

刘天亮¹ 陆泮宇¹ 戴修斌¹ 刘峰¹ 罗杰波²

融合信息化边界和多模态特征的室内空间布局估计

摘要

为感知室内空间布局,提出一种基于信息化边界和多模态特征的场景布局估计方法.首先,采用 VGG-16 全卷积神经网络预测蕴含空间布局先验的信息化边界图.其次,采用 Canny 边缘检测和投票策略估计水平和竖直方向消失点,从消失点等角度间隔引出射线细采样信息化边界能量高的区域.接着,采用 VGG 空间多尺度卷积神经网络估计几何深度和法向特征.然后,积分几何求和候选布局多边形中多模特征描述一元共生,候选布局的表面法向平滑和位置关系确定二元标记约束.最后,采用结构化支持向量机器学习模型,最大布局候选得分以推理布局.实验结果表明,与经典方法相比,本估计方法可以有效改善布局的完整度.

关键词

空间布局;卷积神经网络;场景理解;信息化边界

中图分类号 TP391.41

文献标志码 A

收稿日期 2019-10-20

资助项目 国家自然科学基金(61001152,31200747,61071091,61071166,61172118);江苏省自然科学基金(BK2012437);南京邮电大学校级科研基金(NY214037);国家留学基金(201208320219)

作者简介

刘天亮,男,博士,副教授,硕士生导师,研究方向为图像处理与计算机视觉.liutl@njupt.edu.cn

0 引言

估计室内场景的空域布局结构是计算视觉领域中场景理解的一项重要工作.图像理解是计算机视觉研究的最终愿景之一,能有效帮助和提高计算机分析和认知复杂多变的室内外场景,空间布局估计是近年研究热点^[1-3].估计室内场景的空间布局,是曼哈顿模型假设下根据房间结构划分室内场景并对应左墙、中墙、右墙、天花板和地面这5个空间面,这能为深度恢复、虚拟现实、智能家居、移动机器人导航等应用提供有效空间结构信息^[1-2,4].

室内空间布局通常存在复杂室内场景中容易出现大物体(如家具)遮挡地面和各个墙面、堆积小尺寸目标物体或杂物的干扰、不同相机拍摄角度引起个别多边形面有缺失等现象.而室内布局估计方法在模型设计和特征选取上各有所侧重,且大致分为2类.第1类是分步分治最优选取场景布局候选并结构化预测,第2类是卷积神经网络端到端学习的场景密集几何语义标注.第1类通常利用布局多边形区域面剖分而分步骤提取、拼接视觉感知特征做结构化预测.文献[5]采用3D盒模型结构参数化房间布局并推理布局结构.传统的特征主要分为线结构特征和区域特征.文献[6]在曼哈顿世界假设下利用消失点估计生成布局候选,并结合直线段和蕴含目标结构相关的几何上下文等特征描述布局候选以优化估计.消失点估计法^[7-8]常常利用检测到图像中所有长直线段并投票计算得分,根据得分高低依次赋值给相应消失点.文献[9]进而融入方向图特征提高布局估计精度,而文献[10]则抽取空间布局线索深挖直线段的连接属性.文献[11]将场景深度信息嵌入布局建模以增强特征感知力.文献[12]将室内的家具和装饰品等视为隐变量,采用由消失点引出参数化射线的夹角表征布局候选,并基于4个离散随机变量预测布局.文献[13]拆分目标函数并增加约束,以更少时间求解到更确切结果.

分治策略动机清晰,其子问题或子步骤变得简单、可控且更易解决,但子问题最优未必是全局的最优,每个子步骤最优也不意味全过程看是最优.第2类从全卷积网络模型端到端感知标注整幅场景语义结构.文献[14]仅利用外观图像给出一种规避码本的室内功能区表示与建模,以实现家庭服务机器人对室内服务环境的分类.随着近来卷积神经网络兴盛,故第2类方法从全卷积网络模型感知整幅图像的全局特征.文献[15]利用全卷积神经网络(Fully Convolution Networks,

1 南京邮电大学 江苏省图像处理与图像通信重点实验室,南京,210003

2 罗彻斯特大学 计算机科学系,美国 罗彻斯特,14627

FCNs)^[16]提取房间各种布局边界特征并信息化边界(Informative Edges, IE)的参数有效学习.文献[17]给出基于 FCNs 模型端到端学习直接高质量标注布局多边形区域语义标签,但都要训练数据集足够大且训练周期长.相比分治策略,数据驱动全局优化的端到端学习具有协同增效优势,可能获得全局上更优的解,但整合结构化先验不够灵活.

于是,针对家具和饰品等杂乱干扰布局候选、多模态一元特征以及二元关系利用等问题,本文研究基于信息化边界诱导布局候选和区域多模态特征有机融合的室内场景布局估计法.

