doi:10.6041/j.issn.1000-1298.2015.07.002

基于 what-where 双通道理论的移动机器人场景仿生识别*

王富治[1]　宋昌林[1]　蒋代君[1]　冯代伟[2]

(1.西华大学机械工程学院,成都 610039;
2.电子科技大学机械电子工程学院,成都 611731)

摘要:为解决自主移动农业机器人在复杂工作环境进行视觉导航的场景识别,根据 what-where 双通道理论建立了场景感知模型以及场景表示模型,提出了一种基于概率框架的场景仿生识别方法。基于对比度假设和中心假设计算显著图并建立能量函数对该显著图进行优化;利用专家网络分析视觉注意焦点的内容作为 what 信息,以视觉注意焦点转移形成的扫视序列作为 where 信息;根据动作识别规则,利用 what 信息和 where 信息建立可观测的马尔可夫链模型实现场景识别。移动机器人场景识别过程与人的识别过程相似,实验证明所提方法对室内场景识别性能良好,准确率平均达 87.3%。

关键词:农业移动机器人　场景识别　what-where 双通道理论　视觉注意　马尔可夫模型

中图分类号:TP24　　**文献标识码**:A　　**文章编号**:1000-1298(2015)07-0010-07

Bionic Scene Recognition of Agricultural Mobile Robot Based on what-where Dual Channel Theory

Wang Fuzhi[1]　Song Changlin[1]　Jiang Daijun[1]　Feng Daiwei[2]

(1. *School of Mechanical Engineering*, *Xihua University*, *Chengdu* 610039, *China*
2. *School of Mechatronic Engineering*, *University of Electronic Science and Technology of China*, *Chengdu* 611731, *China*)

Abstract: Scene recognition is the key to visual navigation for the agricultural mobile robot in unknown environment. This paper used what-where dual channel theory to build the models of scene perception, scene representation and scene recognition, and proposed a bionic method of scene recognition on the basis of probabilistic framework. This method first computed the bottom-up saliency map of scene based on the contrast prior and the center prior, which can be further optimized with the global energy function. Then shifted the visual focus of saliency map to obtain the saccade sequence as the "where information", and analyzed the content of the visual focus to obtain the "what information" with the experts network comprised of single layer perceptron. Lastly, according to the action recognition regularity of human, built the discrete and observable Markov model using the "what information" and the "where information". The parameters of the model can be determined by training the frame images from the camera on the mobile robot and can be viewed as the prior knowledge about different scenes, which can be recognized by maximizing the likelihood probability of the Markov recognition model. The whole recognition process is similar to human's. Experimental results show that this method has good performance for indoor scenes and the recognition accuracy averaged out at 87.3%.

Key words: Agricultural mobile robot　Scene recognition　what-where double channel theory　Visual attention　Markov model

收稿日期:2014-10-22　修回日期:2015-03-13

*教育部春晖计划资助项目(12202528)、四川省制造与自动化高校重点实验室开放基金资助项目(SZJJ2011-019)和西华大学校重点项目(Z1120223)

作者简介:王富治,讲师,博士,主要从事机器视觉方向研究,E-mail:redfires_789@163.com

引言

基于视觉的自主移动农业机器人工作于野外，其环境复杂且动态变化，以设置人工路标进行视觉导航的方法对于移动农业机器人来说是不适用的。为保证移动农业机器人对环境的适应性，需在场景识别的基础上建立有效的视觉导航策略。从模式识别的观点来看，场景识别本质上是一个从大量、高维的视觉环境信息中抽取出有效低维特征对不同场景进行模式分类的过程。移动机器人场景视觉感知与识别常用的特征抽取方法主要有图像外观（如边缘，颜色等）、兴趣点（如 Harris 角点、SIFT、SUFT 等）以及直方图特征等；而常用的高维特征降维方法可分为线性方法和非线性方法。线性方法有主成分分析、独立成分分析等；非线性方法有支持向量机、Kernel 主分量分析[1]、以流形学习为基础的特征提取理论和技术等[2]。总的来说，这些方法与生物视觉机制相差较远，具有很大的局限性，体现在：框架的输入是被动的，缺乏主动性；框架中信息加工流程基本自下而上，缺乏反馈；缺乏高层知识的指导作用等。由于这些局限的存在，导致以这些方法为基础的场景识别技术往往只能满足于简单的结构化环境，很难满足真实复杂环境（如室外环境）下农业移动机器人同步定位与建图（SLAM）问题的需要。

