

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 路子赟, 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: TP391.41 文献标识码: A 文章编号: 1006-8961(2012)07-0839-07

论文引用格式: 武晖,于昕,隋尧,张利.融合上下文信息的场景结构恢复[J].中国图象图形学报,2012,17(7):839-845.

## 融合上下文信息的场景结构恢复

武晖,于昕,隋尧,张利

清华大学电子工程系,北京 100084

**摘要:**提出了一种融合场景上下文信息的两级分类算法,从单幅图像中恢复场景结构。室外场景的结构化特征使其3维结构可以粗略地分为3类:“地面”,“天空”以及“竖直物体”。首先,把图像分割成具有灰度和颜色一致性的区域;其次确定特征显著区域(“确定区域”)的结构,将特征不明显的区域标记为“未知区域”;然后根据“未知区域”与“确定区域”的相似性及“确定区域”场景结构对“未知区域”的可能结构进行投票,将投票最多的结构类型赋予“未知区域”;最后介绍场景结构恢复在构造场景3维模型方面的应用。实验结果表明,由于利用了场景结构的上下文信息,该算法场景结构恢复的正确率为92.3%,优于现有算法88.1%的恢复正确率。

**关键词:**场景理解;模式识别;图像分割;上下文信息

### Structure recovery algorithm using contextual information

Wu Hui, Yu Xin, Sui Yao, Zhang Li

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

**Abstract:** A two-level algorithm that is integrating contextual information to recover the structure of a single image is presented. Due to the structural features of outdoor scenes, we can classify the structure of a scene into three categories: sky, ground, and vertical objects. First, we over-segment the image into homogeneous regions. Then, we recognize the regions with significant features as “definite regions”, and the regions we can not classify as “undetermined regions”. Next, every nearby definite region with similar features as the undetermined region will vote for an undetermined region. The class with the most votes is assigned to that undetermined region. Finally, we construct a 3D model of the scene. Experiments show that due to the exploitation of the contextual information, almost 92.3% of the pixels can be recovered successfully, which is better than the performance of the existing method, whose result is 88.1%.

**Key words:** scene understanding; pattern recognition; image segmentation; contextual information

## 0 引言

图像是3维场景在成像平面的几何投影,从2维图像中恢复3维的场景结构是计算机视觉的主要研究任务<sup>[1]</sup>。人可以很轻易地从图像中获取实际场景的结构信息,比如场景中有什么物体,它们的位置关系等。但是对于计算机系统而言,从图像(尤

其是从单幅图像)中恢复3维场景结构仍然是一项富有挑战性的工作。

现有的3维重建算法大多基于立体视觉<sup>[2-3]</sup>原理,研究摄像机成像时的几何投影过程,根据两幅图像中匹配点之间的视差计算该点的空间位置<sup>[4-5]</sup>。运动恢复结构(structure from motion)<sup>[6]</sup>通过获取运动摄像机拍摄的同一场景的多幅图像来提取场景的结构信息。但是这些方法都需要获得不同位置或者

收稿日期:2011-09-15;修回日期:2012-01-10

基金项目:国家自然科学基金项目(61132007)

第一作者简介:武晖(1987—),男,清华大学电子工程系硕士研究生,主要研究方向为图像处理、计算机视觉等。E-mail: wuhui09@mails.tsinghua.edu.cn

不同角度拍摄的多幅图像,如果只有一幅图像,就需要利用其他算法来完成场景 3 维结构的恢复。

阴影恢复形状(shape from shading)<sup>[7]</sup>基于光度学原理,从一幅图像中恢复目标物的 3 维信息,但这种方法对实验环境要求很高,当光照条件不满足标准或者被测物体的颜色、纹理分布不均匀时不能得到满意的结果。Criminisi<sup>[8]</sup>基于投影几何约束提出了一种从单幅图像中计算指定点的空间位置的算法,需要人工指定地面上的一个正方形,一些和地面垂直的平行直线以及其他正交关系。“Tour into the Picture”<sup>[9]</sup>通过人工确定平行直线的消隐点,标记前景物体来构造场景的 3 维模型,实现场景漫游等虚拟现实功能,取得了不错的结果。文献[8-9]算法的共同不足在于需要大量的人工参与。最近,Hoiem<sup>[10-12]</sup>等学者从模式识别的角度出发,在室外场景高度结构化的前提下,通过大量训练样本的有监督学习,实现了对场景粗略 3 维结构的自动恢复。但是该识别算法只考虑了区域的局部特征,没有充分利用场景的上下文信息,因此场景结构的识别正确率不是很高,这是本文工作的出发点和需要解决的问题。

针对 Hoiem<sup>[11]</sup>算法正确率不高的问题,本文提出了一种新颖的两级分类算法,在分类的过程中融合了场景上下文信息,在第一级分类算法中确定那些特征明显、容易分类的区域的场景结构。在第二级分类算法中,利用场景结构的分布具有连续性和缓慢变化性这一特征,根据邻近区域的场景结构以及区域之间特征的相似度,确定出第一级分类算法中没有识别的区域的场景结构。实验结果表明该算法的场景结构识别正确率达到 92.3%,优于 Hoiem 算法 88.1% 的识别正确率。

