

基于局部显著区域的自然场景识别

王璐¹⁾ 陆筱霞¹⁾ 蔡自兴²⁾

¹⁾(中原工学院计算机学院, 郑州 450007) ²⁾(中南大学信息科学与工程学院智能所, 长沙 410083)

摘要 场景识别是移动机器人实现拓扑导航的关键。针对未知环境,提出一种基于视觉局部显著区域的自然场景识别方法。首先,提出带反馈的显著性检测模型(FSDM)自底向上进行图像分析;然后,根据显著位置,基于分形实现自动尺度选择,以构造合适尺寸的局部显著区域。对场景图像中的显著区域采用梯度方向、二阶不变矩、归一化色调3种特征进行不变性表示,并根据其匹配率实现场景识别。实验结果表明,FSDM具有较高的显著性检测精度。而且室内室外环境的多次场景识别实验也表明,该方法与全局外观方法相比能够更好地容忍尺度、视角等变化引起的差异,静态场景识别具有较高的准确性。

关键词 场景识别 局部显著区域 带反馈的显著性检测模型 尺度选择 分形维数

中图法分类号:TP391.41 文献标识码:A 文章编号:1006-8961(2008)08-1594-07

Local Salient Regions Based Natural Scene Recognition

WANG Lu¹⁾, LU Xiao-xia¹⁾, CAI Zi-xing²⁾

¹⁾(School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007)

²⁾(Center of Intelligent System and Software, School of Information Science and Engineering, Central South University, Changsha 410083)

Abstract Scene recognition is a key problem in mobile robot topological navigation. For unknown environments, a natural scene recognition approach based on visual local salient regions is presented. Firstly, a feedback saliency detection model (FSDM) is presented to carry out bottom-up scene image analysis. Then, according to the salient positions, automatic scale selection is realized based on fractal dimension to build the local salient regions with appropriate size. Those salient regions are represented by 3 invariant features of gradient orientation, moment and canonical hue. They are used for scene recognition in terms of their match ratio. Experiments show that FSDM can obtain higher accuracy. The scene recognition experiments in both indoor and outdoor environments show that the approach has high stability and tolerance compared to the method based on global appearance when scale and viewpoint etc changed. The accuracy of recognition for static scene is higher.

Keywords scene recognition, local salient regions, FSDM (feedback saliency detection model), scale selection, fractal dimension

1 引言

移动机器人在未知环境(没有或缺乏先验知识)中导航时,必须通过所携带的传感器感知环境,在探索过程中逐步建立环境模型和精确定位。为实现视觉拓扑定位,机器人需要对当前所在位置(拓

扑节点)获得的场景图像进行分析,并在拓扑地图中寻找最相似的匹配,即场景识别。目前在未知环境中利用视觉图像实现场景识别主要有两类可行方法:基于全局图像外观特征和基于兴趣点等局部图像特征。

Iwan Ulrich等利用全景摄像头感知环境,实现了一个基于图像直方图的场景识别系统,所采用的

基金项目:国家自然科学基金重点项目(60234030)

收稿日期:2008-04-10;改回日期:2008-05-05

第一作者简介:王璐(1972~),男,2007年获中南大学计算机应用技术博士学位,现为中原工学院计算机学院副教授。主要研究方向为人工智能、机器视觉、并行计算等。E-mail:dqx_wl@126.com

外观特征包括 HLS 以及 RGB 共 6 张颜色直方图^[1]。Zhou Chao 使用普通 CCD 摄像头感知环境,采用多维直方图描述整幅图像的外观特征^[2]。此类基于外观特征的方法直接利用整幅图像信息,不需要进一步的图像分析,对于计算机而言容易实现。但在移动机器人导航问题中,由于机器人的移动,摄像机捕获的平面图像很容易发生变化,使得这种针对整幅图像的外观特征不够稳定,难以满足远近尺度、视角等发生变化的场景识别任务。而利用局部图像特征进行匹配与识别,对于遮挡以及环境变化等都具有较好的鲁棒性,兴趣点是这类特征中的典型代表^[3]。在移动机器人导航时可以将检测到的局部图像特征作为路标来代表环境(拓扑法)或实现几何定位(几何法)。尤其在未知环境图像分割实施困难的情况下,这种完全像素级的自底向上的分析策略有着广泛的应用前景。