因此，如何从生物机制中寻求借鉴设计环境特征的有效的抽取模型和算法成为农业移动机器人研究富有吸引力的领域。其中由 Ungerleider 和 Mishkin 根据猴子大脑损伤实验所提出的 what-where 双通道视觉认知理论是一个值得关注的理论[3]。what-where 双通道理论的实现与视觉注意机制密切相关。视觉注意机制通过视觉注意焦点的选择与转移，可以有效解决巨量高维信息与视觉系统有限信息处理能力之间的矛盾，这一特性对于需要处理大量动态环境信息的移动机器人来说，具有重要意义。因而，基于视觉注意机制的移动机器人场景感知与识别研究日趋活跃[4-6]。

从 what-where 双通道理论的观点来看，文献[4-6]的不足之处在于，要么只用到 what 信息，要么 what-where 信息缺乏有机的内在联系，因而这些文献所提出的方法性能难以令人满意。视觉注意机制不仅仅是一种高效信息处理方式，同时也是人类的一种认知形式。欲采用 what-where 双通道理论来提高移动机器人的场景感知能力，一方面需要建立 what-where 信息的内在联系，或者说对 what-where 信息合理建模；另一方面就是如何把自顶向下视觉注意机制和自底向上视觉注意机制结合起来。目前这方面的研究不多，较有代表性的有 Rybak 模型以及 Salah 模型。Rybak 模型根据 what-where 双通道理论建立起对象的“动作识别规则”，实验证实该规则具有识别经平移、旋转和比例缩放复杂图像的能力[7]。Salah 将可观测 Markov 模型引入模拟任务驱动的注意机制中，在数字识别和人脸识别中取得了较好效果[8]；最近，斯坦福大学 Elazarya 和 Itti 提出一个基于视觉注意的视觉搜索和目标识别的 Bayes 模型，也是基于 what-where 双通道理论，用来进行航拍图像目标搜索与识别，取得了较好的效果[9]。

本文主要在 what-where 双通道理论基础上，提出移动机器人场景识别方法。该方法的特点在于在获取场景显著图的基础上，利用可观测的马尔可夫模型对 what 信息和 where 信息进行处理，从而建立关于移动机器人场景识别的概率框架。

1　场景感知模型与场景表示模型

1.1　动作识别规则

what-where 理论认为：当生物感知外部世界时，人类视觉系统可以分为具有层次结构的 2 条视觉通路或 2 个子系统，其中 what 子系统主要与对象的形状、大小、颜色等静态特征（称为 what 信息）有关，用来形成感受和进行对象识别；而 where 子系统则主要用来处理空间位置信息（称为 where 信息）。两条通道的交互作用使得人类能够适应复杂的外部世界。

基于 what-where 双通道理论，Rybak 在文献[7]中阐明了人类的“动作识别规则”。所谓“动作识别规则”，是指人类在视觉感知和识别过程中，注意焦点不断移动并较多停留在图像中包含信息较多的部分；同时积极地选择下一注意焦点。因此，视觉感知和识别可以看作为一个“动作”过程。从动作的观点来看，新场景、新对象的一种内部表示（模型），是在人们有意识地积极观察和检查的过程中建立起来的。积极检查的目的在于，发现和记住在此过程中采取的动作和其结果之间的变化之间的关系。因此，感知对象的内部表示是包含一系列的“动作（where 信息）”和“感觉（what 信息）”的交互过程。当系统能够下意识地处理物体，并能够预测出当采取特定动作时物体的反映时，外部物体就变得“可感知”和“可识别”。

Rybak 建立的识别模型非常复杂繁琐，不利于应用。本文根据马尔可夫链建立了新的基于动作规则的场景感知与识别模型。

1.2 场景感知模型

由动作识别规则可知,移动机器人在积极检查环境的过程中提取关于场景“where 信息”和“what 信息”,当机器人能够根据感知模型结合场景模型测出当采取特定动作时物体的反映时,外部物体就变得“可感知”和“可识别“。本文采用马尔可夫(Markov)链来处理这一问题,因此首先给出 Markov 链的定义:

随机序列 X_n,在任一时刻,它可以处在状态 ω_1,ω_2,…,ω_N,且它在 $m+k$ 时刻所处的状态 q_{m+k} 的概率,只与它在 m 时刻的状态 q_m 有关,而与 m 以前的状态无关,即

$$P(X_{m+k}=q_{m+k}\mid X_m=q_m,X_{m-1}=q_{m-1},\cdots,X_1=q_1)=P(X_{m+k}=q_{m+k}\mid X_m=q_m) \tag{1}$$

这里,$q_1,\cdots,q_m,q_{m+k}\in(\omega_1,\omega_2,\cdots,\omega_N)$,称 X_n 为 k 阶 Markov 链,并且称 $a_{ij}=P(q_{m+k}=\omega_j\mid q_m=\omega_i)$ 为 k 步转移概率。

基于 Markov 链的移动机器人感知模型如图 1 所示。该模型主要分为 4 层:注意层、中间层、联接层和识别层。其中注意层用于生成场景显著图,中间层用于提取 what 和 where 信息,联接层则实现注意层和中间层的联接,识别层则基于马尔可夫概率框架实现对场景的识别。

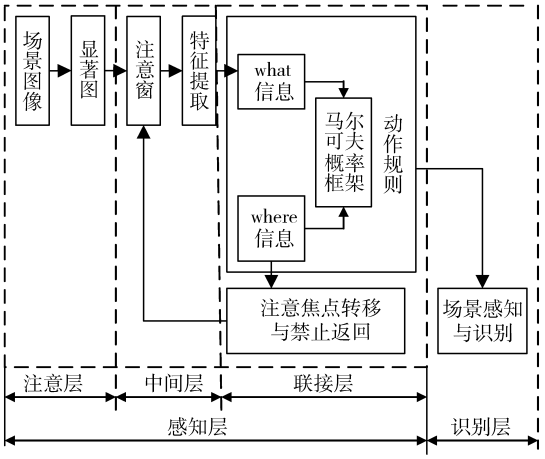

图 1 移动机器人感知模型

Fig. 1 Perception model of mobile robot

1.3 基于马尔可夫链的场景表示模型

要实现移动机器人对场景的识别,除了感知模型,还需要建立与感知模型相适应的场景表示模型。根据图 1 所示的感知模型,可以得到移动机器人基于马尔可夫模型的场景表示模型,如图 2 所示。

图 2 中,ω_j 表示由视觉注意焦点获得的 Markov 链的状态;b_j 表示状态的观测值。该模型将移动机器人场景的识别问题转换为 Markov 链的估值问题,从而有效提高了场景识别的准确性和识别算法的效率。

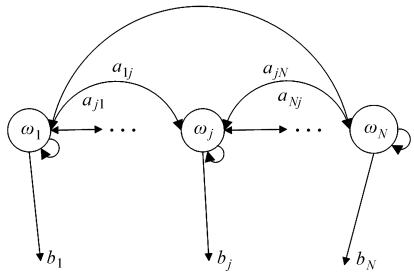

图 2 移动机器人场景模型

Fig. 2 Scene model of mobile robot

2 场景感知实现

2.1 注意层

注意层模拟自底向上的视觉注意机制生成兴趣图。Itti 基于底层视觉特征整合理论的选择性视觉注意算法具有里程碑的意义[10],但该算法也存在若干不足:如背景不能有效抑制;兴趣目标不能一致检测等。Itti 算法不足的原因在于,直接以亮度、颜色、朝向作为底层特征使用,而未考虑到不同特征区域之间的相互位置关系。如直接采用 Itti 算法生成的显著图来进行场景识别,效果并不理想。

为避免 Itti 算法的不足,近来不断有文献在颜色、灰度、朝向等底层特征基础上提出“对比度假设”(Contrast prior)[11-12] 和“中心假设”(Center prior)等来获取视觉显著图[13-14],取得了令人满意的效果。

文献[13-14]提出的“中心假设”基于感兴趣目标处于图像中心。而对于移动机器人来说,感兴趣目标可能处于视野图像中的任何位置。因此,为了获得性能良好的兴趣图,首先采用 SLIC 算法对图像进行超像素(Superpixel)分割[15],然后采用对比度假设和基于凸包技术的居中假设获取初始显著图,其中凸包技术(Convex hull)用来估计每个显著区域的中心[16],最后采用基于图的平滑约束对兴趣图进行全局优化。

