*Europe Conference on Computer Vision*, 2006: 404 - 417.
- [13] KE Y, SUKTHANKAR R. PCA-SIFT: a more distinctive representation for local image descriptors [C]//*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004: II - 506.
- [14] FILLIAT D. A visual bag of words method for interactive qualitative localization and mapping [C]//*IEEE International Conference on Robotics and Automation*, Rome, 2007: 3921 - 3926.
- [15] LEUTE NEGGER S, CHLI M, SIEGWART R Y. BRISK: binary robust invariant scalable key points [C]//*IEEE International Conference on Computer Vision*, Barcelona, 2011: 2548 - 2555.
- [16] HONG Z. BoRF: loop-closure detection with scale invariant visual features [C]//*IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011: 3125 - 3130.