

Journal of Image  
and Graphics

# 中国图象图形学报



ISSN1006-8961  
CN11-3758/TB

2012 7  
Vol.17 No.

中国科学院遥感应用研究所  
中国图象图形学学会主办  
北京应用物理与计算数学研究所

# 中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2012年7月 第17卷 第7期(总第195期)

## 目次

### 综述

中国多媒体技术研究: 2011 ..... 史元春, 徐光祐, 高原(741)

### 图像处理和编码

图像引导滤波的局部多尺度 Retinex 算法 ..... 方帅, 杨静荣, 曹洋, 武鹏飞, 饶瑞中(748)

基于第一主成分方向稳定性的图像零水印 ..... 付剑晶, 王珂(756)

小波变换估计非线性扩散最优停止时间 ..... 蒋平, 张建州(770)

### 图像分析和识别

改进的 Beamlet 与 Canny 相结合提取复杂图像线特征 ..... 曾接贤, 周沥沥, 符祥(775)

动静态信息融合及动态贝叶斯网络的步态识别 ..... 杨旗, 薛定宇(783)

融合音频单词与视觉特征的成人视频检测 ..... 刘毅志, 唐胜, 王向东, 林守勋, 张勇东(791)

基于平面区域跟踪的目标位姿参数自动测量 ..... 回丙伟, 文贡坚, 赵竹新, 钟金荣(798)

3 维图像中边界曲面的分类追踪及抽取 ..... 丁德福, 程柳航, 王利生(806)

局部时空域模型的核密度估计目标检测方法 ..... 王兴宝, 刘纯平, 费兰英, 王朝晖, 季怡(813)

基于相位谱和调谐幅度谱的显著性检测方法 ..... 李崇飞, 高颖慧, 卢凯, 曲智国(821)

### 图像理解和计算机视觉

“目标-场景”语境关联的生成图模型分析 ..... 谢昭, 李姍琦, 高隼(828)

融合上下文信息的场景结构恢复 ..... 武晖, 于昕, 隋尧, 张利(839)

带有局部控制因子的图割光流估计 ..... 路子贇, 唐土生, 高隼, 沈琳, 刘伟(846)

改进的 Harris 亚像素角点快速定位 ..... 何海清, 黄声享(853)

### 计算机图形学

应用最小生成树实现点云分割 ..... 孙金虎, 周来水, 安鲁陵(858)

### 医学图像处理

病毒进化的离散差分进化超声图像特征选择 ..... 张巧荣, 朱长明, 倪军, 刘海波(866)

分水岭优化的 Snake 模型肝脏图像分割 ..... 兰红, 张璐(873)

### 遥感图像处理

Normalized Cut 与分水岭变换在高光谱影像混合像元端元提取中的应用 ..... 许菡, 李小娟(880)