2.1.1 初始显著图

初始显著图通过对比度假设和中央假设获得。设图像经超像素分割后的任一超像素 i,其在 Lab 颜色空间的均值为 $\boldsymbol{c}_i$,其坐标均值为 $\boldsymbol{p}_i$,那么超像素基于对比度假设的显著性定义为[17]

$$S_{co}(i)=\sum_{j\neq i}\|\boldsymbol{c}_i-\boldsymbol{c}_j\|\exp\left(-\frac{\|\boldsymbol{p}_i-\boldsymbol{p}_j\|^2}{2\sigma_p^2}\right) \tag{2}$$

式中 $\exp\left(-\frac{\|\boldsymbol{p}_i-\boldsymbol{p}_j\|^2}{2\sigma_p^2}\right)$——空域加权算子

σ_p——空域加权强度系数

对比度假设常常不能正确地检测出某些背景超

像素，这一问题可通过引入“基于凸包的中心假设”来解决。设对于图像经超像素分割后的任一超像素 i，采用凸包技术获取其中心为 (x_0, y_0)，那么基于“中心假设”的该超像素的显著性定义为

$$S_{ce}(i) = \exp\left(-\frac{\| x_i - x_0 \|^2}{2\sigma_x^2} - \frac{\| y_i - y_0 \|^2}{2\sigma_y^2} \right) \quad (3)$$

式中　x_0、y_0——超像素 i 所有像素行坐标与列坐标的平均值

σ_x、σ_y——行、列坐标的方差，通常取 $\sigma_x = \sigma_y$

最后，根据特征整合理论，由上述 2 个假设所得到的显著图进行相乘，得到初始兴趣图 S_{in}

$$S_{in}(i) = S_{co}(i) S_{ce}(i) \quad (4)$$

2.1.2　初始显著图的优化

由式(4)所得显著图的一致性不太好。一致性是指同属一个目标的不同区域应尽可能具有相同或相似的显著度。而来源于同一目标的不同区域，一般具有相同或相似的色度。因此，图像经超像素分割后，可利用这一特点，对不同超像素的初始显著值建立全局能量函数进行优化[17]。

针对超像素，全局能量函数为

$$E(S) = \sum_i (S(i) - S_{in}(i))^2 + \lambda \sum_{i,j} w_{ij} (S(i) - S(j))^2 \quad (5)$$

这里，w_{ij}为根据前述关于显著度的色度约束所引入的权值，定义为

$$w_{ij} = \exp\left(-\frac{\| \boldsymbol{c}_i - \boldsymbol{c}_j \|}{2\sigma_w^2} \right) \quad (6)$$

式中，$S(i)$、$S(j)$表示 2 个相邻超像素的显著度。式(5)中第 1 项为拟合约束，表示经过优化后的显著度值应与初始值相差很大；式(5)中的第 2 项为平滑约束，表示相邻的且色度相近的超像素应具有相似的显著度。

式(5)的最优解可通过最小化能量函数 $E(s)$ 来获得。令 $E(S)$ 关于 S 的导数为 0，可得到关于 S 的最优解 S^* 为

$$S^* = \mu(\boldsymbol{D} - \boldsymbol{W} + \mu\boldsymbol{I})^{-1} S_{in} \quad (7)$$

这里，$\mu = 1/(2\lambda)$，$\boldsymbol{D}$ 为对角矩阵，其中 $d_{ii} = \sum_j (w_{ij})$；$\boldsymbol{I}$ 为单位阵。

最后，为保证同一目标区域具有相同的显著度，又对式(7)所得到的显著图进行 k-means 分割处理，从而得到最终的显著图。

2.2　中间层

中间层的主要功能是通过专家网络来分析视觉注意焦点的内容或特征，即获取 what 信息，同时通过视觉注意焦点的转移，获取 where 信息。

2.2.1　where 信息获取

显著图产生了一系列的待注意目标，各目标通过胜者为王(winner-take-all)机制吸引注意焦点，焦点在各注意目标之间依注意度从大到小和禁止返回机制原则进行转移，注意焦点的位置产生一系列的 where 信息。

2.2.2　what 信息获取

为获取 what 信息，中间层由单层感知器组成专家网络，其输入为从视觉注意焦点获取的图像特征向量，输出为该信息所属的后验概率向量，即 what 信息。获取 what 信息主要包含如下步骤：