目前有很多兴趣点检测方法,例如各种角点、拐点检测器。兴趣点检测的难点主要体现在稳定性方面,即在视觉条件发生变化情况下,所检测的兴趣点仍然具有远近尺度、旋转、平移、光照、遮挡、视角等不变性。

角点检测所得到的兴趣点最大的缺陷是不具有远近尺度不变的特性。Schmid 等在对已有兴趣点检测器综合评价的基础上,提出了改进的 Harris 角点检测器。该方法在多尺度空间上计算 Harris 兴趣点并只选择在各尺度上具有局部最大值的角点^[3]。通过对角点周围区域的像素点进行特征化描述及仿射变换估计,使所检测的兴趣点具有尺度、旋转、平移及仿射不变的特性,并应用于目标识别及跟踪任务中。还有一类基于视觉注意机制的兴趣点也称为显著点,主要包括两大类:一是基于局部特征的算法,如边缘密度、对称性等;另一类是基于视觉反差的算法,是当前的主要研究方向,其中又以 Lowe 的 SIFT 方法和 Itti 等的算法最具代表性^[4,5]。David Lowe 提出的 SIFT 方法在多尺度空间上计算高斯差(DOG)滤波,并根据其最大最小值决定兴趣点的位置和所在尺度。这类 DOG 方法的缺点是在直线或轮廓的相邻位置也可能检测出局部最大或最小,即存在虚假的兴趣点^[3]。而且 SIFT 方法可以从一幅图像中检测出上千个兴趣点,不利于应用于实时要求的移动机器人导航问题。Itti 等提出采用中心-周围差算子在高斯锥的多尺度空间上计算颜色、亮度、方向对比的显著点检测方法,在针对彩色图像的检

测效果及运算速度上都有突出表现^[5,6]。本文对 Itti 算法的改进及实践表明,所检测的显著点位置具有优良的尺度、旋转、平移和视角不变性。

鉴于局部图像特征在机器人移动过程中表现出的良好稳定性,本文提出一种利用显著的局部图像区域实现自然场景识别的方法,能够在机器人运动引起的远近尺度、旋转、视角变化等情况下保持显著区域的重复检测能力和场景的有效识别。

2 带反馈的显著性检测模型

受注意力选择机制以及 Itti 等人工作的启发,提出了一个基于视觉显著性的自然路标提取方法^[7]。该方法首先利用保细节采样策略建立高斯锥的 4 级多尺度图像空间,然后采用中心-周围差方法在该空间中计算颜色与纹理的特征对比。综合所有对比子图,最终产生一幅显著性指示图。该方法所检测的显著位置具有较好的尺度、旋转、视角、平移不变性,详细设计及稳定性测试参见文献^[7]。尽管该方法采取了如保细节采样、根据傅里叶功率谱峰值确定 Gabor 滤波器参数等措施,并取得了较为满意的检测结果,但在实践中还是发现,该方法中所采用的“归一化”子图综合策略可能造成某些显著位置被忽略,稳定性不如 Itti 提出的“迭代”策略。但“迭代”策略也存在一些问题,例如迭代次数如何确定等。此外,显著点检测方法本身也存在一些固有问题,例如:如何判定显著位置识别的好坏,如何判定子图综合对识别结果的作用等。针对这样一些不足,本文在前述工作的基础上提出一种新的改进方法:带反馈机制的显著性检测模型(feedback saliency detection model, FSDM)。

在子图综合阶段,为克服迭代次数不可控等问题,引入了一条反馈通道。该反馈不仅可以控制迭代次数,而且可以根据特征综合结果调节各种类型特征对综合的影响,以控制各类特征在不同场景图像处理中发挥的作用。