1 室内场景布局估计整体框架

图 1 给出了 2 部分组成的本文方法流程.第 1 部分利用信息化边界由粗到精布局候选项生成 3 步:利用全卷积神经网络计算室内场景的信息边界能量;从边缘检测估计的消失点等角度间距采样粗划分图像区域;从水平和垂直方向分别选取已划分的信息化边界能量较高的区域并细采样得到布局候选.第 2 部分融合区域级多模态异构特征建模场景空间布局并学习推理.首先,利用 VGG 空间多尺度卷积神经网络估计场景表面的几何法向量和深度特征.然后,采用消失射线最大夹角参数来 3D 盒式场景布局的结构化建模.积分几何法累积候选布局多边形中线组特征和几何上下文等多模特征来描述布局候选的一元呈现;布局表面法向平滑和位置关系确定着二元约束关系.最后,采用结构化支持向量机学习模型参数;排序场景布局候选得分,选得分最高者为最终场景布局.

2 信息性边界诱导场景布局候选生成

2.1 VGG-16 全卷积神经网络感知场景信息化边界

受有效解决像素标注问题的全卷积网络 FCNs^[16]启发,本文利用反卷积层上采样粗糙且抽象输出密集像素级标注,联合训练 FCNs 网络完成 2 个任务:生成各种房间边界的信息化边界和标注目标布局的几何上下文标签.采用 Caffe 框架学习 VGG-16 结构的 FCNs 网络,用 32 像素预测步长的 FCNs 训练 NYUD v2 RGBD 数据集中 40 个类别的室内语义分割^[15].文献[15]原始 FCNs 网络有 2 个输入流:RGB 图像和深度特征.本文在预训练模型基础上舍弃深度特征,仅用 RGB 层初始 FCNs 网络.然后,微调网络的学习率和高动量分别为 10^{-4} 和 0.99,新加

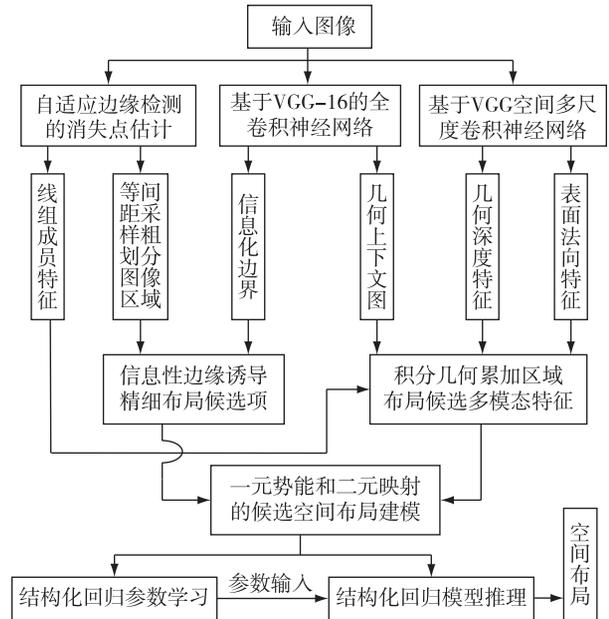


图 1 室内场景布局估计方法流程

Fig. 1 Flowchart of the proposed layout estimation approach for indoor scenes

入的最后一个卷积和反卷积层的学习率设为 10^{-3} .共享 FCNs 除反卷积层外的所有层联合训练.最后反卷积层中输出产生各自对应类别的不同 softmax 概率图.训练时,信息化边界标签预测和几何上下文标签预测这 2 个交叉熵分类损失的总和构成该 FCNs 网络总损失.

2.2 信息化边界诱导由粗到精采样布局候选

首先,利用自适应阈值 Canny 法检测出场景中边缘直线段^[7].然后,用 Rother's 算法和随机一致性采样计算场景图像中 3 个正交的消失点 (p_{v1} 、 p_{v2} 和 p_{v3}),采用投票策略计算点和线段间角度的偏差得分^[8].消失线段 l 与候选消失点 p 的关系:

$$v(l, p) = |l| \cdot \exp\left(-\frac{\alpha}{2\sigma^2}\right), \quad (1)$$

其中 l 和 $|l|$ 表示消失线段及其长度, α 是消失线段 l 与消失点 p 和消失线段中点连线的夹角,鲁棒阈值 $\sigma = 0.1$.接着,从图像水平和垂直方向的消失点等角度间隔引出采样射线进行图像区域的粗划分^[14].图 2b 从消失点 p_{v1} 和 p_{v2} 分别引出 10 条射线粗划分场景区域.

接着,以信息化边界图为空间布局先验,分别从水平和垂直方向计算射线划分图像的各区域中高能量的信息化边界 IE 的像素数量.图 2c 展示信息化边界 IE 中各个布局边界附近有蕴含着室内空间布局

的较高能量.再次,根据消失点 p_{v3} 的位置将水平区域分为上区域和下区域,竖直区域分为左区域和右区域.分别从4个区域选取IE像素数量前K个的区域.最后,再引N条射线自适应细采样所选区域,产生最终较精细布局候选项,如图2d所示.

3 融合区域多模特征和位置约束的布局估计

3.1 场景布局区域级多模特征感知

为了有效确定最佳布局,本文采用直线段成员特征 f_{lm} 、几何上下文特征 f_{gc} 、法向量特征 f_{nv} 和深度特征 f_d 等多模态区域级特征.根据2.2节中估计消失点的长直线段,构成相应直线段成员特征^[5],并根据2.1节中信息化边界估计得到相应几何上下文特征^[5,15].受文献[18]启发,针对输入图像采用级联式

空间多尺度卷积神经网络估计对应的法向量图和深度图,作为多尺度感知的3通道法向量特征和1通道深度特征,如表1和图3所示.不同感知任务决定模型通道数C.若估计法向量,则 $C=3$;若估计深度,则 $C=1$.后2种区域级特征蕴含丰富的3D空间几何信息,能细粒度且稳健地感知并表征场景布局候选中的区域属性.

