

文章编号: 1001-0920(2012)09-1320-05

# 上下文场景识别模型的稀疏贝叶斯判别学习方法

陈雷, 陈启军

(同济大学电子与信息工程学院, 上海 201804)

**摘要:** 在机器人场景识别问题中, 将连续场景的相关性通过基于隐马尔可夫模型的上下文模型进行描述. 采用不同于传统的使用生成模型方法学习上下文场景识别模型的方式, 首先引入稀疏贝叶斯学习机对上下文模型中图像特征的后验概率进行建模, 然后通过贝叶斯原理将稀疏贝叶斯模型与隐马尔可夫模型结合, 提出一种能够实现上下文场景识别模型的判别学习方法. 在真实场景数据库上的实验结果表明, 由该方法得到的上下文场景识别系统具有很好的场景识别能力和泛化特性.

**关键词:** 场景识别; 上下文模型; 稀疏贝叶斯学习; 隐马尔可夫模型  
**中图分类号:** TP24                      **文献标志码:** A

## Sparse Bayesian discriminative learning method for context-based scene recognition model

CHEN Lei, CHEN Qi-jun

(School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China. Correspondent: CHEN Qi-jun, E-mail: qjchen@mail.tongji.edu.cn)

**Abstract:** For the robotics scene recognition problem, the relationship between continuous scenes can be modeled by a hidden Markov model based context model. Unlike traditionally used generative method to learn this model, a sparse Bayesian learning machine is adopted to model the posterior probabilities of image features. Then by combining the sparse model with the hidden Markov model using Bayes theory, a discriminative learning method of the context scene recognition model is proposed. The experiments on a real scene database show that the obtained scene recognition system possesses good recognition performance and generalization ability.

**Key words:** scene recognition; context model; sparse Bayesian learning; hidden Markov model

## 1 引言

传统的机器人定位与拓扑地图构建中, 机器人并不提取环境语义上的信息<sup>[1]</sup>. 然而, 随着研究的深入, 特别是近几年对语义地图构建等问题的研究, 人们已经认识到引入环境语义上的认知对于提高移动机器人的智能以及与环境交互的能力具有重要意义, 而视觉场景识别则是一条重要途径. 移动机器人视觉场景识别要求机器人只通过视觉设备采集当前环境的图像便能够对该环境或场景的类别作出语义上的判断, 例如会议室、咖啡厅和走廊等.

目前, 人们关于场景识别所进行的一些研究工作<sup>[2-5]</sup>已经取得了许多进展, 但这些方法通常将场景图像视为相互独立的数据样本. 对于移动机器人应用

而言, 这种假设显然不符合人们对连续场景图像序列的直观认识. 一般认为, 机器人在一段时间内只可能同一类场景中活动, 前一时刻的场景往往只能反映这一时刻机器人所在场景的各种可能性. 因此, 若直接忽略场景之间内在的联系, 必然会影响场景预测结果的准确性.

为利用上下文信息提高识别的鲁棒性, Torralba等<sup>[6]</sup>提出了一种使用隐马尔可夫模型(HMM)对图像上下文关系进行建模的方法. 该模型使用HMM对上下文关系进行描述, 并使用高斯混合模型GMM学习图像特征的似然概率. 注意到, 这种基于HMM的场景上下文描述模型往往与其他生成模型的识别方法相结合进行学习, 除Torralba等<sup>[6]</sup>使用的GMM, 文献

收稿日期: 2011-04-14; 修回日期: 2011-09-07.

基金项目: 国家863计划项目(2009AA04Z213); 国际科技合作计划项目(2010DFA12210); 上海曙光跟踪项目(10GG11); 上海市科技人才计划项目(11XD1404800).

作者简介: 陈雷(1984—), 男, 博士生, 从事计算机视觉、统计学习等研究; 陈启军(1966—), 男, 教授, 博士生导师, 从事机器人控制、智能系统等研究.