改进非局部均值滤波的 SAR 图像降噪方法 ..... 郑永恒, 程建, 曹宗杰(886)

~~~~~  
第四届国际遥感考古会议 ..... (892)

## 中国图象图形学报

刊名题字: 宋 健

月刊(1996 年创刊)

第 17 卷 第 7 期

2012 年 7 月 16 日出版

**主管单位** 中国科学院

**主 办** 中国科学院遥感应用研究所  
中国图象图形学学会

北京应用物理与计算数学研究所

**主 编** 李小文

**编辑出版** 《中国图象图形学报》编辑出版委员会

北京 9718 信箱 邮编 100101

电子信箱:jig@irsa.ac.cn

电话:010-64807995 010-82614429

网 址:www.cjig.cn

**印刷装订** 北京北林印刷厂

**广告经营许可证** 京朝工商广字第 0346 号

**总 发 行** 北京报刊发行局

**订 购** 全国各地邮局

**国外发行** 中国国际图书贸易总公司

(中国国际书店)

(北京 399 信箱 邮编 100044)

**Superintended by** Chinese Academy of Sciences

**Sponsored by** Institute of Remote Sensing Application,  
CAS China Society of Image and Graphics  
Institute of Applied Physics and Computational  
Mathematics

**Chief editor** LI Xiaowen

**Editor, Publisher** Editorial and Publishing Board  
of Journal of Image and Graphics  
(P. O. Box 9718, Beijing 100101, China)  
E-mail:jig@irsa.ac.cn

**Distributed by** Beijing Bureau for Distribution of Newspapers  
and Journals

**Domestic** All Local Post Offices in China

**Foreign** China International Book Trading Corporation  
(P. O. Box 399, Beijing 100044, China)

**Printed by** Beijing Beilin Printing House

ISSN 1006-8961 CN11-3758/TB CODE ZTTFXZ 国内邮发代号: 82-831 国外发行代号: M1406 国内定价: 45.00 元

# Journal of Image and Graphics

( Monthly , Started in 1996 )

Vol. 17 No. 7 July 2012

## Contents

### Review

Researches on multimedia technology in China, 2011 ..... Shi Yuanchun, Xu Guangyou, Gao Yuan (741)

### Image Processing and Coding

Local multi-scale Retinex algorithm based on guided image filtering  
..... Fang Shuai, Yang Jingrong, Cao Yang, Wu Pengfei, Rao Ruizhong (748)

Image zero-watermark based on direction stability of first principal component vector ..... Fu Jianjing, Wang Ke (756)

Stopping-time estimation for anisotropic diffusion using discrete wavelet transform ..... Jiang Ping, Zhang Jianzhou (770)

### Image Analysis and Recognition

Complex image line feature extraction based on improved Beamlet transform and the Canny operator  
..... Zeng Jiexian, Zhou Lili, Fu Xiang (775)

Gait recognition based on dynamic & static information fusion and dynamic bayesian network ..... Yang Qi, Xue Dingyu (783)

Fusing audio-words with visual features for adult video detection  
..... Liu Yizhi, Tang Sheng, Wang Xiangdong, Lin Shouxun, Zhang Yongdong (791)

Automatic measurement for an object's position and attitude via tracking planar regions  
..... Hui Bingwei, Wen Gongjian, Zhao Zhuxin, Zhong Jinrong (798)

Detection and extraction of boundary surface patches within 3D images ..... Ding Defu, Cheng Liuhang, Wang Lisheng (806)

Foreground object detection method using kernel density estimation of a local spatio-temporal model  
..... Wang Xingbao, Liu Chunping, Fei Lanying, Wang Zhaohui, Ji Yi (813)

Saliency detection method based on phase spectrum and amplitude spectrum tuning  
..... Li Chongfei, Gao Yinghui, Lu Kai, Qu Zhiguo (821)

### Image Understanding and Computer Vision

"Object-Scene" contextual associated generative graph model analysis ..... Xie Zhao, Li Shanqi, Gao Jun (828)

Structure recovery algorithm using contextual information ..... Wu Hui, Yu Xin, Sui Yao, Zhang Li (839)

Graph cut optical flow estimation with a local control factor ..... Lu Ziyun, Tang Tusheng, Gao Jun, Shen Lin, Liu Wei (846)

Improved algorithm for Harris rapid sub-pixel corners detection ..... He Haiqing, Huang Shengxiang (853)

### Computer Graphics

Research on point cloud segmentation using a minimum spanning tree ..... Sun Jinhu, Zhou Laishui, An Luling (858)