3.2 室内3维盒式布局结构化建模

地面 F_1 、左墙 F_2 、前墙 F_3 、右墙 F_4 及天花板 F_5 构成三维盒式布局,其可以采用场景消失点诱导的消失射线最大夹角来参数化^[5].鉴于杂物对空间布局估计干扰较大,室内场景空间布局利用多模态特征的一元共生呈现和二元平滑及其位置约束来结构化建模,公式为

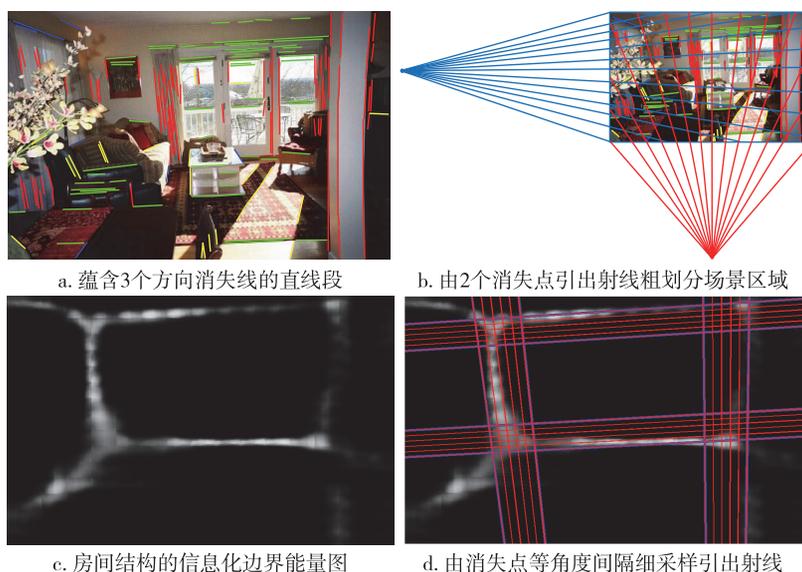


图2 由粗到精室内场景候选布局生成示意图

Fig. 2 Generation diagram of coarse-to-fine layout candidates for indoor scenes

表1 多尺度卷积神经具体网络参数

Table 1 Detailed parameters in multi-scale convolutional neural network

网络层名称	卷积核		网络层名称	卷积核		网络层名称	卷积核	
	尺寸	个数		尺寸	个数		尺寸	个数
卷积层 a1	3×3	64	卷积层 b1	9×9	96	卷积层 c1	9×9	64
卷积层 a2	3×3	128	卷积层 b2	5×5	(96+64)	卷积层 c2	5×5	(64+C)
卷积层 a3	3×3	256	卷积层 b3	5×5	64	卷积层 c3	5×5	64
卷积层 a4	3×3	512	卷积层 b4	5×5	64	卷积层 c4	5×5	64
卷积层 a5	3×3	512	卷积层 b5	5×5	C	池化层 c1~c4	3×3	
池化层 a1~a5	2×2		池化层 b1~b5	3×3				
全连接层 a6		4 096	反卷积层 b6	2×2	C			
全连接层 a7		64						
反卷积层 a8	4×4	64						

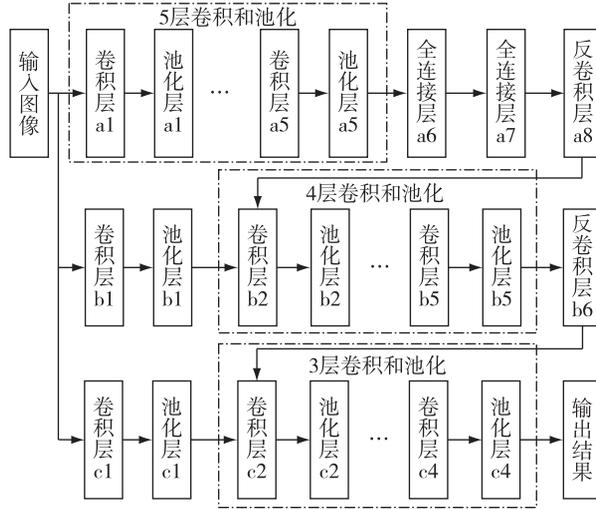


图3 多尺度卷积神经网络模型

Fig. 3 Frame of multi-scale convolutional neural network model

$$f(\mathbf{x}, \mathbf{y}; \boldsymbol{\omega}) = \sum_{\alpha \in F} \boldsymbol{\omega}_\alpha^T \Psi_\alpha(\mathbf{x}, \mathbf{y}_\alpha) + \sum_{\alpha, \beta \in F} \boldsymbol{\omega}_b^T \Psi_b(\mathbf{y}_\alpha, \mathbf{y}_\beta), \quad (2)$$