[7]使用了贝叶斯网络, [8]使用了类似于语音识别领域的一种概率语言模型, 而 [9]则使用了朴素贝叶斯分类器. 近年来, 基于判别模型的图像识别方法(如 SVM, boost 等), 以其出色的识别性能和效率越来越得到人们的广泛关注. 但是, 这些判别模型算法的主要目的是作出决策, 缺乏具有概率意义的输出, 而基于 HMM 的上下文识别框架中需要对图像特征的似然概率进行建模, 因此一般无法直接将这些判别模型算法与该上下文模型结合进行学习训练.

本文基于 Torralba 等<sup>[6]</sup>的上下文描述框架, 提出一种利用稀疏贝叶斯判别模型——相关向量机<sup>[10]</sup>对模型进行训练的方法. 该方法一方面保留了相关向量机计算效率高、泛化性能好的优点; 同时又充分利用了其具有对图像特征后验概率建模的特点, 结合贝叶斯原理实现了对上下文场景识别模型的判别学习. 在模拟机器人运动的真实环境图像数据库<sup>[9]</sup>中的实验表明, 相比通过生成模型学习的上下文模型, 本文提出的模型具有更高的识别率与泛化特性, 同时具有很高的计算效率.

## 2 判别学习的上下文场景识别模型

本文首先基于 Torralba<sup>[6]</sup>的上下文模型框架, 构建了场景识别系统. 直观上可以想象, 连续的场景之间必然存在一定的相关性. 对于具有  $N_q$  类场景的连续图像, 用  $Q_t, \mathbf{x}_t$  分别表示  $t$  时刻图像对应的场景类别和特征, 其中  $Q_t \in \{1, 2, \dots, N_q\}$ . 理想的场景识别模型预测结果应不仅仅基于当前的最新图像观测, 还需综合考虑过去时刻观测的信息. Torralba<sup>[6]</sup>提出的上下文识别模型便使用了隐马尔可夫模型(HMM)的方法建立了前后场景图像之间的相关性, 并将视觉系统预测的目标由传统的估计条件概率分布  $P(Q_t|\mathbf{x}_t)$  推广为估计  $P(Q_t|\mathbf{x}_{1:t})$ .

移动机器人在环境中运动, 从一个场景进入另一个场景, 可看作隐马尔可夫模型中一个状态至另一个状态的切换; 因此, 可将图像所属的场景类别视为潜在变量, 图像特征向量视为观测变量. 假设当前时刻场景的概率只与前一时刻的场景有关, 即前后图像所属的场景具有一阶马尔可夫性,  $P(Q_t|Q_{1:t-1}) = P(Q_t|Q_{t-1})$ , 则通过 HMM 建模可得到如下的场景识别公式:

$$P(Q_t = q|\mathbf{x}_{1:t}) \propto P(\mathbf{x}_t|Q_t = q)P(Q_t = q|\mathbf{x}_{1:t-1}) = P(\mathbf{x}_t|Q_t = q) \sum_{q'} A(q', q)P(Q_{t-1} = q'|\mathbf{x}_{1:t-1}). \quad (1)$$

其中:  $A(q', q) = P(Q_t = q|Q_{t-1} = q')$  为 HMM 的转移概率矩阵,  $P(\mathbf{x}_t|Q_t)$  为观测似然概率. 通过求使  $P(Q_t = q|\mathbf{x}_{1:t})$  取最大值的  $q$ , 可得到

$$\tilde{q} = \arg \max_q P(Q_t = q|\mathbf{x}_{1:t}). \quad (2)$$

式(2)即为图像最可能对应的场景类别.

注意到, 该模型中转移概率矩阵代表了 HMM 中状态与状态之间转移的概率. 在本文的背景下, 它代表从一种场景进入另一种场景可能发生的概率, 因此可通过统计训练图像中一类场景到另一类场景的转换次数求得该矩阵. 为了使该矩阵不出现为零的元素, 可采用文献[6]的方法使用 Dirichlet 分布对该矩阵进行平滑处理.