### Medical Image Processing

Virus-evolutionary discrete differential evolution algorithm for feature selection of cervical lymph nodes in ultrasound images  
..... Zhang Qiaorong, Zhu Changming, Ni Jun, Liu Haibo (866)

Liver image segmentation algorithm based on the Snake model And optimized by watershed transformation  
..... Lan Hong, Zhang Lu (873)

### Remote Sensing Image Processing

Endmember extraction for hyperspectral image based on normalized cut and watershed transformation  
..... Xu Han, Li Xiaojuan (880)

SAR image denoising via improved non-local means filter ..... Zheng Yongheng, Cheng Jian, Cao Zongjie (886)

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2012)07-0828-11

论文引用格式: 谢昭, 李姗姗, 高隽. “目标-场景”语境关联的生成图模型分析[J]. 中国图象图形学报, 2012, 17(7): 828-838.

## “目标-场景”语境关联的生成图模型分析

谢昭, 李姗姗, 高隽

合肥工业大学计算机与信息学院, 合肥 230009

**摘要:** 图像场景分析是目前计算机视觉领域的研究热点, 体现了场景与目标之间的包含关系。在分析过程中合理的使用基于上下文关系的知识可以提高场景分析模型的适用性和目标识别的准确率。从“图像集——场景——目标——部分——视觉词汇”这种层次的角度进行场景分析, 将全局上下文信息和局部上下文信息同时融入到基于 HDP 的生成图模型中, 在场景层和目标层这两个不同的层次上, 共同作用于场景分析。场景分析的结果可以用来约束目标识别, 目标识别的结果可以反馈作用于场景分析。

**关键词:** 场景分析; 生成图模型; 上下文信息; 反馈机制; 目标识别

### “Object-Scene” contextual associated generative graph model analysis

Xie Zhao, Li Shanqi, Gao Jun

School of Computer & Information, HeFei UniVersity of Technology, Hefei 230009, China

**Abstract:** Currently, image analysis is the central issue in the field of computer vision. It reflects the inclusion relationships between the scene and the object. In the process of scene analysis, properly used contextual knowledge can improve the applicability of the scene analysis model and the accuracy of the object recognition. We do our research on the scene analysis from a hierarchical perspective of “database-scene-object-part-visual words”, and we add global contextual information and local contextual information into a generative graph model based on a hierarchical dirichlet process, analyzing the scene at the scene-level and the object-level at the same time. In this way, the result of the scene analysis can be used to constrain object recognition, while the object recognition result feedback effect on the scene analysis.

**Key words:** scene analysis; generative graph model; contextual information; feedback mechanism; object recognition

## 0 引言

场景分析是图像理解的重要研究内容, 侧重于对图像场景整体的认知和理解, 主要强调场景特征的检测与描述, 概率推理模型的建立, 场景对目标分析的约束机制等。在图像理解领域, 目标识别和场景分析近年来都取得了许多进展<sup>[1-5]</sup>, 但关于建立目标识别和场景分析之间反馈过程方面的研究却不

多。而在现实世界中, 目标的出现都不是相互孤立的, 且与其周围环境、场景有着密切的联系。场景中的目标识别侧重对局部区域的理解, 场景分析侧重对图像景物的整体描述与理解, 两者相互影响、共同作用。

基于 HDP(hierarchical dirichlet process) 图结构模型<sup>[6]</sup>, 结合层次思想, 运用 Gibbs Rao-Blackwellized 采样方法, 构建一种场景分析生成图模型, 用于建立图像场景分析与目标识别间的反馈机制。

收稿日期: 2011-07-11; 修回日期: 2012-03-13

基金项目: 国家自然科学基金项目(60905005, 60875012)

第一作者简介: 谢昭(1980—), 男, 副研究员, 2007年7月获合肥工业大学计算机应用技术专业博士学位, 主要研究方向为计算机视觉、图像理解。Email: xiezhao1980@126.com

具体实现思想是在生成图模型中融入场景级全局知识和目标级局部知识。将场景看成一个整体,通过对场景全局特征的感知进行场景分析,场景分析结果可以用来指导目标识别过程;然后将场景看成目标的组合分布,通过目标与场景之间以及目标与目标之间的先验信息结合局部知识对场景分析结果进行验证。

### 1 生成图模型下的场景表达及学习推理

借鉴 HDP 图模型的思想,从“图像集—场景—目标—部分—视觉词汇”这种层次的角度,建立视觉特征,目标类别与场景类别之间的概率关系,生成不同目标类别和场景类别信息的参数化表达,并通过迭代进行模型的参数化学习<sup>[7-8]</sup>。

构建理想的生成图模型不仅可以有效地分析场景,而且场景分析的结果可作为全局知识为目标识别提供有效的指导。

#### 1.1 基于 HDP 的生成图模型

场景分析生成图模型如图 1 所示。图像数据集中一般包含若干类别的场景图像,如办公室场景、街道场景等。令场景类别集合  $s = \{s_1, s_2, \dots, s_J\}$ ,  $J$  为场景类别个数,每幅场景图像都属于一种场景类别,不同类别的场景图像所包含的目标类别也不相同。如图 2 中,办公室场景中含有计算机、键盘等目标类别,街道场景中有公路、汽车等目标类别。

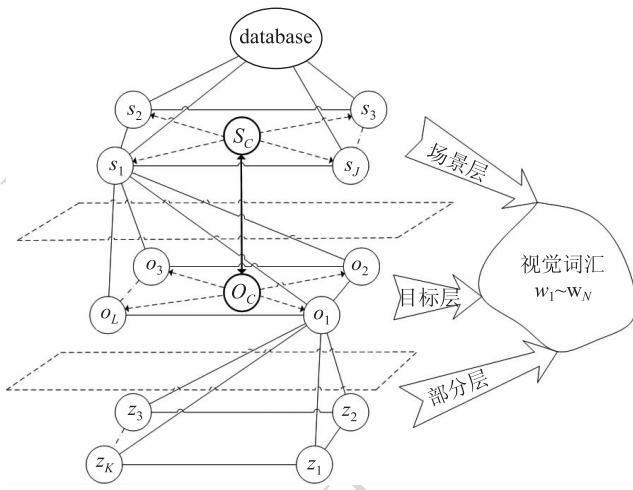


图 1 场景分析生成图模型

Fig. 1 Generative graph model for scene analysis

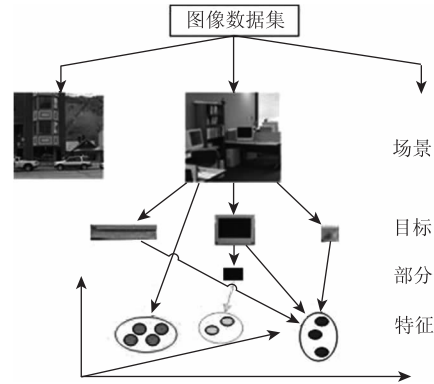