其中 $\alpha, \beta \in F = \{F_1, F_2, F_3, F_4, F_5\}$, \mathbf{x} 是输入, \mathbf{y}_α 和 \mathbf{y}_β 表示布局候选中某个多边形区域, $\Psi_\alpha(\mathbf{x}, \mathbf{y}_\alpha)$ 表示布局候选多边形中区域级特征, $\boldsymbol{\omega}_\alpha$ 表征 $\Psi_\alpha(\mathbf{x}, \mathbf{y}_\alpha)$ 一元共生特征权重. $\Psi_b(\mathbf{y}_\alpha, \mathbf{y}_\beta)$ 表征布局候选两两多边形之间位置关系, $\boldsymbol{\omega}_b$ 表征 $\Psi_b(\mathbf{y}_\alpha, \mathbf{y}_\beta)$ 二元平滑约束权重.

式(2)第1项采用4种模态特征刻画空间布局相关区域级属性的一元共生呈现关系:

$$\Psi_\alpha(\mathbf{x}, \mathbf{y}) = \sum_{\alpha \in F} \Psi_\alpha(\mathbf{x}, \mathbf{y}_\alpha) = \Psi_{LM}(\mathbf{x}, \mathbf{y}) + \Psi_{GC}(\mathbf{x}, \mathbf{y}) + \Psi_N(\mathbf{x}, \mathbf{y}) + \Psi_D(\mathbf{x}, \mathbf{y}), \quad (3)$$

其中 $\Psi_{LM}(\mathbf{x}, \mathbf{y})$ 表示直线段组成员特征势能^[5], $\Psi_{GC}(\mathbf{x}, \mathbf{y})$ 表示 FCNs 网络得到的几何上下文势能^[15], $\Psi_N(\mathbf{x}, \mathbf{y})$ 是法向量特征约束多边形的方向势能^[18], $\Psi_D(\mathbf{x}, \mathbf{y})$ 是多边形中深度特征势能^[18]. 利用几何积分图像法累积求和布局候选 \mathbf{y} 可以得到这些区域级特征.

式(2)第2项表示二元平滑和位置约束的特征映射关系^[13]:

$$\sum_{\alpha, \beta \in F} \Psi_b(\mathbf{y}_\alpha, \mathbf{y}_\beta) = \sum_{\alpha, \beta \in F} \Psi_N(\mathbf{y}_\alpha, \mathbf{y}_\beta) + \sum_{\alpha, \beta \in F} \Psi_{LO}(\mathbf{y}_\alpha, \mathbf{y}_\beta), \quad (4)$$

其中 $\Psi_N(\mathbf{y}_\alpha, \mathbf{y}_\beta)$ 表示两两多边形区域之间的法向量内积 IP 的映射关系. 若 $IP > 0$, 说明夹角小于 90° , 即两平面同向; 若 $IP = 0$, 说明正交, 即两平面相互垂

直; 若 $IP < 0$, 说明夹角大于 90° , 即两平面反向. $\Psi_{LO}(\mathbf{y}_\alpha, \mathbf{y}_\beta)$ 表示布局候选多边形对应室内三维空间墙面间的位置关系, 如表2所示. 表2中, A表示列区域名称, B表示行区域名称, 数值表示A与B的位置关系. 数值1表示A不可能高于B; 2表示A不可能低于B; 3表示A不可能在B右边; 4表示A不可能在B左边; 5表示A不可能在B前面; 6表示A不可能在B后面. 这些位置约束能有效理清空间布局候选中两两多边形之间在三维空间上的结构依赖.

表2 室内场景布局的多边形区域之间二元位置关系

Table 2 Pairwise location relationship between polygon regions in indoor scene layout

B	A				
	天花板	地面	左墙	前墙	右墙
天花板		2	2,4	2,5	2,3
地面	1		1,4	1,5	1,3
左墙	1,3	2,3		3,5	3
前墙	1,6	2,6	4,6		3,6
右墙	1,4	2,4	4	4,5	

3.3 结构化输出学习空间布局模型

给定训练集 $D = \{\mathbf{x}_i, \mathbf{y}_i\}$ ($1 \leq i \leq I, I$ 为总数), 采用结构化输出 SVM 学习框架来最大化分类决策间隔^[5,20], 以测度最优的候选布局与基准布局之间的结构化损失来学习布局模型参数 \mathbf{w} :

$$\min_{\boldsymbol{\omega}, \xi} \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \lambda \sum_i \xi_i, \quad \text{s.t. } \xi_i \geq 0, \forall i, \\ \boldsymbol{\omega}^T \Psi(\mathbf{x}_i, \mathbf{y}_i) - \boldsymbol{\omega}^T \Psi(\mathbf{x}_i, \mathbf{y}) \geq \Delta(\mathbf{y}_i, \mathbf{y}) - \xi_i,$$