定义初始状态为  $P(Q_1 = q|\mathbf{x}_1) = 1/N_p$ , 则每次输入新的测试图像时都可以采用式(1)更新得到新的  $P(Q_t = q|\mathbf{x}_{1:t})$ . 可以发现, 对于  $t$  时刻更新式(1),  $P(Q_{t-1} = q'|\mathbf{x}_{1:t-1})$  已经在  $t-1$  时刻的更新中计算得到. 因此, 在实际执行中, 式(1)的更新只被执行一次, 不会给系统带来额外的负担.

### 2.1 稀疏贝叶斯判别学习

在大多数已有的工作中, 观测似然概率  $P(\mathbf{x}_t|Q_t = q)$  大都通过生成模型的方法训练得到, 如 GMM<sup>[6]</sup>, 贝叶斯网络<sup>[7]</sup>和朴素贝叶斯分类器<sup>[9]</sup>等. 但是, 在场景识别问题中, 由于样本数量的限制以及计算效率和实际应用中表现出的泛化性能等, 判别模型分类算法往往能比生成模型方法取得更好的分类效果. 由于大多数判别模型学习方法本质上不能提供具有概率意义上的预测, 为了将判别模型的各种优点融入基于上下文信息的识别系统, 本文将通过引入一种稀疏贝叶斯模型——相关向量机<sup>[10]</sup>来达到以判别模型方式对上下文场景识别系统进行学习的目的.

相关向量机是一种基于核函数的学习机, 它通过将模型参数的多数元素设置为零来控制模型的复杂度, 从而达到避免过学习, 提高模型泛化能力的目的. 对于式(1)表达的上下文模型,  $N_q$  类场景的识别问题有  $N_q$  个似然概率  $P(\mathbf{x}|Q = q)$  需要计算, 因此本文首先采用了“一对多”策略对模型进行训练, 即对于每一个场景类别, 都分别训练一个两分类相关向量机, 用于求解对应的似然概率.

针对每一个两分类相关向量机模型, 假设训练集包含  $N$  个样本  $\{\mathbf{x}_n, Q_n\}$ .  $\mathbf{x}$  为场景的特征描述向量,  $Q_n \in \{1, 0\}$  则分别代表样本是否属于类别  $q$ . 相关向量机首先使用 S 形函数(sigmoid)对样本特征  $\mathbf{x}$  属于场景  $q$  的后验概率进行建模, 有

$$P(Q = 1|\mathbf{x}) = \sigma(f(\mathbf{x})) = \frac{1}{1 + e^{-f(\mathbf{x})}}, \quad (3)$$

其中  $f(\mathbf{x})$  就是相关向量机的判别函数. 类似所有的稀疏判别模型, 其定义为

$$f(\mathbf{x}) = f(\mathbf{x}; \mathbf{w}) =$$

$$\sum_{i=1}^N w_i K(\mathbf{x}, \mathbf{x}_i) + w_0 = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}). \quad (4)$$

其中:  $K(\mathbf{x}, \mathbf{x}_i)$  为核函数;  $N$  为训练样本的个数, 同时也是核函数的个数;  $\mathbf{w} = (w_0, \dots, w_N)^T$  为相关向量机的权值向量;  $\boldsymbol{\phi}(\mathbf{x}) = [1, K(\mathbf{x}, \mathbf{x}_1), \dots, K(\mathbf{x}, \mathbf{x}_N)]^T$ .

注意到, 该定义中权值参数的个数与训练样本的个数相同, 如果直接通过求训练样本的似然函数最大化来估计权值向量  $\mathbf{w}$ , 则模型将出现严重的过学习. 为了防止这种情况, 相关向量机模型引入了一组超参数  $\{\alpha_i\}_{i=0}^N$ , 用于对权值参数的条件概率分布进行约束. 相关向量机设定  $w_i$  满足 0 均值、方差为  $\alpha_i^{-1}$  的高斯分布, 即

$$P(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=0}^N N(w_i|0, \alpha_i^{-1}), \quad (5)$$

其中  $\boldsymbol{\alpha}$  为  $N+1$  维的超参数向量, 并令  $\alpha_i$  满足 Gamma 分布. 此约束产生的效果是, 最终得到的模型中大部分权值参数  $w_i$  将趋于零. 若将小于一定阈值的权值参数设为零, 则该权值在判别式  $f(\mathbf{x})$  中将不起作用, 因此只需计算那些权值参数不等于 0 的部分即可, 从而实现模型的稀疏性. 可以说, 为每一个权值参数设置独立的超参数可使相关向量机模型具有稀疏特性, 从而防止过学习.