(1) 选取视觉注意焦点的图像视觉特征

采用与 Itti 类似的方法，选取颜色、亮度和朝向 3 种。其中颜色特征为 4 个宽调谐颜色通道，即红色：$R = r - (g + b)/2$；绿色：$G = g - (r + b)/2$；蓝色：$B = b - (r + g)/2$；黄色：$Y = (r + g)/2 - |r - g|/2 - b$。亮度图像由 $I = (r + g + b)/3$ 得到。朝向特征使用 Gabor 小波对 I 进行分解，得到 4 个不同朝向，$\theta \in \{0°, 45°, 90°, 135°\}$。

(2) 生成特征图

由于视网膜中的感受野一般是由中心的兴奋区域和周边的抑制区域构成的同心圆结构，故可以通过中心-边缘操作模拟该同心圆结构。对于已选取的视觉特征，可在此基础上通过计算图像不同特征通道的精细金字塔尺度 c 和周边粗糙金字塔尺度 s 的“中央周边差”运算得到特征图。

考虑到图像尺寸，选取 $c \in \{3, 4, 5\}$，$s \in \{6, 7, 8, 9\}$ 中的 6-3，7-3，7-4，8-4，8-5，9-5 共计 6 种组合。针对亮度、颜色、朝向 3 种特征，其特征图分别定义为：

亮度特征图

$$I(c,s) = |I(c) \Theta I(s)| \quad (8)$$

颜色特征图

$$\begin{cases} RG(c,s) = |(R(c) - G(c)) \Theta (G(s) - R(s))| \\ BY(c,s) = |(B(c) - Y(c)) \Theta (Y(s) - B(s))| \end{cases} \quad (9)$$

朝向特征图

$$O(c,s,\theta) = |O(c,\theta) \Theta O(s,\theta)| \quad (10)$$

共计 42 幅特征图。

(3) 由专家网络获得 what 信息

采用单层感知器组成专家网络，用来分析视觉注意焦点的内容。输入为由全部 42 幅特征显著图(选金字塔尺度为 4)的视觉焦点窗口(选 5×5)经 PCA 处理后所获取的 80 维特征向量，输出为该信息所属类别的后验概率向量，即 what 信息。

对某一场景图像所获得的特征图像如图 3 所

示。由于中间层采用单层感知器作为专家网络,降低了计算的复杂性同时提高了精度。

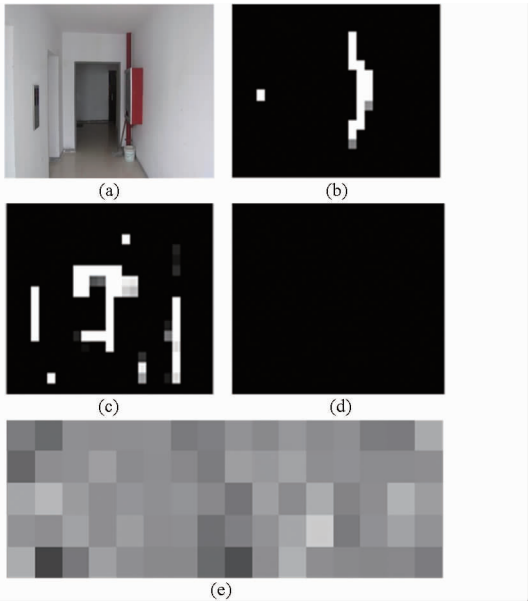

图 3 场景图像特征图

Fig. 3 Feature map of scene image

(a) 楼道场景图 (b) 颜色特征图 (c) 亮度特征图
(d) 朝向特征图 (e) 获取的特征向量

2.3 联接层

联接层利用离散的可观测马尔可夫链联接注意层和中间层,被注意焦点访问过的区域作为马尔可夫模型的状态(where 信息),专家网络的输出(what 信息)作为状态的观测值,根据这 2 种信息调整单个马尔可夫链的概率而最大化某个训练样本形成的特定的扫视路径序列的似然值。

在学习过程中,因为每个状态都是可以观测的,所以可以通过计数的方式得到状态转移概率 a_{ij} 和初始状态分布概率 π_i 以及每个样本在每个状态下的专家网络输出得到状态观察值 $b_j(k)$。因此,观测序列的概率为

$$P(O,L|\lambda)=\pi_{L_i}b_{L_i}\prod_{i=2}^{N}a_{L_{i-1}L_i}b_{L_i}(O_i) \tag{11}$$