首先对所有子图进行一次 DOG 滤波,并根据不同特征分别进行综合,形成颜色和纹理特征各自的显著性子图 I_c 和 I_t , 如下述公式所示。其中, $RG(c, s)$ 、 $BY(c, s)$ 分别为中心级和周围级红/绿对比子图和蓝/黄对比子图。 $T(c, s, \theta)$ 为纹理对比子图。 c 代表中心级(scale 1、2), s 代表周围级(scale 3、4)。

$$I_C = \bigoplus_{c=1}^2 \bigoplus_{s=3}^4 [DOG(\mathbf{RG}(c,s)) + DOG(\mathbf{BY}(c,s))]$$

$$I_T = \sum_{\theta \in \{10^\circ, 45^\circ, 90^\circ, 135^\circ\}} \bigoplus_{c=1}^2 \bigoplus_{s=3}^4 [DOG(\mathbf{T}(c,s,\theta))]$$

计算 I_C 和 I_T 各自的最大值 $\max(I_{fi})$ 与均值 $\text{MEAN}(I_{fi})$, 用以度量每类特征的贡献, 即权值 w_{fi} 。这里, $fi = \{C, T\}$, 分别表示颜色和纹理特征。

$$w_{fi} = (\max(I_{fi}) - \text{mean}(I_{fi}))^2$$

综合 I_C 和 I_T , 形成显著性指示图 \mathbf{SM} 。

$$\mathbf{SM} = w_C I_C \oplus w_T I_T$$

上述公式中, “ \oplus ”代表逐像素求和。由于判断显著位置是否准确是个很难的问题, 首先什么是显著的很难定义, 另外也没有很好的定量计算方法。为此, 这里采用显著位置的稳定性作为评判显著性检测的度量, 即若本次迭代得到的平均显著位置与上次迭代得到的平均位置之间的偏差 δ_{loc} 在允许范围内 ($\leq R_\delta$), 认为显著位置是可信的; 否则, 返回第 1 步再次进行 DOG 滤波重新迭代。每幅显著性指示图中的平均显著位置计算如下:

$$\overline{loc} = \sum_{i=1}^N \sqrt{(x_i^2 + y_i^2)}$$

式中, N 代表该显著性指示图中显著位置的总数, x_i, y_i 分别表示第 i 个显著位置的 x, y 坐标。

FSDM 很好地兼顾了迭代次数与显著性检测效果之间的有机结合, 在实践中取得了令人满意的检测结果。

3 显著区域构造

确定显著位置后, 一种方法是将所有显著位置即兴趣点进行不变性表示后输出, 可能会产生数千个特征点。该方法对于目标识别发挥较好的作用, 但用于代表场景则过于繁杂。因此, 本文根据显著位置通过尺度自动选择技术确定显著区域, 从而合并若干显著点, 有效减少了参与识别的局部图像特征数量。

由于远近尺度、视角等环境变化, 显著区域的尺寸不可能固定。自动选择尺度可以采用两种方法, 一种是根据不同尺寸显著区域与周围邻域的特征差异来设计自适应选择显著区域大小的算法^[8]。这种方法的物理意义明确, 但计算量较大; 另一种方法是将尺度发生变化时局部图像中保持不变的特征作为显著区域大小选择的目标。实现中需要基于某些

能够代表局部图像结构的特征设计评价函数, 并根据函数随区域尺寸变化出现的极值选择合适的区域大小^[3]。此方法由于不涉及与尺寸外区域的比较, 因此方法简洁, 计算量较小。

Schmid 等运用 LOG 算子进行自动尺度选择, 并得出结论: 采用 LOG 函数最有可能得到正确的特征化尺度。但该算子的尺度指的是高斯尺度, 并非区域大小。尽管给出了一个通用的与实际区域尺度的对应关系, 即高斯尺度的 3 倍, 但实际上这种选择往往与人类观察不相符合。