通过“一对多”学习, 系统将得到  $N_q$  个形如式 (3) 的概率判别式, 分别对应图像属于某类场景的后验概率. 注意到, 尽管具有概率输出的特点, 相关向量机仍是一种判别模型, 它直接建模的对象为后验概率  $P(Q|\mathbf{x})$ , 而非似然概率  $P(\mathbf{x}|Q)$ . 但是, 通过贝叶斯原理  $P(Q_t = q|\mathbf{x}_t) \propto P(\mathbf{x}_t|Q_t = q)P(Q_t = q)$ , 可将式 (1) 表示的上下文模型进一步扩展为

$$\begin{aligned} P(Q_t = q|\mathbf{x}_{1:t}) &\propto \\ P(Q_t = q|\mathbf{x}_t)P(Q_t = q)^{-1} &\cdot \\ \sum_{q'} A(q', q)P(Q_{t-1} = q'|\mathbf{x}_{1:t-1}). &\quad (6) \end{aligned}$$

其中:  $P(Q_t = q)$  为图像属于场景  $q$  的先验概率, 此概率可通过计算该场景在训练样本中所占的比例求得; 而  $P(Q_t = q|\mathbf{x}_t)$  则可由对应该场景的相关向量机模型进行预测计算得到. 这样便实现了通过判别模型的学习来得到 HMM 上下文模型的目的. 对于任意一幅新获取的场景图像, 只要将其图像描述特征向量  $\mathbf{x}_t$  代入式 (6) 进行计算, 然后通过式 (2) 便可得到该幅图像所属的场景类别.

## 2.2 参数学习

下面进一步说明本文对相关向量机模型进行训练的过程. 设  $\mathbf{Q} = (Q_1 \dots Q_N)^T$  为训练样本标签组成的向量, 因为训练的是二分类分类器, 所以设  $Q_n \in \{0, 1\}$ . 相关向量机学习的目标是, 根据训练样本求解

在约束条件 (5) 下模型判别函数 (4) 的权值参数  $\mathbf{w}$ . 由于参数  $\mathbf{w}$  的后验概率密度  $P(\mathbf{w}|\mathbf{Q}, \boldsymbol{\alpha})$  无法通过直接积分求得解析解, 本文采用文献 [10] 中的 Laplace 近似方法来迭代求解模型的参数, 得到的更新方程为

$$\bar{\mathbf{w}} = \mathbf{A}^{-1} \boldsymbol{\Phi}^T (\mathbf{Q} - \mathbf{y}), \quad (7)$$

$$\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A})^{-1}, \quad (8)$$

$$\alpha_i^{\text{new}} = \frac{\rho_i}{(\bar{w}_i)^2}. \quad (9)$$

其中:  $\mathbf{y} = (y_1, \dots, y_N)^T$ ,  $y_n = \sigma\{f(\mathbf{x}_n; \mathbf{w})\}$ ;  $\rho_i = 1 - \alpha_i \Sigma_{ii}$ ;  $\mathbf{A} = \text{diag}(\alpha_i)$ ;  $\mathbf{B}$  是一个  $N \times N$  的对角阵, 对角元素为  $b_n = y_n(1 - y_n)$ ; 而  $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\mathbf{x}_1), \dots, \boldsymbol{\phi}(\mathbf{x}_N)]^T$  是  $N \times (N+1)$  维的设计矩阵.

训练过程只要在初始化各参数后, 通过式 (7) ~ (9) 便可迭代求解  $\boldsymbol{\alpha}$  和  $\bar{\mathbf{w}}$  的值, 直到模型满足收敛条件. 最后,  $\bar{\mathbf{w}}$  的终值就是期望求得的判别式 (4) 中的权值参数向量  $\mathbf{w}$ , 其中非零值对应的训练样本特征向量就是模型的相关向量. 由相关向量机的稀疏性原理可知, 式 (4) 判别模型中将只有少量的  $\mathbf{w}$  非零, 因此能达到非常高的计算效率和泛化性能.