式中,L 表示状态序列,O 表示观测序列;$\lambda=\{\pi_i,a_{ij},b_j(k)\}$ 表示 Markov 链的参数,$i,j=1,2,\cdots,N$ 对应状态的个数,或者说为焦点转移次数。$k=1,2,\cdots,M$ 对应观测样本的状态数。

3 场景识别

当到某一时刻,设已经有 t 个区域被注意过,判断待识别图像属于某个场景的概率记为

$$a_t(c)=P(O_1,\cdots,O_t,L_1,\cdots,L_t|\lambda_c) \tag{12}$$

式中,$O_1,\cdots,O_t$ 为观测序列,$L_1,\cdots,L_t$ 为状态序列。焦点转移停止的标准是焦点区域的显著度已低于事先规定的阈值,或该概率已达到能做出决策的置信度。因此,在识别过程中,只需要让待识别图像经历有限次数的焦点转移,就能做出正确的识别判断。定义到 t 时刻,图像属于某场景 c 的后验概率定义为

$$a_t^*(c)\overset{\text{def}}{=}P(c|O_1,\cdots,O_t,L_1,\cdots,L_t)=\arg\max_c\frac{a_t(c)}{\sum_{j=1}^{k}a_t(j)} \tag{13}$$

设置信度为 τ,则焦点停止转移的标准为

$$a_t^*(c)\geqslant\tau \tag{14}$$

4 实验与分析

4.1 实验设计

对于农业移动机器人视觉导航,考虑到问题的难度,首先考虑室内环境。实验是在 Xparter 移动机器人实验平台上进行的。与人在进行场景识别时类似,让机器人依房间→门口→走廊→楼道的顺序在一定范围内对这几个场景(这几个场景是进行移动机器人室内视觉导航的几个关键场景)进行移动摄像,从摄像头视频中对视频帧按一定间隔进行选取。该移动机器人所用摄像头的帧率为 30 帧/s,按 3 帧/s 的间隔选取。同一场景的不同帧应包含场景的主要区域。

实验分为 2 步,第 1 步获取如上关键场景的样本图像,由式(7)得到各图像的显著图,然后可以这些显著图为样本对识别模型中的一些参数进行确定。这些参数可作为关于场景的知识保存起来,以用于之后的场景识别;第 2 步从样本图像中选取一部分图像,或者让机器人重新移动获取场景图像以进行识别,以检测本文方法的效果。

4.2 参数确定

所采用的可观测马尔可夫模型包含的参数为:$\lambda=\{\pi_i,a_{ij},b_j(k)\}$。根据前文所述,这几个参数的确定方法为

$$\begin{cases}a_{ij}=\dfrac{N_{L_i\to L_j}}{N_{L_i}}\\ \pi_i=\dfrac{N_{L_i(t=1)}}{\sum_i N_{L_i(t=1)}}\\ b_j(k)=\dfrac{N_{L_j|O_i}}{N_{L_j}}\end{cases} \tag{15}$$

式中 N_{L_i}——从 L_i 开始的转移次数

$N_{L_i\to L_j}$——从状态 L_i 转移到状态 L_j 的次数

$N_{L_i(t=1)}$——状态序列从 L_i 开始的次数

$N_{L_j|O_i}$——观察 O_i 时 L_j 出现的次数

当确定了 $b_j(k)$ 后,可以利用 $b_j(k)$ 对单层感知

器基于 δ 学习规则进行训练确定输出层与输入层之间的联接权值。感知器网络的输入层具有 80 个神经元节点，输出层神经元节点数量与离散马尔可夫状态数及计算精度有关，本文取为 10。通过训练确定模型相关参数后，即可利用该模型应用于移动机器人场景识别。

4.3　实验结果与分析

针对楼道、房间、走廊、门口等场景，在移动机器人的某次移动过程中从视频帧中共选取了 409 帧图像，其中 211 帧用以进行训练，198 帧用来进行识别。首先获取显著图，如图 4 所示。从结果可知，显著图与人的视觉经验一致，从而为场景识别打下了良好基础。图 4d 显示了某楼道图像的显著图及其根据 WTA 准则的视觉焦点转移的过程。这些显著区域实质是关于场景的关键信息，根据这些信息所得到的移动机器人在该次移动过程中的识别结果如表 1 所示。