$$\forall i, \forall \mathbf{y} \in Y/y_i, \quad (5)$$

其中 \mathbf{x}_i 表示训练集 \mathbf{D} 中第 i 个场景, ζ_i 为对应 \mathbf{x}_i 的松弛变量, λ 为平衡系数(本文 $\lambda = 1$), 权重向量 $\boldsymbol{\omega} = [\boldsymbol{\omega}_a, \boldsymbol{\omega}_b]$, Y 表示布局候选集合, \mathbf{y} 为集合 Y 的元素(即第 i 个场景的布局候选), \mathbf{y}_i 表示基准标注布局, $\Psi(\mathbf{x}_i, \mathbf{y})$ 和 $\Psi(\mathbf{x}_i, \mathbf{y}_i)$ 分别表示提取 \mathbf{x}_i 对应的布局候选 \mathbf{y} 和基准标注布局 \mathbf{y}_i 的对应特征映射. $\Delta(\mathbf{y}_i, \mathbf{y})$ 量化 \mathbf{y}_i 和 \mathbf{y} 这 2 个布局之间的测度损失, 公式为:

$$\begin{aligned} \Delta(\mathbf{y}_i, \mathbf{y}) &= \Delta_l(\mathbf{y}_i, \mathbf{y}) + \Delta_c(\mathbf{y}_i, \mathbf{y}) + \Delta_p(\mathbf{y}_i, \mathbf{y}), \\ \Delta_l(\mathbf{y}_i, \mathbf{y}) &= \sum_{k \in [1, 5]} \delta(F_{ik}, F_k), \\ \Delta_c(\mathbf{y}_i, \mathbf{y}) &= \sum_{k \in [1, 5]} \|c_{ik} - c_k\|^2, \\ \Delta_p(\mathbf{y}_i, \mathbf{y}) &= \sum_{k \in [1, 5]} \left(1 - \frac{\text{Area}(F_{ik} \cap F_k)}{\text{Area}(F_{ik} \cup F_k)}\right), \end{aligned} \quad (6)$$

其中 $F_k \in \{F_1, F_2, F_3, F_4, F_5\}$ 为布局候选 \mathbf{y} 的标注标签; $F_{ik} \in \{F_{i1}, F_{i2}, F_{i3}, F_{i4}, F_{i5}\}$ 为基准布局 \mathbf{y}_i 的标注标签, $\Delta_l(\mathbf{y}_i, \mathbf{y})$ 表示多边形面缺失误差之和, $\Delta_c(\mathbf{y}_i, \mathbf{y})$ 表示多边形质心位置误差之和, $\Delta_p(\mathbf{y}_i, \mathbf{y})$ 表示多边形像素误差之和.

3.4 最大化候选得分推理空间布局

给定空间布局模型的参数权重向量 $\boldsymbol{\omega}$, 输入图像 \mathbf{x} 以及 2.2 节产生的多个场景布局候选 \mathbf{y} , 排序布局候选得分, 得分最高者作为最优布局 \mathbf{y}^* [5], 则

$$\mathbf{y}^* = \arg \max_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}; \boldsymbol{\omega}), \quad (7)$$

其中 \mathbf{y}^* 表示最优布局, $f(\mathbf{x}, \mathbf{y}; \boldsymbol{\omega}) = \mathbf{w}^T \Psi(\mathbf{x}, \mathbf{y})$ 表示场景 \mathbf{x} 和其布局候选项 \mathbf{y} 之间的映射关系.

4 实验结果与分析

采用 Hedau 与 LSUN 室内场景数据集主观与客观评测. Hedau 数据集 [5] 包括 314 幅含有多种室内场景且稠密标注的彩色图像及对应的布局图像(200 组训练图像对和 114 组测试图像对), 包含 5 种结构化语义类别(地面、左墙、中墙、右墙、天花板). LSUN 数据集包含 5 394 幅拍摄于多种室内场景且稠密标注的彩色图像及对应的布局图像(4 000 组训练图像对、394 组验证图像对以及 1 000 幅测试图像), 提供结构化布局边缘线交叉的基准角点.

4.1 主观评估

图 4 和图 5 分别给出本文方法在 Hedau 数据集上不同模态特征组合的估计结果和 LSUN 数据集上与 Hedau 方案布局的估计结果. 实验方案包括 4 种模态特征(直线段组成员特征(LM)、几何上下文特征(GC)、深度特征(DP)、法向量特征(NV))和二元

约束关系(PW). 由图 4 可知, LM+GC+DP 特征融合估计的布局图 4d 明显优于 LM+GC 特征融合估计的图 4c. 其中图 4IV 的右墙(蓝色标注部分)的墙面有凸出三角形区域和书柜等干扰, 引入深度特征后识别右墙区域的效果明显. 其原因是深度特征能刻画场景中空间几何的景深关系且不受空间中物体遮挡的影响. 而本文方法(LM+GC+DP+NV+PW)融合 4 种模态特征并结合多边形之间结构位置约束关系的图 4e 效果, 也明显优于 LM+GC+DP 方法的图 4d. 图 4III 中左墙和前墙均为白色且不易区分, 同时有椅子等干扰, 融合法向量特征并且考虑多边形之间位置结构关系后, 场景布局边界预测更加准确. 其原因是法向量特征决定着图像多边形区域所在的平面方向且多边形之间的二元位置关系约束着多边形空间位置结构的可能组合.