## 3 场景图像特征与描述

对于场景识别问题, 如何获取图像描述向量  $\mathbf{x}$  同样是一个重要问题. 本文首先采用 SIFT 描述子<sup>[11]</sup>对图像局部区域特征进行提取. 对于一幅场景图像, 首先使用密集采样 SIFT 对整幅图片进行特征提取, 然后构建“视觉词典”<sup>[12]</sup>, 也就是采用  $k$  means 聚类方法对采集的 SIFT 特征进行聚类. 这样, 每一个 SIFT 特征都将量化为一个“视觉单词”. 若  $K$  为  $k$  means 的聚类数, 则任何一幅场景图像都可以用一个  $K$  维向量来表示“视觉单词”在一幅图像中出现的频率, 即形成关于“视觉单词”的直方图. 为进一步对图像特征在图像上的空间分布进行建模, 本文采用 Lazebnik 等<sup>[13]</sup>提出的空间金字塔匹配 (SPM) 模型对图像进行表示. 每张场景图像都按不同层次进行了分割, 图像在第  $l$  层被等分成  $2^l \times 2^l$  个区块, 然后对  $l$  层的每一个区块分别建立“视觉单词”直方图, 按着顺序连接便构成了这一层的图像描述. 图 1 所示为本文使用的三层图像描述. 再将不同层次的直方图向量按一定的权值<sup>[13]</sup>串连起来并规一化, 便可得到全图的描述向量.

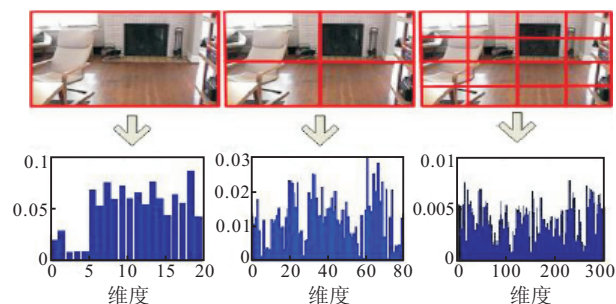


图 1 图像的空间金字塔直方图表示

## 4 实验

本文实验在VPC家庭环境图像库<sup>[9]</sup>上进行, 该数据库通过手持摄像机在6户真实家庭中采集了11类连续场景图像. 由于各个家庭在装修风格、年代、面积和布局上各不相同, 因此该数据库是一个非常具有挑战性的场景分类库. 然而, 因为不是每户家庭都存在完整的11类场景, 所以本文采用文献[9]中的建议, 仅选择每户家庭都具备的5类场景进行实验, 即卧室、浴室、厨房、起居室和餐厅. 图2所示为其中一户家庭中5类场景的示例图像.

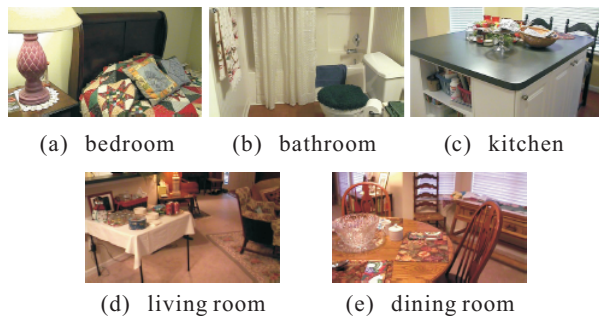


图2 VPC家庭环境图像库中的5类场景图像示例

由于该数据库中图片的像素较高, 为减少图像特征提取所花费的时间, 本文采用的训练和测试图片均先预处理为宽度为400像素的灰度图像. 图像特征提取采用步长为10个像素的密集采样SIFT特征的方式. 为了考虑多个尺度上的信息, 每个采样点分别在4, 8, 12和16这4个半径尺度上提取SIFT描述子. “视觉词典”生成使用 $k$  means聚类, 词典的大小即聚类的 $K$ 值统一设置为600.