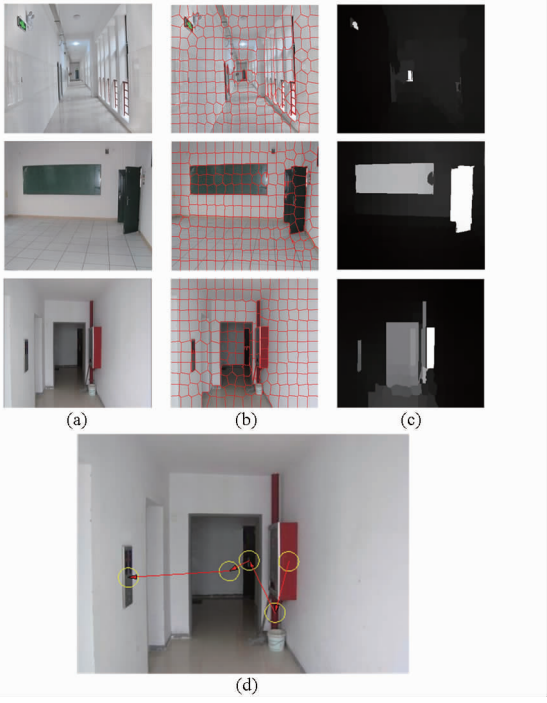

图 4　识别过程中的视觉注意焦点转移

Fig. 4　Shifting of visual focus for recognition

（a）场景图　（b）超像素分割结果图

（c）显著图　（d）楼道场景视觉注意焦点转移

为了更好反映移动机器人场景仿生识别过程，在利用样本进行参数确定获得关于场景知识的基础上，让机器人重新对楼道场景进行移动摄像（移近后移开），在移动过程中取 8 帧图像，根据式（11）所获得的观测概率如图 5 所示（该图为基于楼道 1、房间 1、门口 1 和走廊 1 所获得的数据）。对于其他场景的实验，具有类似的结果。图 5 说明，当移动机器人已获得关于楼道场景的知识，这些知识对其他场景具有良好的区分作用，进而识别楼道，但这样的识别过程受到移动机器人所处方位、背景光等的影响，其对场景的识别往往是一个动态的过程，即对场景的识别概率是一个由小到大的过程，在某一方位观测概率具有最大值，根据该最大概率可实现对场景的可靠识别。

表 1　室内场景识别结果

Tab. 1　Recognition results for indoor scene

场景类型	正确识别数	错误识别数	测试正确率/%	训练正确率/%
房间 1	22	3	88.0	93.7
房间 2	20	2	90.9	94.8
房间 3	24	4	85.7	95.3
楼道 1	17	2	89.5	93.6
楼道 2	17	3	85.0	92.1
走廊 1	18	4	81.8	94.2
走廊 2	23	5	82.1	93.8
门口 1	15	1	93.8	94.5
门口 2	16	2	88.9	93.2

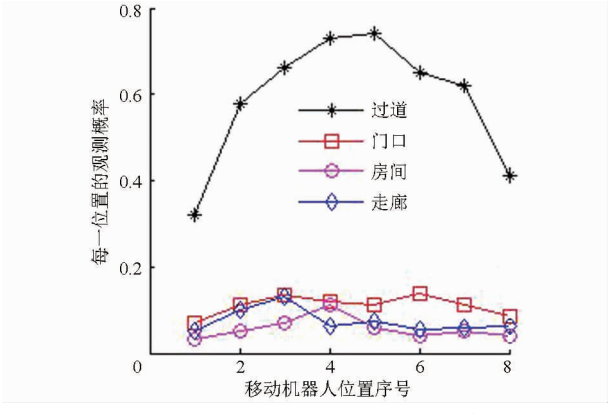

图 5　楼道口场景的动态识别概率

Fig. 5　Dynamic recognition of lane scene of mobile robot

由此可见，对于室内场景来说，本文的方法具有较高的识别准确率。这一方面由于本文基于概率模型进行识别，另一方面本文方法是在获取场景显著图的基础上建立马尔可夫模型进行识别。因而，视觉显著图的质量对于识别的可靠性产生决定性影响。本文产生视觉显著图的方法具有如下特点：由于采用了“对比度假设”和“中心假设”，2 种假设不像 Itti 算法一样，显著性计算直接依赖于亮度、颜色等底层特征，因而对环境光变化引起的识别具有一定的抵抗作用，同时能较好地抑制背景，并能较好保持目标区域的一致性。但是，这种识别方法较一般方法运算量要大得多。