## 1 问题阐述

因为不同的 3 维场景可能投影出相同的 2 维图

像,从几何投影理论来讲,由单幅图像恢复 3 维场景是一个病态问题,有无穷多个解。但是当看到一幅图像,人却可以很轻松地在大脑中形成原始的 3 维场景。这是因为世界是结构化的,有一定的规律性,人不仅掌握几何投影知识,也有强大的学习能力。正常情况下,天空是蓝色的,位于图像的上方。草地是绿色的,纹理比较规律。建筑物在重力的作用下直立于地面上。行人在地面上行走而不会跑到空中。这些规律都可以从图像本身得到。利用图像中丰富的有鉴别能力的特征,可以识别出无穷种可能性中最合理的 3 维场景结构。将室外场景的 3 维结构分为 3 类<sup>[10]</sup>:“地面”,包括马路、草坪、沙滩等平行于地平面的物体;还有位于地面之上,与地面成垂直关系的“直立物体”,包括建筑物,树木,行人等;第 3 种是“天空”。通过对大量室外场景图像的观察,Hoiem<sup>[10]</sup>发现这 3 种根据方向区分的结构类别可以在很大程度上表达一个室外场景的 3 维结构。恢复场景的 3 维结构就转化对图像内容进行分类的模式识别问题,基于这个假设,本文提出了两级识别算法,融合场景结构的上下文信息,解决了场景 3 维结构的恢复问题。本文的实验部分用数据说明了这个假设的合理性。

## 2 图像表示

图像以像素矩阵的形式保存。这种组织方式的缺点是只能利用每个像素的颜色信息和位置信息作为分类特征。为了提取具有鉴别能力的特征,同时降低分类算法的复杂度,将图像以超像素的形式组织。超像素是一类基于 Graph-Cut 的图像分割算法的统称,对定义在输入图像上的能量函数<sup>[13]</sup>进行最优化,从而将其分割成一些具有灰度和颜色一致性的离散的区域。图 1(b)为将图 1(a)分割成 234 个

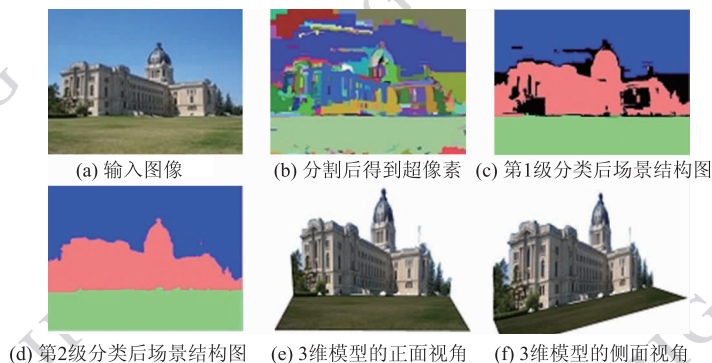


图 1 算法的完整流程和结果

Fig. 1 The overall procedure of the algorithm and the result

超像素,每个超像素用一个彩色区域表示。由于超像素内部具有一致性,可以认为它们在实际3维场景中属于相同的场景结构。因此,本文的分类算法是基于超像素的,确定了超像素的结构也就确定了像素的场景结构。将图像组织成超像素也可以减少需要进行分类的元素个数,图1(a)所示图像尺寸为 $550 \times 412$ ,如果直接对像素进行分类,将要执行226 600次分类操作,而以超像素为分类对象,只需234次,节省了大约1 000倍的分

### 3 特征选择

特征选择是模式识别问题的一个重要方面。直接关系到分类准确性。表1列出了提取的用于3维场景结构分类的特征。用于分类器训练的超像素的特征向量 $I = [C_1, C_2, C_3, P_1, S_1, T_1]$ 。

表1 基于超像素提取的分类特征

Table 1 Features extracted based on super pixels

| 特征种类 | 特征定义                    |
|------|-------------------------|
| 颜色特征 | $C_1$ HSV 颜色空间 H 分量的平均值 |
|      | $C_2$ HSV 颜色空间 S 分量的平均值 |
|      | $C_3$ HSV 颜色空间 V 分量的平均值 |
| 位置特征 | $P_1$ 归一化 Y 坐标值的平均值     |
| 尺寸特征 | $S_1$ 像素数量占图像像素总数量的比例   |
| 纹理特征 | $T_1$ 像素处的梯度模值的平均值      |

#### 3.1 颜色特征

室外场景包含丰富的颜色信息。一般情况下,天空是蓝色或者白色的,地面是绿色的(草地)或褐色的(马路)。颜色的这种规律性有助于区分场景结构,如红色区域属于“天空”的可能性很小,而属于“垂直物体”的概率就很大。存在很多颜色空间(RGB, HSV, CMY, ...), HSV 颜色空间用色度、饱和度和亮度来表示一种颜色,最符合人对颜色的直接生理感知,而颜色的 RGB 属性和人对颜色的感觉没有直接关系,所以提取 HSV 颜色空间中各个分量的平均值作为超像素的颜色特征,对应于表1中的3个特征 $C_1, C_2, C_3$ 。