实验过程如下: 采用文献[9]中类似“留一法”交叉检验的方式对VPC数据库进行实验, 即首先使用其中5户家庭的场景图像作为训练集, 剩下的1户家庭场景图像作为测试集; 然后逐个地对每一户家庭进行测试, 以验证由不同环境训练得到的模型对新环境的识别能力. 识别结果用平均识别率(%)表示, 即正确识别的场景个数除以测试图像总数. 此外, 为进一步验证分类算法本身对场景识别的优劣, 实际记录了模型在选择“不使用上下文信息”和“使用上下文信息”两种情况下的识别效果, 其中不使用上下文信息时只需将HMM中的转移概率矩阵设置为均一值即可(设置 $A(i, j) = 1/N_p$ ).

为了与以往的基于生成模型的上下文场景识别方法进行比较, 本文同时实现了以下两种基准算法, 并进行了实验测试.

1) 使用SIFT的空间金字塔模型对场景进行图像描述, 然后使用高斯混合模型(GMM)对场景的特征描述向量进行建模, 模型中高斯组份的均值和混合权重通过EM算法求得<sup>[6]</sup>. 在测试阶段, 将图像特征向量

代入学习得到的高斯混合模型便可得到相应的观测似然概率.

2) 使用SIFT的空间金字塔模型对场景进行图像描述, 然后使用朴素贝叶斯分类器训练分类模型, 最后与HMM上下文模型相结合.

另外, 将本文的实验结果与文献[9]中公布的结果进行了对比, 对比结果分别如表1和表2所示. 文献[9]采用的也是基于朴素贝叶斯分类器的生成模型, 但其在实验中使用了与本文不同的图像特征描述方法, 即CENTRIST特征<sup>[9]</sup>.

表1 VPC的平均识别率, 模型不使用上下文信息

测试集	平均识别率/%			
	GMM	naive Bayes	文献[9]	本文算法
home 1	39.66	38.42	44.77	40.92
home 2	29.87	31.79	33.33	36.76
home 3	35.22	37.62	40.68	55.91
home 4	46.22	42.84	43.28	52.76
home 5	40.03	40.33	41.10	42.42
home 6	46.37	44.26	48.07	50.46
average	39.57	39.21	41.87	46.44

表1 对VPC的平均识别率, 模型使用上下文信息

测试集	平均识别率/%			
	GMM	naive Bayes	文献[9]	本文算法
home 1	42.32	41.73	44.58	43.21
home 2	32.63	35.51	35.89	39.66
home 3	40.29	41.67	40.96	59.08
home 4	47.67	44.28	49.93	53.22
home 5	45.33	45.55	46.91	47.62
home 6	50.74	48.38	55.46	55.73
average	43.16	42.85	45.62	49.75

注意到, 因为本文实验的测试样本和训练样本是在完全不同的家庭环境中分别采集的, 所以对于场景识别系统而言具有非常高的挑战性. 从表1可以看到, 在不考虑场景上下文关系的情况下, 本文识别算法的平均识别率为46.44%. 结果表明: 首先, 本文模型的平均识别率高于基准生成模型的识别率, 也好于文献[9]的实验结果; 其次, 虽然测试家庭对于机器人而言是完全陌生的环境, 但通过一定量的训练, 本文模型仍能有效地用于新环境的场景识别. 这意味着本文方法具有很好的泛化特性, 而泛化特性则表征了在有限的数据下训练出来的模型是否具有能够正确识别新数据的能力.