分形学是以局部和整体具有自相似的复杂事物为研究对象探索其复杂性的科学。分形维数越大, 客体就越复杂越粗糙, 反之亦然。基于此特点, 用分形维数作为区分不同类别纹理的参数^[9]。在尺度选择时, 若不同尺度图像区域中包含了相似物体, 虽然尺度不断放大但其纹理特征相似, 则分形维数基本保持在很小的变化区间。但是, 若区域中包含了新的物体或背景, 则纹理将更加复杂, 而分形维数也将发生较大的变化。因此, 本文选择分形维数发生变化时所对应的区域尺度来建立合适的显著区域。

大多数图像的分形维数计算都是基于灰度开展的, 这源于 Pentland 假设^[9]。但实际上, 梯度方向特征具有更好的尺度、视角不变性。因此, 本文中首先计算图像中每个像素点的梯度方向, 然后再采用微分计盒维数法计算该尺度图像的分形维数^[10]。具体计算过程如下:

(1) 计算每像素的梯度方向特征^[4], 其中 A 代表像素的灰度值。

$$GO_{i,j} = \arctan\left(\frac{A_{i,j+1} - A_{i,j}}{A_{i,j} - A_{i+1,j}}\right)$$

(2) 针对 $M \times N$ 的 2 维图像, 结合梯度方向等级 (360 级) 作为 Z 轴, 形成一个 3 维空间。进一步, 将该空间划分为若干规格长、宽、高为 $\frac{Me}{N} \times e \times \frac{360e}{N}$ 的长方形小盒。假设 $N \leq 360 \leq M$, e 取 4。

(3) 若图像的第 (i,j) 个小格内所包括的梯度方向最大等级和最小等级分别落在该小格上面编号为 h_a 和 h_b 的小盒内部, 则 $k_e(i,j) = h_a - h_b + 1$, 其中, $k_e(i,j)$ 表示第 (i,j) 个小格对 K_e 的贡献量。在此基础上, 计算 $K_e = \sum_{i,j} k_e(i,j)$ 。

(4) 计算微分计盒维数:

$$D = \frac{\log K_e}{\log\left(\frac{1}{e}\right)}$$

实践表明,根据梯度方向特征计算的分形维数在图像尺度发生变化但包含相似物体的情况下表现更加稳定。当包含新物体或背景时,该分形维数出

现比较明显的跃变。不同尺度下分形维数的变化情况如图 1 所示,可根据出现跃变的位置选择合适的图像区域尺度。

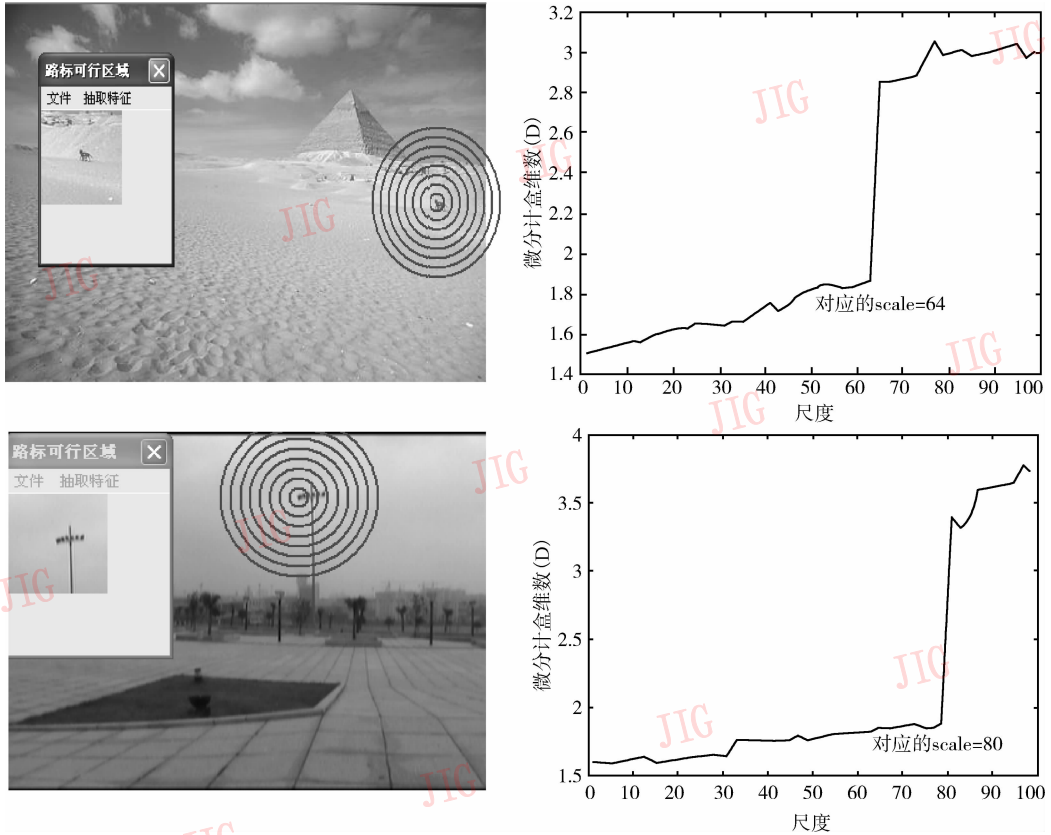


图 1 基于分形维的路标尺度选择

Fig. 1 Scale selection of landmark based on fractal dimension

4 基于显著区域的场景识别

结合外观特征对所检测的显著区域进行不变性表示。选择类内方差较小且类间距离较大的特征表示图像,有利于更好地进行分类识别。因此,对在许多文献中经常使用并受到好评的的图像特征(包括直方图特征、梯度、梯度方向、不变矩、归一化的色调值、RANK 等)分别针对 Berkeley 图像库、室内环境及室外广场环境图像进行了大量类内及类间测试。最终,作者选择梯度方向、二阶不变矩、归一化的色调作为表示方案,即 $LMK(GO, \phi_1, SH)$, 定义如下。

(1) 梯度 G 和梯度方向 $GO^{[4]}$:

$$G_{i,j} = \sqrt{(A_{i,j} - A_{i+1,j})^2 + (A_{i,j} - A_{i,j+1})^2}$$

$$GO_{i,j} = \arctan\left(\frac{A_{i,j+1} - A_{i,j}}{A_{i,j} - A_{i+1,j}}\right)$$

(2) 区域特征-不变矩: $\phi_1 = \eta_{20} + \eta_{02}$

(3) 色彩特征-归一化的色调值: $SH = \frac{H_{max}}{H_{min}}$

这 3 个特征总计只有 12 个字节,因此,存储需求较小。视觉皮层中神经元的响应特性表明,视角、尺度甚至图像拉伸变化等对局部图像的方向和空间频率影响较小^[4]。因此,梯度方向特征 GO 对于尺度、视角、旋转、光照等都具有很好的不变性,而归一化的色调 SH 则对于旋转、仿射等具有较好的不变性。

在从自然环境图像中检测出显著路标后,环境识别问题可以描述为:将待识别环境图像中检测到的路标与环境图像库的图像中检测出的路标进行匹配,以搜索与待识别图像最相近环境的过程。实践中,从待识别图像和环境库中图像各自检测出一组路标。因此,最相近环境意味着与待识别图像中路标匹配数量最多的环境库中图像,即最多相似性的图像。用匹配率衡量匹配的数量,即匹配的路标个

数占路标总数 ($\min\{N_{I_1}, N_{I_2}\}$, 其中, N_{I_1} 、 N_{I_2} 分别表示从待识别图像 I_1 和环境库中图像 I_2 所检测出的显著路标的数量) 的百分比。