图 2 场景分析生成图模型参数对应关系示意图

Fig. 2 Visualization representation of generative scene analysis graph model's parameters

目标类别集合  $o = \{o_1, o_2, \dots, o_L\}$ ,  $L$  为目标类别个数。某目标类别中的每个目标实例可视为若干个部分  $z$  (parts) 组成的集合,由于直接统计目标的特征计算量非常庞大,这里引入部分的概念<sup>[8]</sup>。部分由特征  $w$  聚类形成,相似的特征可根据聚类算法集合为一个部分。对从图像中提取出的特征进行  $K$ -means 聚类,训练出一个具有鲁棒性和代表性的视觉词汇库。有了这个视觉词汇库后就可通过一幅图像中提取出的特征来判断该图像属于哪类场景,其中包含哪些目标种类。

图 1 中  $s, o, z$  分别表示场景、目标以及部分标记,  $S_c$  表示场景间的全局知识,  $O_c$  表示目标间的局部知识。  $K$  表示某类目标实例的部分个数(部分个数可由 stick-breaking 过程计算),  $N$  表示图像集中视觉词汇总数。该生成图模型每一层的概率分布符合 Dirichlet 分布,层次性符合 HDP。

#### 1.2 生成图模型的学习推理

构建生成图模型后,完成场景分析任务即是为了保证场景  $s$  和目标  $o$  关于特征变量  $w$  的后验概率  $p(s, o | w)$  为最大值。通过 Bayesian 公式,将后验概率转换成似然概率和先验概率的乘积:  $p(s, o | w) \propto p(w | s, o)p(s, o)$ , 其中  $p(w | s, o)$  表示通过学习得到的观测变量  $s, o$  关于视觉词汇  $w$  的概率分布。

下面从目标级和场景级两个层次详细说明该生成图模型如何建立视觉词汇与目标间的映射关系,视觉词汇与场景间的映射关系以及目标与场景间的映射关系。

##### 1.2.1 生成图模型下目标级学习推理

HDP 图模型节点间包含了分层的思想:观测节点、隐节点、最终标记节点。将其应用于目标识别,

图像中目标分层的思想为：“目标—部分—特征”，即“ $o_j \rightarrow z_j \rightarrow N_j$ ”。

生成图模型下的目标表达如图 3 所示。其中灰色节点表示观测变量；白色节点表示隐变量；矩形方框表示对其中包含的变量进行迭代采样，右下角的参数表示其迭代次数。

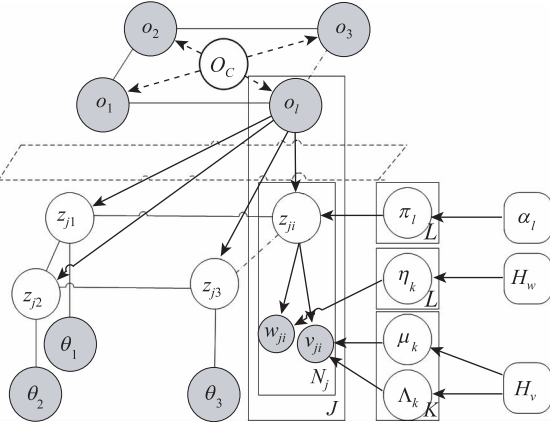


图 3 生成图模型下的目标表达

Fig. 3 Generative graphical representation for object recognition

部分  $z$  是图像中外观特征和位置特征的聚类，其中  $w_j$  表示外观特征， $w_j = (w_{j1}, \dots, w_{jN_j})$ ， $v_j$  表示位置特征  $v_j = (v_{j1}, \dots, v_{jN_j})$ ；隐变量  $z_{ji}$  表示观测特征变量 ( $w_j, v_j$ ) 属于某一部分的索引，整幅图像所有特征对应索引向量为  $z_j = (z_{j1}, z_{j2}, \dots, z_{jN_j})$ ；部分参数  $\theta_l = (\eta_k, \mu_k, \Lambda_k)$ ，其超参（即初始分布）为  $\lambda_l$ 。每个部分的外观特征  $w_{ji}$  服从多项式分布  $\eta_k \in \Pi_{W-1}$ ，其超参为  $H_w$ ；每个部分的位置特征  $v_{ji}$  服从 Normal-inverse-Wishart 分布  $D(\mu_k, \Lambda_k)$ ，其超参为  $H_v$ ；每个部分  $z_{ji}$  服从 Dirichlet 分布  $\pi_l$ ，其超参为  $\alpha_l$ 。

Rao-Blackwellized Gibbs 采样可用于对部分参数  $\theta_l = (\eta_k, \mu_k, \Lambda_k)$  进行联合概率采样。用  $z_{ji}$  表示除  $z_{ji}$  之外的图像中所有部分特征索引， $w_{ji}$  表示除  $w_{ji}$  之外所有外观特征集合， $v_{ji}$  含义类似。分别对部分的外观参数和位置参数进行迭代采样，在每次迭代过程中，固定  $z_{ji}$ ，然后对  $z_{ji}$  进行索引采样。根据目标级模型的 Markov 特性，部分特征索引的后验概率分布为

$$P(z_{ji} | z_{\setminus ji}, w, v, \theta) \propto P(z_{ji} | o_j) P(w_{ji} | z_j, w_{\setminus ji}) P(v_{ji} | z_j, v_{\setminus ji}) \quad (1)$$

用  $C_{kw}^{-i}$  表示除第  $i$  个特征之外部分  $k$  中所有外观特征描述子为  $w$  的特征个数； $N_{lk}^{-i}$  表示除第  $i$  个特征之外部分  $k$  的总特征个数；这里对特征个数的统计是

针对图像数据集中目标类别为  $l$  的所有图像。用  $w$  表示图像数据集中的某个预定义的视觉词汇， $k$  表示图像数据集中所有目标类别的部分个数， $W$  表示图像数据集中的特征词汇总数； $\bar{\mu}_k, \bar{\Lambda}_k$  分别表示部分  $k$  中除第  $i$  个特征之外，所有特征正规化矩匹配 (RMM) 的位置平均值和方差值。式 (1) 右边的每一项均可采用 Dirichlet 先验标准表达，即

$$P(z_{ji} = k | z_{\setminus ji}, o_j = l) = \frac{N_{lk}^{-i} + \alpha_l / K}{\sum_{k'} N_{lk'}^{-i} + \alpha_l} \quad (2)$$

$$P(w_{ji} = w | z_{ji} = k, z_{\setminus ji}, w_{\setminus ji}) = \frac{C_{kw}^{-i} + \lambda_l / W}{\sum_{w'} C_{kw'}^{-i} + \lambda_l} \quad (3)$$

$$P(v_{ji} | z_{ji} = k, z_{\setminus ji}, v_{\setminus ji}) = P(v_{ji} | \{v_{j'i'} | z_{j'i'} = k, (j', i') \neq (j, i)\}) \approx D(v_{ji}; \bar{\mu}_k, \bar{\Lambda}_k) \quad (4)$$

基于 Rao-Blackwellized Gibbs 的目标部分采样算法（见 2.2 节）最终可计算得到某种目标类别下部分的分布情况和某部分下的视觉词汇的分布情况。

### 1.2.2 生成图模型下场景级学习推理

场景分析的目的是对场景进行分类以及得到该场景中的目标类别。若将某一场景类别看成 1.2.1 节中的单目标类别，而将场景中的目标实例看成 1.2.1 节目标中的部分，如此一来，整幅场景就类似于一个单目标。图像中场景分层的思想为“场景—目标—特征”，即“ $s_j \rightarrow o_j \rightarrow N_j$ ”。

生成图模型下的场景表达如图 4 所示，其表达结构和学习推理过程与目标级类似。

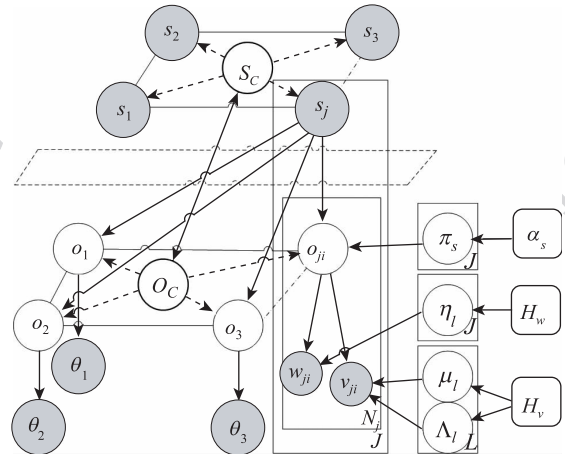


图 4 生成图模型下的场景表达

Fig. 4 Generative graphical representation for scene analysis