图 5 整体呈现了本文方法和文献[5]方法的布局估计结果的视觉效果. 结合图 5d 和图 5c 综合表现可知, 本文方法在室内场景布局估计的视觉表现上有明显优势. 其原因是文献[5]仅从消失点引出消失射线生成布局候选, 同时只采用简单的线组成员 LM 和几何上下文 GC 特征; 而本文结合信息性边缘由粗到精优化布局候选, 同时综合利用多模态特征融合和多边形二元项约束关系来实现布局估计推理, 其更鲁棒和判别力更强.

4.2 客观评估

采用像素误差率(Pixel Error, PE)和角点误差率(Corner Error, CE)来客观评价实验效果. 前者由布局估计错误标注的所有像素数与图像中所有像素数的比值确定; 后者由布局估计的角点与基准角点之间的距离与图像对角线长度的比值确定. 二者数值越小, 实验效果均越好. 表 3 给出了本文布局估计方法在 Hedau 数据集上多模态特征组合下的像素点误差和测试耗时情况. 其测试耗时是相同测试配置环境下各算法未做整体优化的平均运行时间. 由表 3 可知, 引入深度特征的 LM+GC+DP 方法的像素点误差相较 LM+GC 法有显著地降低(降了 0.66 个百分点). 这说明深度特征蕴含场景的几何空间关系, 能更好表征场景布局的区域表面结构, 同时降低遮挡物对布局估计的影响. 相较于 LM+GC+DP 法, 依次引入法向量 NV 特征和多边形二元约束 PW 的 LM+GC+DP+NV 法和 LM+GC+DP+NV+PW 法的像素点误差分别降低了 0.25 和 0.43 个百分点, 其性能提升空间有限; 其主要原因是本文利用信息性边缘由

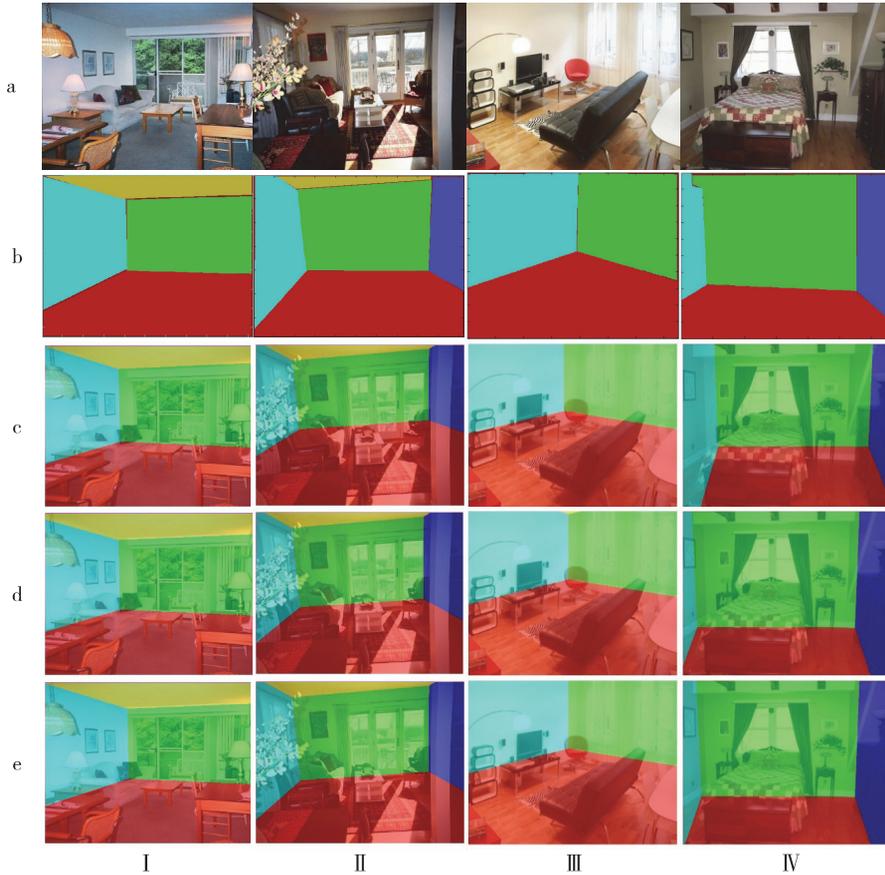


图4 Hedau 数据集上的布局估计结果

a.输入图像;b.基准布局;c.LM+GC^[5];d.LM+GC+DP;e.本文方法结果

Fig.4 Results of layout estimations on Hedau dataset, (a) input images, (b) ground truth layouts, (c) results by LM+GC method in [5], (d) results by LM+GC+DP method, and (e) results by our proposed method

粗到精生成布局候选已经在某种程度上做过布局筛选,使其对相应推理结果影响不明显.由表3中测试耗时可知,与其他方法相比,本文方法虽然耗时相对较高,但在合理范围且相差不大.本文方法显著提升了像素标注的误差性能.