本文产生视觉显著图的方法仍然属于底层特征驱动法，对室外场景来说，如街道以及移动农业机器人所工作的野外环境等，往往包含各种特征，背景光变化复杂，尚不能取得理想的显著图，因而识别效果

尚不能令人满意,需要引入任务驱动的视觉显著性进行改善。

识别过程有一参数 τ。τ 值愈大,识别精度愈高,但识别时间愈长。τ 的大小取决于 2 个因素:焦点转移的次数和识别精度。在室内场景识别的实验中,取 $\tau=0.75$,焦点转移平均次数仅为 3.12,说明在场景识别中,仅仅需要部分信息就可实现对场景的可靠识别。

5 结束语

场景识别是未知环境下农业移动机器人视觉导航的一个非常重要问题。受生物视觉机制的启发,提出了一种基于 what-where 双通道理论的移动机器人场景识别方法。建立了机器人的感知模型以及与此感知模型相适应的场景表示模型,其具体识别过程是首先利用对比度假设和中心假设获取场景显著图,并对该显著图利用平滑约束进行优化,从而可以更好地抑制背景,同时显著目标具有更好的一致性;然后以该显著图为基础,使用 Gaussian 金字塔分解和 Garbor 滤波提取视觉注意焦点的图像特征,经单层感知器专家网络处理后获得 what 信息,视觉注意焦点的转移序列形成 where 信息;最后利用离散的可观测马尔可夫模型,根据 what 信息和 where 信息调整 Markov 链的概率,从而实现场景的正确识别。该方法本质上是将注意焦点的知识保存在 Markov 链的状态信息中,根据当前注视点的内容确定下一注意焦点的知识记忆保存在转移概率中,在识别过程中,这些保存的记忆信息作为自顶向下的信息指导扫视路径,使得机器人具有良好的场景识别能力。

参 考 文 献

1 Scholkopf B, Smola A. Learing with kernels[M]. Cambridge: MIT Press, 2002.

2 Smola A J, Mika S, Scholkhopf B, et al. Regularized principal manifold[J]. Journal of Machine Learning Research, 2001, 1(3):179 - 209.

3 Ungerleider L G, Mishkin M. Two cortical visual systems[M]// Ingle D J, Goodale M A, Mansfield R J W. Analysis of Visual Behavior. Cambridge, MA: The MIT Press, 1982:549 - 588.

4 Christian Siagian, Laurent Itti. Rapid biologically-inspired scene classification using features shared with visual attention[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 300 - 312.

5 Fernando Lopez-Garcia, Xose Ramon Fdez-Vidal, Xose Manuel Pardo, et al. Scene recognition through visual attention and image features: a comparison between SIFT and SURF approaches[G]// Object Recognition. INTECH Open Access Publisher, 2011:185 - 198.

6 Julio Vega, Eduardo Perdices. Robot evolutionary localization based on attentive visual short-term memory[J]. Sensors, 2013, 13(1): 1268 - 1299.

7 Rybak I A, Gusakova V I, Golovan A V, et al. A model of attention-guided visual perception and recognition[J]. Vision Research, 1998, 38(15 - 16):2387 - 2400.

8 Salah A A, Alpaydin E, Akarun L. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligeuce, 2002, 24(3):429 - 425.

9 Lior Elazary, Laurent Itti. A bayesian model for efficient visual search and recognition[J]. Vision Research, 2010, 50(14):1338 - 1352.

10 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.

11 Perazzi F, Krahenbuhl P, Pritch Y, et al. Saliency filters: contrast based filtering for salient region detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012:733 - 740.

12 Cheng M M, Zhang G X, Mitra N J, et al. Global contrast based salient region detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2011:409 - 416.

13 Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 32(10):1915 - 1925.

14 Jiang H, Wang J, Yuan Z, et al. Automatic salient object segmentation based on contex and shape prior[C]// Proceedings of the British Machine Vision Conference Dundee: BMVA Press, 2011: 110.1 - 110.12.

15 Achanta R, Smith K, Lucchi A, et al. SLIC superpixels[R]. EPFL Technical Report, 2010.

16 Xie Y L, Lu H C. Visual saliency detection based on Bayesian model[C]// 2011 18th IEEE Internetional Conference on Image Processing, 2011:645 - 648.

17 Chuan Yang, Lihe Zhang, Huchuan Lu. Graph-regularized saliency detection with convex-hull-based center prior[J]. IEEE Sigal Processing Letters, 2013, 20(7):637 - 640.