Rao-Blackwellized Gibbs 采样可用于对目标参数  $\theta_s = (\eta_l, \mu_l, \Lambda_l)$  进行联合概率采样。用  $\mathbf{o}_{v_i}$  表示除  $o_{v_i}$  之外的图像中所有目标特征索引,分别对目标的外观参数和位置参数进行迭代采样,在每次迭代过程中,固定  $\mathbf{o}_{v_i}$ ,然后对  $o_{v_i}$  进行索引采样。根据场景级模型的 Markov 特性,目标特征索引的后验概率分布为

$$P(o_{v_i} | \mathbf{o}_{v_i}, \mathbf{w}, \mathbf{v}, \mathbf{s}) \propto P(o_{v_i} | s_j) P(w_{v_i} | o_{v_i}, \mathbf{w}_{v_i}) P(v_{v_i} | o_{v_i}, \mathbf{v}_{v_i}) \quad (5)$$

用  $C_{lw}^{-i}$  表示除第  $i$  个特征之外目标  $l$  中所有外观特征描述子为  $w$  的特征个数;  $N_{sl}^{-i}$  表示除第  $i$  个特征之外目标  $l$  的总特征个数;这里对特征个数的统计是针对图像数据集中场景类别为  $s$  的所有图像。 $\bar{\mu}_l, \bar{\Lambda}_l$  分别表示目标  $l$  中除第  $i$  个特征之外,所有特征正规化矩匹配的位置平均值和方差值。式(5)中右边的每一项均可采用 Dirichlet 先验标准表达,即

$$P(o_{v_i} = l | \mathbf{o}_{v_i}, s_j = j) = \frac{N_{sl}^{-i} + \alpha_s / L}{\sum_{l'} N_{sl'}^{-i} + \alpha_s} \quad (6)$$

$$P(w_{v_i} = w | o_{v_i} = l, \mathbf{o}_{v_i}, \mathbf{w}_{v_i}) = \frac{C_{lw}^{-i} + \lambda_s / W}{\sum_{w'} C_{lw'}^{-i} + \lambda_s} \quad (7)$$

$$P(v_{v_i} | o_{v_i} = l, \mathbf{o}_{v_i}, \mathbf{v}_{v_i}) = P(v_{v_i} | \{v_{j'v'} | o_{j'v'} = l, (j', i') \neq (j, i)\}) \approx D(v_{v_i}; \bar{\mu}_l, \bar{\Lambda}_l) \quad (8)$$

调整目标部分采样算法,基于 Rao-Blackwellized Gibbs 的场景目标采样算法(见 2.1 节)最终可计算得到某种场景类别下目标的分布情况和某目标的视觉词汇分布情况。

## 2 生成图模型的反馈机制

场景中的目标分析侧重于对局部区域的理解,场景分析侧重于对图像景物的整体分析,是对场景内各区域之间布局关系的表示<sup>[9]</sup>。场景分析得到约束场景中目标识别的全局知识,目标识别结果得到反馈场景描述的局部知识,形成可进一步进行场景中目标识别与场景分析的先验信息。目标识别与场景分析两者相互影响,共同作用,其反馈过程实现框架如图 5 所示。

图 5 中的局部知识可由目标部分采样算法的输出得到,大致过程如下:

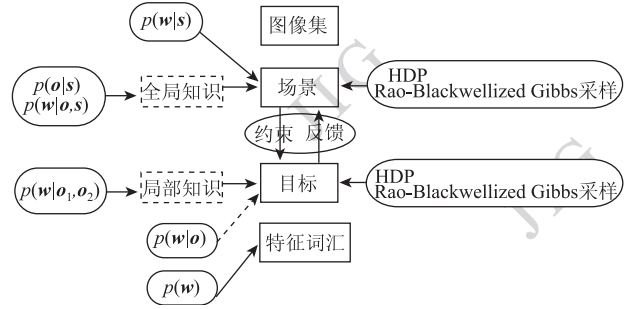


图 5 反馈过程框架图

Fig. 5 Feedback process map

目标识别的最终目的是计算图像特征向量集合属于哪一类目标的最大的后验概率  $p(\mathbf{o} | \mathbf{w})$ 。将其转换为似然概率和先验概率的乘积,由此需要计算第  $j$  幅图像中  $N_j$  个特征由某一目标类别  $o_j$  生成的似然概率:

$$P(\mathbf{w}_j, \mathbf{v}_j | o_j, \mathbf{J}) =$$

$$\int P(\mathbf{w}_j, \mathbf{v}_j | o_j, \boldsymbol{\pi}_l, \boldsymbol{\theta}_l) P(o_j, \boldsymbol{\pi}_l, \boldsymbol{\theta}_l | \mathbf{J}) d\boldsymbol{\pi}_l d\boldsymbol{\theta}_l \quad (9)$$

式中,  $\mathbf{J}$  表示同类场景的训练集图像;  $\boldsymbol{\pi}_l$  表示某种目标类别下部分的分布情况,  $\boldsymbol{\theta}_l = \{\eta_k, \mu_k, \Lambda_k\}_{k=1}^K$  表示部分特征参数。利用 Rao-Blackwellized Gibbs 采样器对特征的部分索引进行采样,假设一共采样  $T$  次,第  $t$  次采样结果为  $\mathbf{z}_j^{(t)}$ ,则似然概率的离散形式为

$$P(\mathbf{w}_j, \mathbf{v}_j | o_j, \mathbf{J}) \approx \frac{1}{T} \sum_{t=1}^T P(\mathbf{w}_j, \mathbf{v}_j | o_j = l, \boldsymbol{\pi}^{(t)}, \boldsymbol{\theta}^{(t)}) \quad (10)$$

对式(10)中所有特征进行离散化最大似然概率的计算可转化为

$$P(\mathbf{w}_j, \mathbf{v}_j | o_j = l, \boldsymbol{\pi}^{(t)}, \boldsymbol{\theta}^{(t)}) = \prod_{i=1}^{N_j} \sum_{k=1}^K \bar{\pi}_{lk} \bar{\eta}_k(w_{v_i}) D(v_{v_i}; \bar{\mu}_k, \bar{\Lambda}_k) \quad (11)$$

由此,从目标部分采样算法的输出可以得到目标类别下的视觉词汇概率分布  $p(\mathbf{w} | \mathbf{o})$ ,以及目标类别间共享的视觉词汇其概率分布  $p(\mathbf{w} | o_i, o_j)$ 。其中,目标类别间共享视觉词汇指的是不同的目标类别(如猫和狗)其部分(如它们的眼睛)在视觉特征很相似的情况下就可以使用相同的视觉词汇表示该部分<sup>[10]</sup>。

全局知识可由场景目标采样算法的输出得到。场景分析的最终目的是计算图像特征向量集合属于哪一类场景的最大的后验概率  $p(\mathbf{s} | \mathbf{w})$ 。类似地,计算第  $j$  幅图像中  $N_j$  个特征由某一场景类别  $s_j$  生成的似然概率为

$$P(\mathbf{w}_j, \mathbf{v}_j | s_j, \mathbf{J}) =$$

$$\int P(\mathbf{w}_j, \mathbf{v}_j | s_j, \boldsymbol{\pi}_s, \boldsymbol{\theta}_s) P(s_j, \boldsymbol{\pi}_s, \boldsymbol{\theta}_s | \mathbf{J}) d\boldsymbol{\pi}_s d\boldsymbol{\theta}_s \quad (12)$$

式中,  $\pi_s$  表示某种场景类别下目标的分布情况,  $\theta_s = \{\eta_l, \mu_l, \Lambda_l\}_{l=1}^L$  表示目标特征参数。利用 Rao-Blackwellized Gibbs 采样器对特征的目标索引进行采样, 假设一共采样  $T$  次, 第  $t$  次采样结果为  $\mathbf{o}_j^{(t)}$ , 则似然概率的离散形式为

$$P(\mathbf{w}_j, \mathbf{v}_j | s_j, \mathbf{J}) \approx \frac{1}{T} \sum_{t=1}^T P(\mathbf{w}_j, \mathbf{v}_j | s_j = s, \pi^{(t)}, \theta^{(t)}) \quad (13)$$