表4给出本文方法相较于其他传统经典方法在Hedau和LSUN数据集上像素点误差和角点误差的对比评估.鉴于文献[13,19]方案出现在LSUN数据集之前,本文未列出其在LSUN数据集上实验结果.文献[5]方法直接由消失点产生室内场景布局候选并采用线组成员特征估计场景布局;文献[13]在Hedau框架基础上增加方向图特征估计最终场景布局;文献[15]利用深度学习预测信息边界特征,将信息边界特征和几何上下文特征融合产生场景布局估计.由表4可知,本文方法采用布局信息性边缘优化布局候选并利用几何上下文、深度、法向量等多种模态特征构建模型以约束估计布局,在像素点误差和

角点误差方面均有显著提高.相关方法在LSUN数据集上像素点误差PE明显高于其在Hedau数据集的相应PE,其原因主要是LSUN数据集中室内场景较为杂乱,比如存在易干扰的广告海报图等;图5Ⅲ右墙镂空且后面还有一个空间,易导致布局估计效果欠佳.LSUN数据集的角点误差CE看,利用信息性边缘由粗到精生成布局候选,角点精度随布局候选精度的提高而提高.

表3 不同子模块组合下本文方法在Hedau数据集上的实验结果比较

Table 3 Comparison of experimental results with different submodels on Hedau dataset

方法	像素点误差率 PE/%	运行时间/s
LM+GC	11.72	29.3
LM+GC+DP	11.06	34.6
LM+GC+DP+NV	10.81	40.2
本文最终方法 (LM+GC+DP+NV+PW)	10.63	52.1

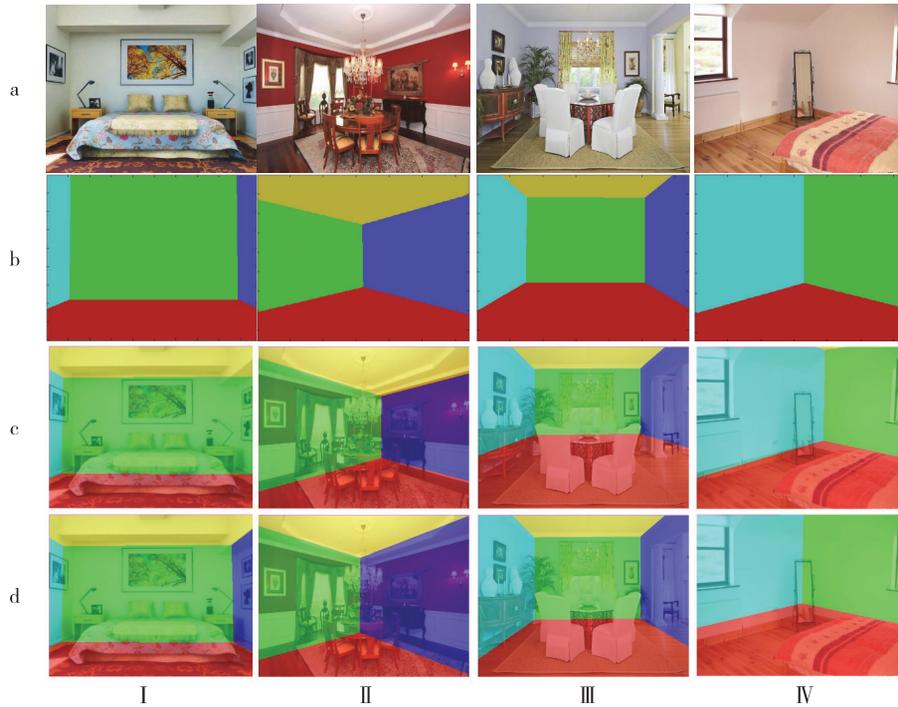


图5 LSUN数据集上的布局估计结果

a.输入图像;b.基准布局;c.文献[5]方法结果;d.本文方法结果

Fig.5 Results of layout estimations on LSUN dataset, (a) input images, (b) ground truth layouts, (c) results by LM+GC method in [5], and (d) results by our proposed method

表4 本文方法与其他经典方法的综合比较

Table 4 Comparison between our proposed method and traditional methods %

布局估计方法	Hedau 数据集		LSUN 数据集	
	像素点误差率 PE	角点误差率 CE	像素点误差率 PE	角点误差率 CE
本文	10.63	N/A	14.35	9.88
文献[15]	12.83	N/A	16.71	11.02
文献[13]	13.59	N/A	N/A	N/A
文献[20]	16.20	N/A	N/A	N/A
文献[5]	21.20	N/A	24.23	15.48

5 结语

本文提出了一种基于信息化边界和多模态异构特征融合与二元平滑空间位置约束的室内场景布局估计方法.相比于其他经典方法,本文方法能够估计到的布局整体视觉效果更佳,空间布局估计准确度更高.

参考文献

References

[1] 姚拓中,左文辉,宋加涛,等.结合物体先验和空域约束的室内空域布局推理[J].自动化学报,2017,43(8):1402-1411

YAO Tuozhong, ZUO Wenhui, SONG Jiatao, et al. Estimating spatial layout of cluttered rooms by using object prior and spatial constraints [J]. Acta Automatica Sinica, 2017, 43(8): 1402-1411

[2] 庄严,卢希彬,李云辉,等.移动机器人基于三维激光测距的室内场景认知[J].自动化学报,2011,37(10):1232-1240

ZHUANG Yan, LU Xibin, LI Yunhui, et al. Mobile robot indoor scene cognition using 3D laser scanning [J]. Acta Automatica Sinica, 2011, 37(10): 1232-1240