5 实验及分析

5.1 显著性检测的稳定性比较

对来自 Berkeley 图像库的 20 幅图像进行了测试, 图像分辨率为 480×320 。建立纹理子图平均耗时 370.6 ms, 建立颜色子图平均耗时 123.2 ms, 全部运行时间不超过 600 ms。与体现 Itti 算法的 Eduardo 自然路标检测方法进行了比较, 其显著点检测的运行时间超过 3.7s^[6]。虽然考虑编程能力的差异, 该时间不一定完全反映问题, 但同等条件下本文方法显著性检测更迅速。从检测效果看, 由于图像分辨率高且亮度分布较均匀, 应用本算法能够有效地检测出显著性区域, 成功率可以达到 95% (19 幅图像的检测结果中包括了人类视觉选择的显著区域)。而对于显著性检测的准确性, 由于目前无法从定量的角度判断所检测的显著点是否显著, 因此, 本文仅采用人类视角的方法进行粗略判断^[3]。图 2 显示本文方法及 Eduardo 方法各自产生的显著性指

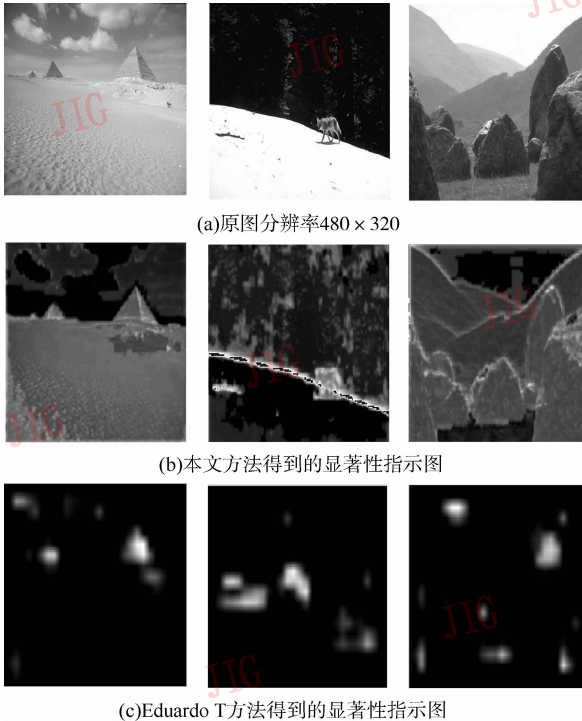


图 2 针对 Berkeley 图像库的显著性检测比较
Fig. 2 Comparison of saliency detection on Berkeley image benchmark

示图(灰度图), 图中越白的位置代表显著性越强, 可见本文方法得到的显著位置更准确。

5.2 自然环境识别结果及比较

为测试局部显著路标的场景识别能力, 分别在室内室外环境取 5 类图像构成场景图像库, 如图 3 所示。然后, 重新进入这些环境拍摄待识别图像, 再到场景库中寻找最好的匹配。待识别图像及识别结果如图 4、5 所示。



图 3 5 类图像构成的场景库

Fig. 3 Scene library including 5 class images



图 4 待测试图像

Fig. 4 Images to be test

采用最近邻方法进行匹配测试, 匹配判据采用 Jeffrey 距离^[2]。从图 5 的识别结果可以看出, 第 1、2 幅待识别图像所在的环境基本属于静态环境, 视角、尺度等变化对显著路标的检测影响较小, 利用本文方法在场景库中准确地找到了匹配图像。但对于第 3 幅图像, 场景中出现了原本不属于该场景的目标, 该目标的出现也直接影响了显著路标的提取和构造, 从而造成识别结果显得模棱两可, 即无法准确识别。