得到上述实验结果的主要原因在于本文算法同时具有判别模型和稀疏模型的优点. 一方面, 对于一般生成模型的方法, 需对图像特征的观测似然概率进行建模, 这在许多实际应用中, 训练样本在数量上的限制使得它们无法很好地表征研究对象真实的特征分布情况, 从而也限制了生成模型建模的准确性. 然而, 通过判别学习, 直接求对样本的期望误差或代价

函数最小化,往往能在实际应用中取得更高的准确度.另一方面,过多参数的分类模型往往会过分地拟合训练样本,使得对新样本的预测识别率大大降低.从2.1节可以看出,本文方法通过对模型参数 $w_i$ 设定零均值高斯分布约束,使得学习得到的大部分模型参数趋于零,进而在整个模型中不起作用.这样,可通过只保留小部分的模型参数来控制模型的复杂度,从而避免过学习,起到提高泛化性能的作用.

从表2的结果可以看出,当模型考虑了场景之间的上下文关系后,平均识别率又有了进一步的提升.这说明,一方面引入场景的上下文关系能够提高机器人对环境的认识;另一方面,本文采用了泛化能力更好的相关向量机判别模型对上下文模型进行学习,从而可以更好地结合二者的优势,实现更高的场景识别率.值得注意的是,由于相关向量机模型的高稀疏性,每次识别过程中只需考虑模型中很少量的相关向量,计算效率很高,非常适合机器人的实时识别.经实验测试,本文实验训练出的模型相关向量数量在250~350之间,若去除特征提取的耗时,则识别单张场景图像的时间不大于0.01s.

## 5 结 论

场景的上下文信息对于移动机器人进行场景识别具有指导意义.一般的上下文场景识别系统大多通过隐马尔可夫建模结合生成模型的方法进行训练,本文则通过引入一种稀疏贝叶斯模型——相关向量机,使用判别模型的训练方法得到上下文识别模型.在真实场景数据库上的实验结果表明,采用本文方法训练得到的上下文模型能够获得比传统生成模型方法更好的识别率和泛化性能,并具有很高的计算效率.

### 参考文献(References)

- [1] Lategahn H, Geiger A, Kitt B. Visual SLAM for autonomous ground vehicles[C]. IEEE Int Conf on Robotics and Automation (ICRA). Shanghai: IEEE, 2011: 1732-1737.
- [2] Fei-Fei L, Perona P. A Bayesian hierarchical model for learning natural scene categories[C]. IEEE Conf on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005: 524-531.
- [3] Bosch A, Zisserman A, Muoz X. Scene classification using a hybrid generative/discriminative approach[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712-727.
- [4] Quattoni A, Torralba A. Recognizing indoor scenes[C]. IEEE Conf on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 413-420.
- [5] Brown M, Susstrunk S. Multi-spectral SIFT for scene category recognition[C]. IEEE Conf on Computer Vision and Pattern Recognition(CVPR). Colorado Springs: IEEE, 2011: 177-184.
- [6] Torralba A, Murphy K P, Freeman W T, et al. Context-based vision system for place and object recognition[C]. IEEE 9th Int Conf on Computer Vision. Nice: IEEE, 2003: 273-280.
- [7] Im S, Cho S. Context-based scene recognition using Bayesian networks with scale-invariant feature transform[C]. The 8th Int Conf on Advanced Concepts for Intelligent Vision Systems. Antwerp: Springer Verlag, 2006: 1080-1087.
- [8] Ando R, Shinoda K, Furui S, et al. Robust scene recognition using language models for scene contexts[C]. The 8th ACM Multimedia Int Workshop on Multimedia Information Retrieval. Santa Barbara: ACM, 2006: 99-106.
- [9] Wu J, Christensen H I, Rehg J M. Visual place categorization: Problem, dataset, and algorithm[C]. IEEE Int Conf on Intelligent Robots and Systems. St Louis: IEEE, 2009: 4763-4770.
- [10] Tipping M E. Sparse Bayesian learning and the relevance vector machine[J]. J of Machine Learning Research, 2001, 1(3): 211-244.
- [11] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. Int J of Computer Vision, 2004, 60(2): 91-110.
- [12] Sivic J, Zisserman A. Video google: A text retrieval approach to object matching in videos[C]. IEEE 9th Int Conf on Computer Vision(ICCV). Nice: IEEE, 2003: 1470-1477.
- [13] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]. IEEE Conf on Computer Vision and Pattern Recognition. New York: IEEE, 2006: 2169-2178.