对式(13)中所有特征进行离散化最大似然概率的计算可转化为

$$P(\mathbf{w}_j, \mathbf{v}_j | s_j = s, \pi^{(t)}, \theta^{(t)}) = \prod_{i=1}^{N_j} \sum_{l=1}^L \bar{\pi}_{sl} \bar{\eta}_l(w_{ji}) D(v_{ji}; \bar{\mu}_l, \bar{\Lambda}_l) \quad (14)$$

由此, 从场景目标采样算法的输出可以得到: 场景类别下目标的概率分布  $p(\mathbf{o} | \mathbf{s})$ ; 场景类别下的视觉词汇概率分布  $p(\mathbf{w} | \mathbf{s})$ ; 场景类别和该场景类别下目标类别的视觉词汇概率分布  $p(\mathbf{w} | \mathbf{s}, \mathbf{o})$ 。至此, 由以上分析就可求出  $p(\mathbf{s}, \mathbf{o} | \mathbf{w})$  的最大值

$$p(\mathbf{s}, \mathbf{o} | \mathbf{w}) \propto \frac{p(\mathbf{w} | \mathbf{s}, \mathbf{o}) p(\mathbf{o} | \mathbf{s}) p(\mathbf{s})}{p(\mathbf{w} | \mathbf{s})} \quad (15)$$

## 2.1 场景分析约束机制下的目标识别

全局知识能够提高目标识别的准确率, 因此可用场景分析结果指导目标识别过程, 即在特定场景中某些特定目标出现的概率是大致固定的, 如果某类目标在某类场景中频繁出现, 则可认为该类目标与此类场景是相关的。

场景分析结果约束目标识别过程如下:

1) 场景级生成图模型下完成场景目标采样

输入: 当前  $j$  图像中的  $N_j$  个特征的目标索引向量为  $\mathbf{o}_j^{(t-1)}$ ;  $j$  图像中的场景类别为  $s_j = j$ ; 采样迭代次数  $T$ 。

(1) 将  $N_j$  个视觉特征进行随机排列为  $\tau(\cdot)$ ;

(2) 令  $\mathbf{o}_j = \mathbf{o}_j^{(t-1)}$ , 对每个  $i \in \{\tau(1), \dots, \tau(N_j)\}$  重新采样其索引  $o_{ji}$ ;

(3) 设定  $\mathbf{o}_j^{(t)} = \mathbf{o}_j$ , 目标的权重分布和参数分布可采样为

$$\pi_s^{(t)} \sim \text{Dir}(N_{s1} + \alpha_s/L, \dots, N_{sL} + \alpha_s/L)$$

$$\eta_l^{(t)} \sim \text{Dir}(C_{l1} + \lambda_s/W, \dots, C_{lW} + \lambda_s/W)$$

输出: 某种场景类别下目标的分布情况  $\pi_s$ ; 某目标的视觉词汇分布情况  $\eta_l$ 。

2) 利用场景目标采样算法得到的全局知识约束目标识别过程

输入:  $p(s_j | \mathbf{w}), p(\mathbf{s}, \mathbf{o}_j | \mathbf{w}), j = 1, \dots, L$ 。

(1) 由场景分析得到的  $p(\mathbf{s} | \mathbf{w})$ , 目标与场景的

联合概率分布  $p(\mathbf{s}, \mathbf{o} | \mathbf{w})$  知

$$p(\mathbf{o} | \mathbf{s}, \mathbf{w}) \propto \frac{p(\mathbf{s}, \mathbf{o} | \mathbf{w})}{p(\mathbf{s} | \mathbf{w})}$$

(2) 对于每个目标  $\mathbf{o}_j, j = 1, \dots, L$ , 求目标条件概率最大值  $p_{\max}(\mathbf{o} | \mathbf{s}, \mathbf{w})$ 。

输出:  $\max\{p(o_1 | s_j, \mathbf{w}), \dots, p(o_L | s_j, \mathbf{w})\}$ 。

若两幅场景图像中的区域目标具有完全一致的视觉外观, 识别过程不能迅速区分这两类目标, 如图6所示。但融入完成场景分析后的全局知识便能解决两个目标的视觉外观歧义问题<sup>[11]</sup>。由此可见场景分析可用于指导目标识别。



(a) 室内浴室场景, 目标为电吹风机 (b) 室外街道场景, 目标为钻孔机

图6 目标视觉外观歧义问题

Fig. 6 The problem of object appearance ambiguity

## 2.2 目标识别反馈机制下的场景分析

场景中的目标识别是为了对场景进行更好的解释, 其目的是为了描述场景, 它是场景分析的基础<sup>[12-13]</sup>。可用目标识别结果反馈作用于场景分析过程, 当初步的图像场景分析完成后, 继续分析得到的图像目标分布。当图像的多个目标标记结果与场景类别相互矛盾时, 选择与该场景目标分布最为接近的其他场景类别重新进行分析。

目标识别结果反馈作用场景分析过程如下:

1) 目标级生成图模型下完成目标部分采样

输入: 当前  $j$  图像中的  $N_j$  个特征的部分索引向量为  $\mathbf{z}_j^{(t-1)}$ ;  $j$  图像中的目标类别为  $\mathbf{o}_j = l$ ; 采样迭代次数  $T$ 。

(1) 将  $N_j$  个视觉特征进行随机排列为  $\tau(\cdot)$ ;

(2) 令  $\mathbf{z}_j = \mathbf{z}_j^{(t-1)}$ , 对每个  $i \in \{\tau(1), \dots, \tau(N_j)\}$  重新采样其索引  $z_{ji}$ ;

(3) 设定  $\mathbf{z}_j^{(t)} = \mathbf{z}_j$ , 部分的权重分布和参数分布可采样为

$$\pi_l^{(t)} \sim \text{Dir}(N_{l1} + \alpha_l/K, \dots, N_{lK} + \alpha_l/K)$$

$$\eta_k^{(t)} \sim \text{Dir}(C_{k1} + \lambda_l/W, \dots, C_{kW} + \lambda_l/W)$$

输出: 某种目标类别下部分的分布情况  $\pi_l$ ; 某部分的视觉词汇分布情况  $\eta_k$ 。

2) 利用目标部分采样算法得到的局部知识反馈作用于场景分析过程

输入: $p(o_j | w), p(s_j, o | w), j = 1, \dots, J$ 。

(1)由目标分析得到的 $p(o | w)$ ,目标与场景的联合概率分布 $p(s, o | w)$ 知:

$$p(s | o, w) \propto \frac{p(s, o | w)}{p(o | w)}$$

(2)对于每个场景 $s_j, j = 1, \dots, J$ ,求场景条件概率最大值 $p_{\max}(s | o, w)$ 。

输出: $\max\{p(s_1 | o_j, w), \dots, p(s_j | o_j, w)\}$ 。

若两类场景图像具有十分类似的视觉特征,没有采用反馈机制时模型是无法迅速区分出场景类别的,如图7所示。但融入完成目标识别后的局部知识可知:计算机通常出现在办公室图像中;汽车通常出现在街道图像中。从而得知图7(a)为办公室场景,图7(b)为街道场景。由此可见目标识别可反馈作用于场景分析。

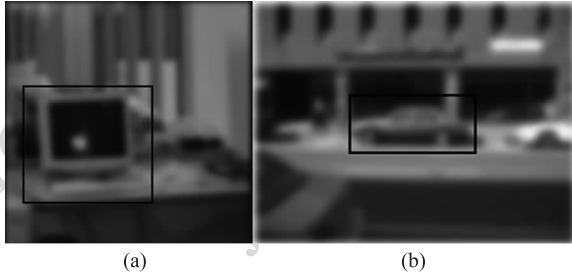


图7 场景视觉歧义问题

Fig. 7 The problem of scene appearance ambiguity