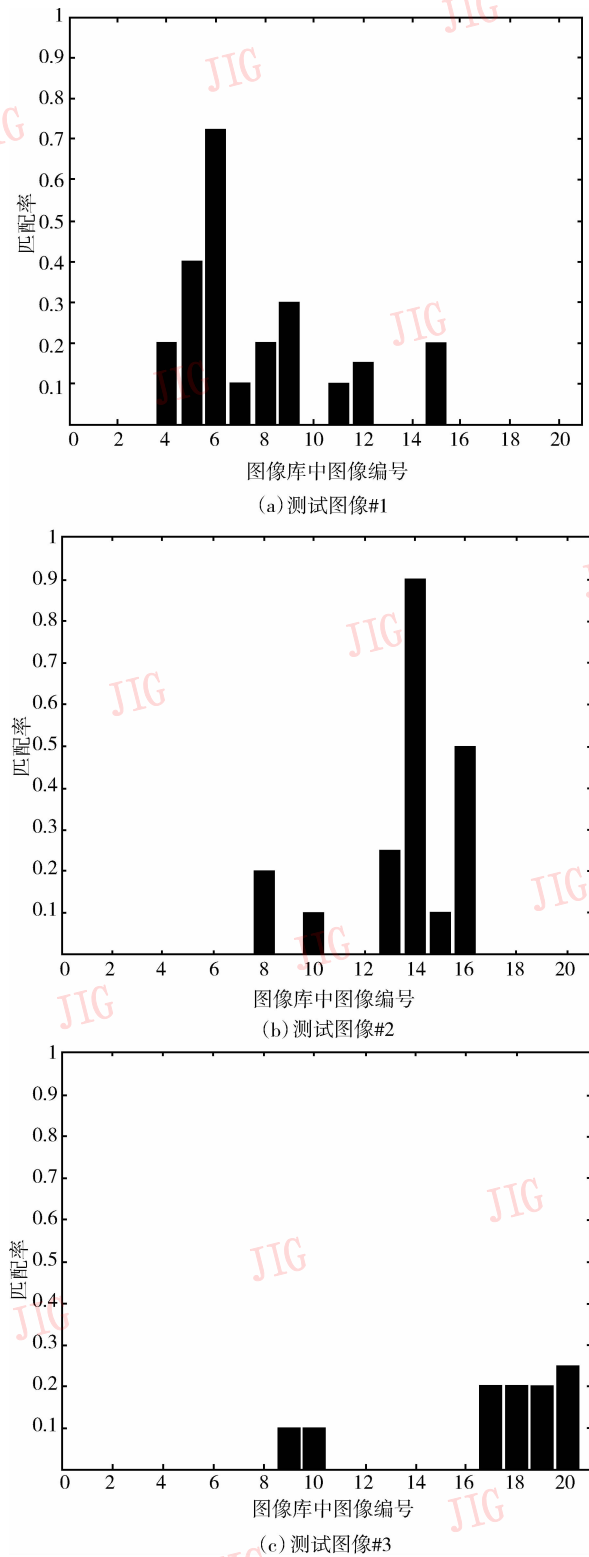


图 5 待测试图像与场景库中图像匹配的相似度
Fig. 5 Similarity between the image in library and the image to be test

代替全局图像描述子,本文方法在静态场景识别时具有较高的稳定性,较好地容忍尺度、视角等变化引起的差异,识别准确率较高,但对于动态场景则难以保证识别的可靠性。

进一步丰富目标环境图像库,在 5 类环境中各取 20 个位置,利用移动机器人携带的摄像头 (Sony EVI-D100) 在每个位置以间隔 36° 扫视环境 (即在每个位置得到 10 幅图像)。在所构成的图像库中,很多图像特别是同类环境图像间的差异更小。然后分别利用本文方法 (Method 1, M1) 和基于全局图像外观的识别方法 (Method 2, M2) 进行了 3 项环境识别实验。其中 M2 方法综合了几个场景识别系统的外观特征表示方案,包括颜色直方图、梯度、边密度和纹理^[1,2]。

实验 1 为随机从 3 个室内环境中任取 30 个位置上的图像,再回到这 3 个环境中进行识别。实验 2 为随机从走廊和室外两个环境中任取 30 个位置上的图像,再回到这 2 个环境中进行识别。实验 3 为随机从 5 类环境中任取 30 个位置上的图像,再回到这 5 类环境中进行识别。定义识准率“Truth”为待识别图像确实匹配到正确环境图像的比例。定义误识别“Error”为待识别图像匹配到错误环境图像的比例。定义不确定“Uncertainty”表示无法得到唯一判断的比率。识别结果如表 1 所示。