[3] 刘天亮,冯希龙,顾雁秋,等.一种由粗至精的 RGB-D 室内场景语义分割方法[J].东南大学学报(自然科学版),2016,46(4):681-687

LIU Tianliang, FENG Xilong, GU Yanqiu, et al. Coarse-to-fine semantic parsing method for RGB-D indoor scenes [J]. Journal of Southeast University (Natural Science Edition), 2016, 46(4): 681-687

[4] Nedovic V, Smeulders A W M, Redert A, et al. Depth information by stage classification [C] // IEEE 11th International Conference on Computer Vision, 2007: 1-8

[5] Hedau V, Hoiem D, Forsyth D. Recovering the spatial layout of cluttered rooms [C] // IEEE 12th International Conference on Computer Vision, 2009: 1849-1856

[6] Hoiem D, Efros A A, Hebert M. Geometric context from a single image [C] // Tenth IEEE International Conference on Computer Vision (ICCV'05), 2005: 654-661

[7] 许宏科,秦严严,陈会茹.一种基于改进 Canny 的边缘检测算法[J].红外技术,2014,36(3):210-214

- XU Hongke, QIN Yanyan, CHEN Huiru. An improved algorithm for edge detection based on Canny [J]. *Infrared Technology*, 2014, 36(3): 210-214
- [8] 梅雪, 夏良正, 李久贤, 等. 一种三维场景的消失点检测算法 [J]. *信号处理*, 2007, 23(6): 924-926.
- MEI Xue, XIA Liangzheng, LI Jiuxian, et al. A vanishing point detection algorithm for 3D scene [J]. *Signal Processing*, 2007, 23(6): 924-926
- [9] Lee D C, Hebert M, Kanade T. Geometric reasoning for single image structure recovery [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 2136-2143
- [10] Ramalingam S, Pillai J K, Jain A, et al. Manhattan junction catalogue for spatial reasoning of indoor scenes [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 3065-3072
- [11] Zhang J, Kan C, Schwing A G, et al. Estimating the 3D layout of indoor scenes and its clutter from depth sensors [C] // *IEEE International Conference on Computer Vision*, 2013: 1273-1280
- [12] Wang H Y, Gould S, Roller D. Discriminative learning with latent variables for cluttered indoor scene understanding [J]. *Communications of the ACM*, 2013, 56(4): 92-99
- [13] Schwing A G, Hazan T, Pollefeys M, et al. Efficient structured prediction for 3D indoor scene understanding [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 2815-2822
- [14] 吴培良, 李亚南, 杨芳, 等. 一种基于 CLM 的服务机器人室内功能区分类方法 [J]. *机器人*, 2018, 40(2): 188-194
- WU Peiliang, LI Yanan, YANG Fang, et al. A CLM-based method of indoor affordance areas classification for service robots [J]. *Robot*, 2018, 40(2): 188-194
- [15] Mallya A, Lazebnik S. Learning informative edge maps for indoor scene layout prediction [C] // *IEEE International Conference on Computer Vision (ICCV)*, 2015: 936-944
- [16] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015: 3431-3440
- [17] Dasgupta S, Fang K, Chen K, et al. DeLay: robust spatial layout estimation for cluttered indoor scenes [C] // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 616-624
- [18] Eigen D, Fergus R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture [C] // *IEEE International Conference on Computer Vision (ICCV)*, 2015: 2650-2658
- [19] Lee D C, Gupta A, Hebert M, et al. Estimating spatial layout of rooms using volumetric reasoning about objects and surfaces [C] // *24th Annual Conference on Neural Information Processing Systems*, 2010: 1288-1296
- [20] Tsochanaridis I, Joachims T, Hofmann T, et al. Large margin methods for structured and interdependent output variables [J]. *Journal of Machine Learning Research*, 2005, 6(2): 1453-1484

Indoor spatial layout estimation using informative edges and multi-modality features

LIU Tianliang¹ LU Panyu¹ DAI Xiubin¹ LIU Feng¹ LUO Jiebo²

¹ Jiangsu Provincial Key Lab of Image Processing and Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003

² Department of Computer Science, University of Rochester, Rochester 14627, USA

Abstract To perceive indoor spatial layout, we present a scene layout estimation method based on informative edges and multi-modality features. First, the VGG-16 full convolutional neural network is applied to predict informative edge map with the prior of spatial layout. Then, Canny edge detection and voting strategy are utilized to estimate the horizontal and vertical vanishing points, while the rays led at equal intervals from the given vanishing points finely resample the divided regions with high informative edge energies for the layout candidates. Next, the spatial multi-scaled VGG-16-based convolutional neural network is adopted to estimate the related geometric depth and normal vectors on the scene surfaces. And then, integral geometry is applied to accumulate the multi-model regional features as unary occurrence potential in the polygons of candidate layouts, and the pairwise label constrains are reflected by surface normal smooth and the location relationship of candidate layouts. Finally, the mode parameters can be learned by structural SVM learning, and the scene layout can be inferred by maximizing the related scores of the layout candidates. Experimental results show that, compared with traditional methods, this proposed estimation method can effectively improve the completeness of the resulting spatial layouts.

Key words spatial layout; convolutional neural network; scene understanding; informative edge