### 2.3 反馈过程的基本步骤

真实世界场景中,其图像内容大多存在上下文关系,这种关系形成了典型的语境知识<sup>[14]</sup>,利用这种知识可使图像分析的中层任务(目标识别)和高层任务(场景分析)之间形成反馈机制,其具体过程如图8所示。

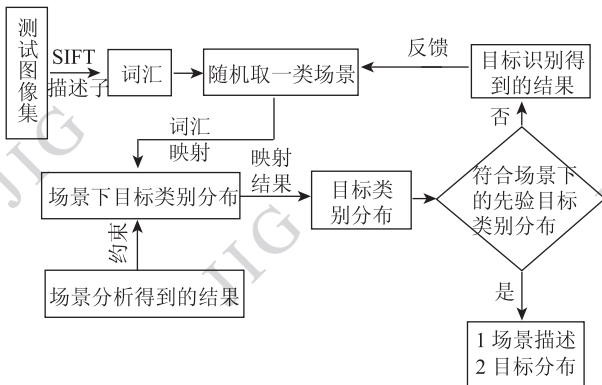


图8 反馈过程示意图

Fig. 8 Illustration of the feedback process

在测试阶段,1)输入测试图像中提取出的特征词汇,从图像集的场景类别集中随机选择一类场景;

2)对所有特征词汇进行词汇映射,得到该场景下的目标分布;3)利用场景分析得到的结果对上一步得到的目标分布进行约束处理,得到约束关系处理后的目标分布;4)将该目标分布结果与此类场景下的先验目标分布作对比,若出现目标是互斥的情况,则去除识别结果概率较低的那个目标,然后根据目标识别结果的反馈信息重新选择场景,回到第2)步进行处理。该过程形成了分层反馈的机制,以此实现图像集下场景分析和目标识别的统一框架。

## 3 实验及结果分析

### 3.1 图像数据集

采用 Lotus Hill 场景图像数据库中4类场景子集进行实验。包括2类室外场景:1)312幅街道场景,均含2类目标,车辆、公路;2)278幅郊外场景,均含2类目标,树木、草地。2类室内场景:3)294幅客厅场景,均含2类目标,沙发、茶桌;4)306幅办公室场景,均含2类目标,计算机屏幕、键盘。如图9所示。



图9 Lotus Hill 4类场景图像

Fig. 9 Example images from Lotus Hill dataset containing 4 scene categories

实验过程中4类场景的图像集分别采用独立的词汇字典,每类集合词汇个数均为 $W = 1\ 600$ ;每类场景图像均选取200幅作为各自的训练图像集,其余作为测试图像集。

### 3.2 实验结果

采用算法1和算法2中的 Rao-Blackwellized Gibbs 采样器进行生成图模型参数学习时,对已经标记好的特征参数和目标参数进行不同次数的迭代采样。图10给出了生成图模型对两类场景图像(室外街道场景、室内办公室场景)的目标识别效果,其中(a)(h)表示原始图像;(b)——(e)(i)——(l)表示采样迭代次数分别为200次、300次、400次、500次时,该场景类别下目标部分在图像中对应的后验概率较高的特征标记;(f)(m)表示500次迭代采样时该场景类别下目标类别的部分分布;(g)(n)表示500次迭代采样时部分对应的目标类别标记。

由图10可以看出,迭代采样的次数影响着目标识别的精确度,但增加迭代采样次数会使计算代价

增大。然而随着迭代采样次数的增加,迭代曲线呈收敛趋势。图 11 给出了本文实验中生成图模型下

的迭代收敛曲线,由此可依据目标识别的精确度和计算量的大小来选择合适的迭代采样次数。

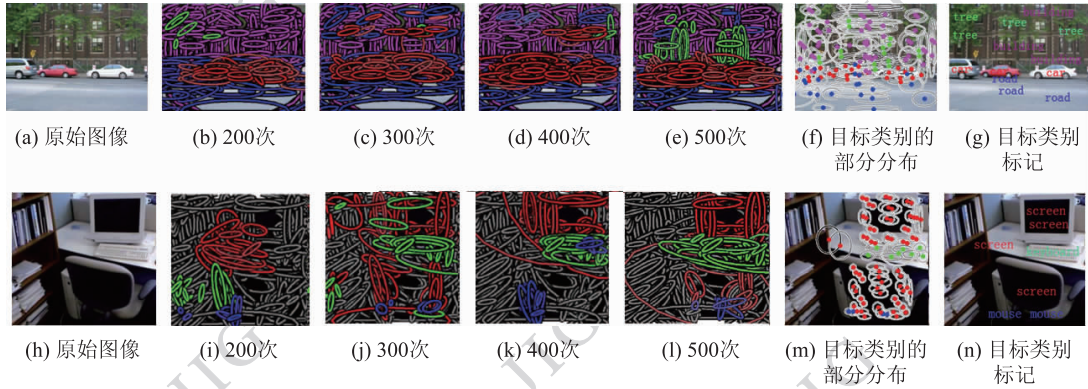


图 10 生成图模型下的目标识别

Fig. 10 Object recognition produced by generative graph model

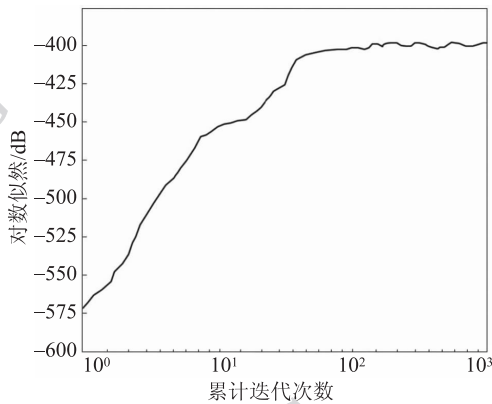


图 11 迭代收敛曲线

Fig. 11 Iteration convergence curve

通过使用本文的生成图模型,4 类场景的平均分辨率和 8 类目标的平均识别率分别达到了 91.25% 和 90.625%,其中场景分辨率由目标识别结果反馈作用于场景分析过程后计算得到,即 2.2 节中的  $p_{\max}(s | o, w)$ ;目标识别率由场景分析结果约束目标识别过程后计算得到,即 2.1 节中的  $p_{\max}(o | s, w)$ 。图 12 和图 13 给出了场景分析与目标识别的混淆矩阵结果,平均分辨率/识别率( $a_p$ )即是混淆矩阵主对角线各数据和的平均值,混淆矩阵的行表示模型识别出的场景/目标类别,列表示真实场景/目标类别。

图 14 显示了生成图模型反馈方法和当前先进的 HDP 图模型方法各自检测 4 类场景和 8 类目标的 ROC 曲线图以及这两种方法的 ROC 平均性能曲线对比图:(a) — (d) 为检测 4 类场景的 ROC 曲线对比;(e) — (l) 为检测 8 类目标的 ROC 曲线对比,(m) 为两种方法检测场景与目标的 ROC 平均性能

$a_p = 91.25\%$

|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 街道  | .91 | .01 | .02 | .07 |
| 郊外  | .08 | .92 |     |     |
| 客厅  | .04 |     | .89 | .07 |
| 办公室 | .03 |     | .04 | .93 |
|     | 街道  | 郊外  | 客厅  | 办公室 |

图 12 4 类场景分析实验的混淆矩阵

Fig. 12 Confused matrix of 4 scene category

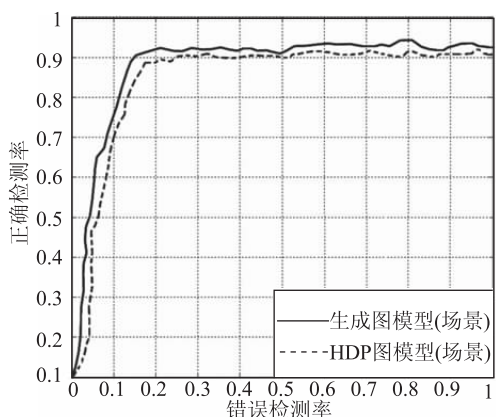
$a_p = 90.625\%$

|    |     |     |     |     |     |     |    |    |
|----|-----|-----|-----|-----|-----|-----|----|----|
| 车辆 | .96 | .04 |     |     |     |     |    |    |
| 公路 | .02 | .91 |     | .07 |     |     |    |    |
| 树木 |     | .88 |     | .12 |     |     |    |    |
| 草地 | .09 | .06 | .85 |     |     |     |    |    |
| 沙发 |     |     | .89 | .06 | .03 | .02 |    |    |
| 茶桌 |     |     | .07 | .90 | .03 |     |    |    |
| 屏幕 |     |     | .03 | .02 | .94 | .01 |    |    |
| 键盘 |     |     | .03 | .02 | .03 | .92 |    |    |
|    | 车辆  | 公路  | 树木  | 草地  | 沙发  | 茶桌  | 屏幕 | 键盘 |

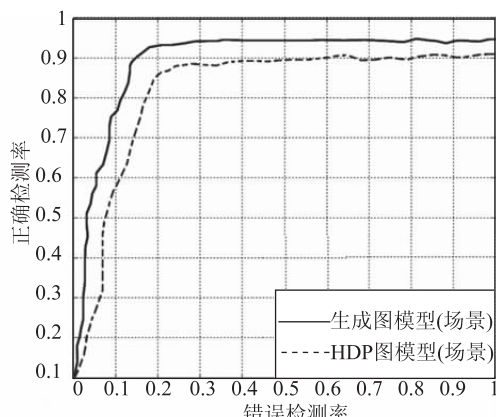
图 13 8 类目标识别实验的混淆矩阵

Fig. 13 Confused matrix of 8 object category

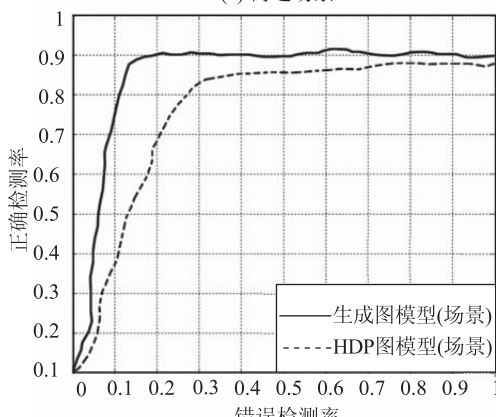
曲线对比。其中错误检测率为场景/目标分析结果与实际不同,正确检测率为场景/目标分析结果与实际相同。实验结果表明,生成图模型反馈方法下的场景分析和目标分析均优于 HDP 模型,同 HDP 相



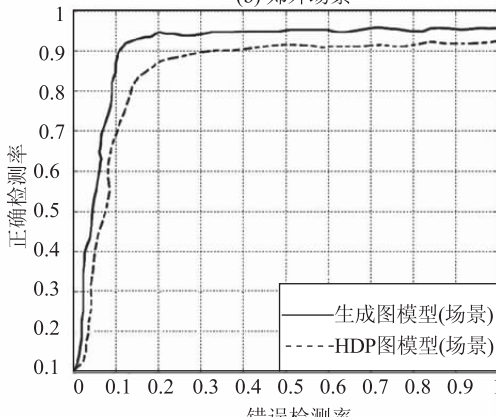
(a) 街道场景



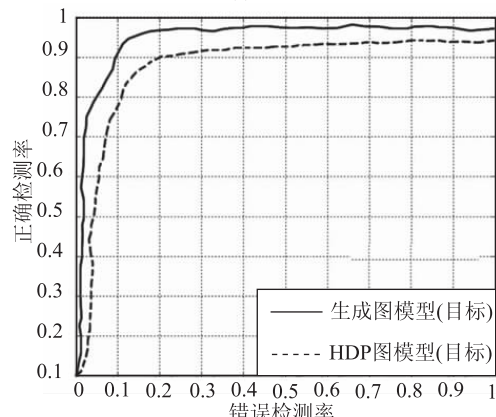
(b) 郊外场景



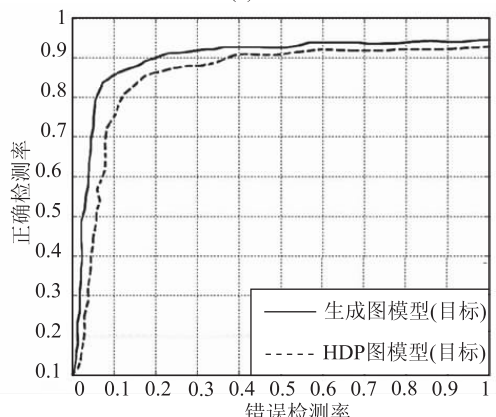
(c) 客厅场景



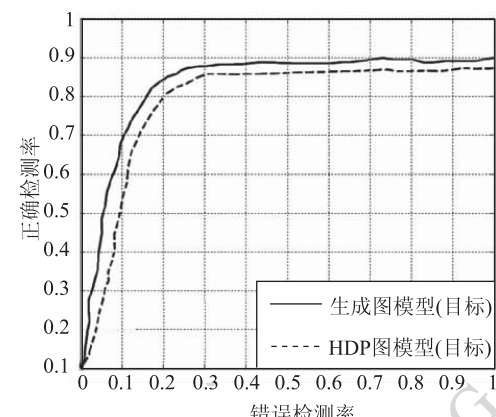
(d) 办公室场景



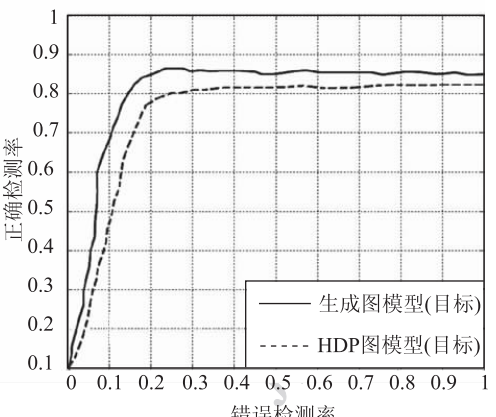
(e) 车辆



(f) 公路



(g) 树木



(h) 草地

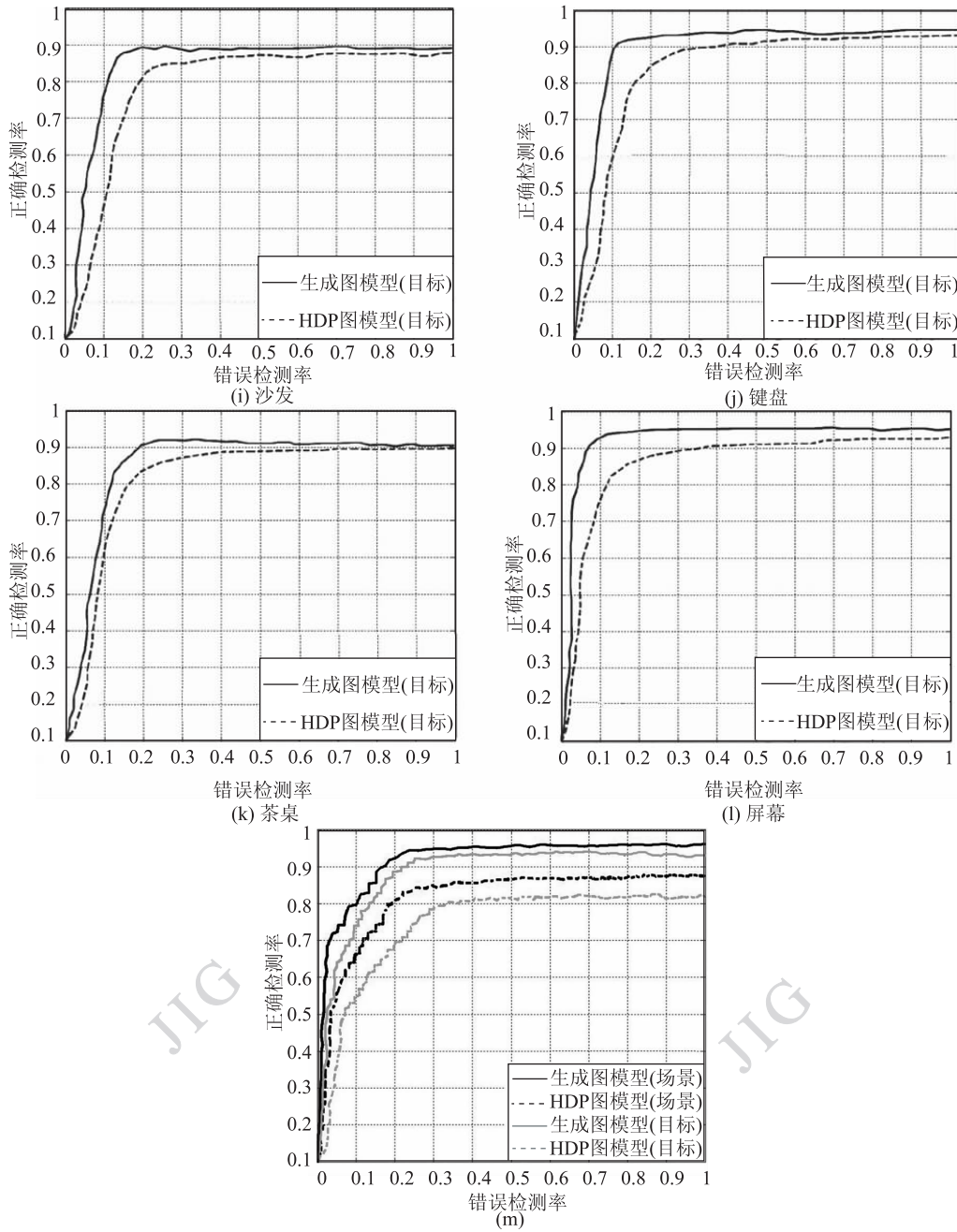


图14 生成图模型与HDP模型的ROC性能曲线对比

Fig. 14 Comparison ROC recognition performance curves of generative graph model (solid lines) and HDP model (dashed lines)