表 1 环境识别结果比较

Tab. 1 Comparison of environment recognition results

单位:%

识别	实验 1		实验 2		实验 3	
	M1	M2	M1	M2	M1	M2
识准率	86.67	83.33	80	90	86.67	80
误识别率	0	16.67	0	10	0	20
不确定	13.33	0	20	0	13.33	0

从表 1 可以看出,当待识别的环境差别较大 (如实验 2 所示) 时,采用基于全局图像外观的识别方法也可以获得较高的识别精度。当然,出现这种结果的另一个原因是,利用本文方法在走廊环境中得到的显著路标较少,因而分辨能力有所下降。但对于大部分待识别环境 (如实验 1、3 所示),本文方法都能够取得较基于全局外观识别方法更好的识别性能。

上述实验结果表明,由于采用局部图像描述子

6 结 论

本文基于局部特征具有更好不变性的思想,提出一种基于视觉局部显著区域的场景识别方法。为提高显著位置的稳定性,提出了带反馈的显著性检测模型 FSDM。基于分形进行自动尺度选择,以构造参与识别的局部显著区域。室内室外环境的场景识别实验表明,该方法与全局外观方法相比能够更好地容忍尺度、视角等变化引起的差异,静态场景识别时具有较高的稳定性。后续研究将针对动态环境,通过检测运动目标并剔除,来提高显著路标提取的稳定性。

参考文献 (References)

- 1 Ulrich I, Nourbakhsh I. Appearance based obstacle detection with monocular color vision [A]. In: Proceedings of the IEEE International Conference on Robotics and Automation [C]. San Francisco, CA, USA 2000: 1023 ~ 1029.
- 2 Zhou Chao, Wei Yu-chen, Tan Tie-niu. Mobile robot self-localization based on global visual appearance features [A]. In: Proceedings of the IEEE International Conference on Robotics and Automation [C]. Taipei, China, 2003: 1271 ~ 1276.
- 3 Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors [J]. International Journal of Computer Vision, 2004, **60**(1): 63 ~ 86.
- 4 Lowe D. Object Recognition from Local Scale-Invariant Features [A]. In: Proceedings of the 7th International Conference on Computer Vision (ICCV'99) [C]. Corfu, Greece, 1999: 1150 ~ 1157.
- 5 Itti L, Koch C, Niebur E. A Model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, **20**(11): 1254 ~ 1259.
- 6 Todt E, Torras C. Detection of nature landmarks through multiscale opponent features [A]. In: Proceedings of the International Conference on Pattern Recognition [C]. Barcelona, Spain, 2000: 3988 ~ 3991.
- 7 Wang Lu, Cai Zi-xing. Saliency based natural landmarks detection under unknown environments [J]. Pattern Recognition and Artificial Intelligence (Chinese), 2006, **19**(1): 100 ~ 105. [王璐, 蔡自兴. 未知环境下基于视觉显著性的自然路标检测 [J]. 模式识别与人工智能, 2006, **19**(1): 100 ~ 105.]
- 8 Zhang Peng, Wang Run-sheng. Detecting salient regions based on location shift and extent trace [J]. Journal of Software, 2004, **15**(6): 891 ~ 898. [张鹏, 王润生. 基于视点转移和视区追踪的图像显著区域检测 [J]. 软件学报, 2004, **15**(6): 891 ~ 898.]
- 9 Pentland A P. Fractal based Description of Nature Scenes [J]. IEEE Transactions on PAMI, 1984, **6**(6): 661 ~ 674.
- 10 Liu Jing, Zhong Wei-cai. Image retrieval based on fractal Dimensions [J]. Journal of Computer Research and Development, 2004, **41**(7): 1182 ~ 1187. [刘静, 钟伟才. 基于分数维的图像检索新方法 [J]. 计算机研究与发展, 2004, **41**(7): 1182 ~ 1187.]