比,本文方法特点如下:建立了场景分析和目标识别间的反馈机制,将中低层的数据处理分析与高层的知识表达推理有效结合,误分情况可以得到有效的控制。

该生成图模型反馈方法的具体实验过程如图15和图16所示。

1) 场景分析指导目标识别

首先,将输入图像看作一个整体,类似于一个单目标,分析出其场景类别;其次,将各个目标实

例看作是场景的部分,得到目标初始标记;然后,采用场景全局知识约束目标识别结果,将与场景内容冲突的目标重新识别,选择其他概率较高的目标类别作为该目标的标记。如图15,电脑桌的视觉特征和街道的视觉特征类似,在识别过程中错误的被标记为“road”,而识别完成后采用图像的场景信息“office”去约束识别结果,可将其修正为背景目标。

2) 目标识别结果反馈作用于场景分析

模型对输入的图像仍然先分析其场景类别,然后将场景图像中的目标当作常规意义的目标部分进行标记。当图像所属的场景类别的概率分布较为平

均,图像的场景类别信息就可能出现误差,若此时图像的多个目标标记结果与场景类别相互矛盾,且标记结果的概率较高,则可根据这些标记目标重新分析图像所属场景类别,如图16。

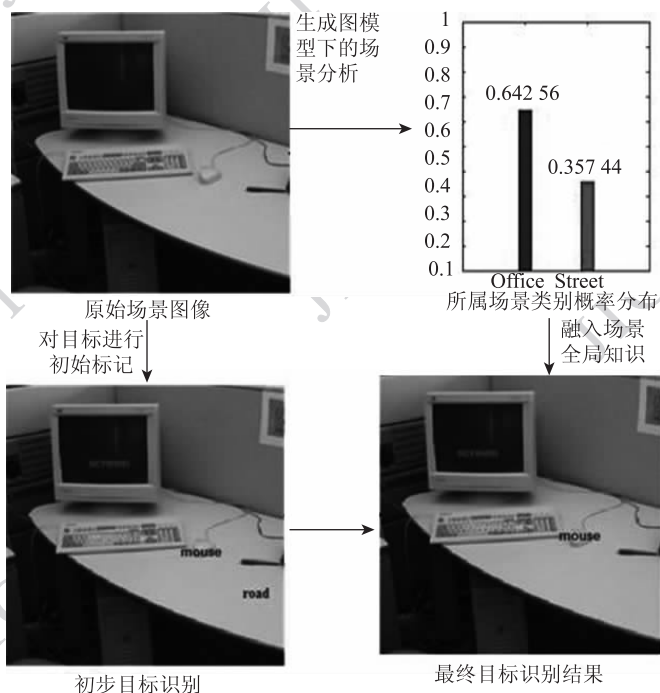


图15 场景分析指导目标识别

Fig. 15 Illustration of scene analysis guide object recognition

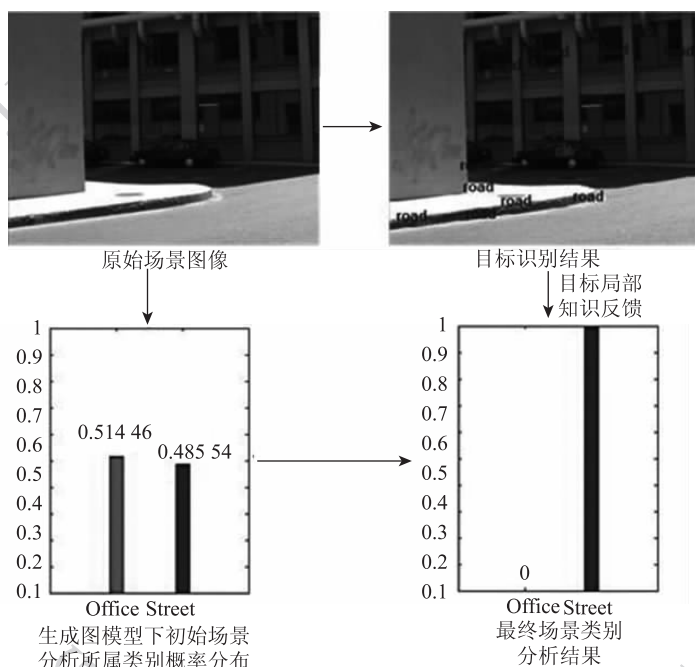


图16 目标识别结果反馈作用于场景分析

Fig. 16 Illustration of object recognition feedback effect on scene analysis

## 4 结 论

构建了一种目标识别和场景分析之间的反馈过程框架,并在此基础上探讨和分析了整体的场景分析与局部的目标分析之间的反馈作用机制,尤其是得到的场景类别先验信息对目标分析的影响,从而体现了整体对局部分析的影响。该框架以生成图模型和概率论为理论基础,结合分层思想,运用 Gibbs Rao-Blackwellized 采样方法,同时融入了场景级全局知识和目标级局部知识。通过实验表明构建的生成图模型具有较好的场景分析效果,并对目标分析有一定的指导作用。该生成图模型的反馈机制属首次尝试,之后的工作将改进模型的学习推理方案,在训练过程中减少监督,用更广泛的实验来验证模型。

### 参考文献 (References)

- [ 1 ] L Fei-Fei, P Perona. A bayesian hierarchical model for learning natural categories [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, California: IEEE Computer Society, 2005: 524-532.
- [ 2 ] L.-J. Li, R. Socher, L Fei-Fei. Towards total understanding classification, annotation and segmentation in an automatic framework [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida: IEEE Computer Society, 2009: 2036-2043.
- [ 3 ] Myung J C, Joseph J. Lim. Exploiting hierarchical context on a large database of object categories [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, California: IEEE Computer Society, 2010: 129-136.
- [ 4 ] Pedro F F, Olga Veksler. Tiered labeling with dynamic programming [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, California: IEEE Computer Society, 2010: 3097-3104.
- [ 5 ] Leo Z, Chen Y H, Alan Yuille, et al. Latent hierarchical structural learning for object detection [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, California: IEEE Computer Society, 2010: 1062-1069.
- [ 6 ] Yee W T, Michael I J, Matthew J B, et al. Hierarchical dirichlet processes [ J ]. American Statistical Association, 2006, 101(476): 1566-1581.
- [ 7 ] M I Jordan. An Introduction to Graphical Models [ M ]. Cambridge: Center for Biological and Computational Learning, MIT Press, 1997.
- [ 8 ] Erik B S. Graphical models for visual object recognition and tracking [ D ]. Cambridge: Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, 2006.
- [ 9 ] Xie Z. Researches for key issues and methods in image understanding [ D ]. Hefei: Hefei University of Technology, 2007. [ 谢昭. 图像理解的关键问题和方法研究 [ D ]. 合肥: 合肥工业大学, 2007. ]
- [ 10 ] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: Workshop on Generative Model Based Vision, 2004.
- [ 11 ] Gao J, Xie Z. Image Understanding Theory and Methods [ M ]. Beijing: Science Press, 2009. [ 高隽, 谢昭. 图像理解理论与方法 [ M ]. 北京: 科学出版社, 2009. ]
- [ 12 ] Aude O. The role of context in object recognition [ J ]. Trends in Cognitive Sciences, 2007, 11(12), 520-527.
- [ 13 ] Carolina G, Brian M, Gert L. Multi-class object localization by combining local contextual interactions [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, California: IEEE Computer Society, 2010: 113-120.
- [ 14 ] Marcus R, Michael S. What helps where - and why? semantic relatedness for knowledge transfer [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, California: IEEE Computer Society, 2010: 910-917.