#### 3.2 位置特征

室外场景是高度结构化的,天空位于建筑物和地面的上方,建筑物竖立于地面上。反映在图像中,天空位于图像的顶部,地面位于底部,而建

筑物等垂直物体位于中间。像素在图像中的高度值有助于判断其属于3种结构的哪一种,如图1(d)所示。用超像素中所有像素 Y 坐标的平均值作为特征可以提供较好的鉴别能力。由于不同图像存在尺寸差异,所以计算得到的 Y 坐标平均值要用图像高度进行归一化,得到表1中的特征 $P_1$ 。

#### 3.3 尺寸特征

图像是依据像素间的灰度和颜色的一致性进行分割的。天空和地面在灰度和颜色上变化不大,因此得到的超像素尺寸比较大,而垂直物体由于颜色灰度变化丰富,得到的超像素比较小,如图1(b)所示。这个直观的性质可以用超像素所包含的像素个数(超像素的面积)在图像的总像素个数(图像面积)中所占比例表示,对应表1中的 $S_1$ 。

#### 3.4 纹理特征

本算法采用基于 Sobel 算子的像素梯度模值表示图像纹理特征。像素梯度是图像纹理变化大小的一种度量,使用像素梯度作为局部特征可以识别纹理不同的区域。因为天空灰度变化缓慢,所以梯度模值比较小,而地面和垂直物体纹理丰富,梯度模值大。定义超像素的纹理特征 $T_1$ 为所有像素处梯度模值的平均值。

## 4 分类器训练

分类器设计是模式识别的核心。本文使用基于决策树的 Adaboost<sup>[14-17]</sup>分类器。Adaboost 算法通过迭代,对 $T$ 个弱分类器构成的集合 $\{h_t(I), t=1, \dots, T\}$ 进行自适应的提升( $I$ 为待分类的特征向量),计算每个弱分类器的权重 $\{a_t\}$ ,得到一个强分类器

$$H_{Ada}(I) = \text{sgn} \left( \sum_{t=1}^T a_t h_t(I) \right) \quad (1)$$

权重是通过对训练样本的有监督学习得到的。对于标准的 Adaboost 分类器,其输出为

$$H_{Ada}(I) = \begin{cases} 1 & \text{样本 } I \text{ 属于此类别} \\ -1 & \text{样本 } I \text{ 不属于此类别} \end{cases} \quad (2)$$

标准 Adaboost 分类器只能处理两类标签问题。本文中需要将样本分到3个类别,因此训练4个 Adaboost 分类器: $H_{sky}(I), H_{ground}(I), H_{vertical}(I), H_{same\_label}(I, J)$ ( $J$ 为待分类的特征向量),分别对“天空”,“地面”,“垂直物体”3种类别进行分类,以

及判断两个样本是不是属于同一类别。这4个分类器的功能如式(3)–(6)所示。

$$H_{\text{sky}}(I) = \begin{cases} 1 & I \text{ 属于“天空”} \\ -1 & I \text{ 不属于“天空”} \end{cases} \quad (3)$$

$$H_{\text{ground}}(I) = \begin{cases} 1 & I \text{ 属于“地面”} \\ -1 & I \text{ 不属于“地面”} \end{cases} \quad (4)$$

$$H_{\text{vertical}}(I, J) = \begin{cases} 1 & I, J \text{ 属于“垂直物体”} \\ -1 & I, J \text{ 不属于垂直物体} \end{cases} \quad (5)$$

$$H_{\text{same\_label}}(I, J) = \begin{cases} 1 & I, J \text{ 属于相同类别} \\ -1 & I, J \text{ 不属于相同类别} \end{cases} \quad (6)$$

前3种分类器的训练过程如下:

输入:  $P_1, P_2, \dots, P_N$  室外场景图像的训练集合。

输出:  $H_{\text{sky}}(I), H_{\text{ground}}(I), H_{\text{vertical}}(I)$ 。

1) for  $i = 1, \dots, N$

将  $P_i$  分割成超像素  $S_{ij} (j \in \{1, \dots, M\}, M$  是图像  $P_i$  的超像素数量), 计算每个超像素  $S_{ij}$  的特征向量  $I_{ij}$ 。

end

2) 所有样本标记为“天空”和“非天空”两类, 训练  $H_{\text{sky}}(I)$ 。

3) 所有样本标记为“地面”和“非地面”两类, 训练  $H_{\text{ground}}(I)$ 。

4) 所有样本标记为“垂直物体”和“非垂直物体”两类, 训练  $H_{\text{vertical}}(I)$ 。

前3个分类器只考虑单个超像素, 忽略了场景结构分布的上下文信息。场景上下文信息表现在两个方面: 首先, 相同类别的超像素的特征向量在特征空间内的距离比不同类别的特征向量之间的距离小, 这是分类器工作的前提; 其次, 在物理空间上越接近的像素, 它们属于同一场景结构的概率越大, 因为场景结构的分布具有连续性和缓慢变化性, 图1(d)可以清楚地说明这一点。

基于以上讨论的场景上下文信息, 设计了“类别一致性”分类器  $H_{\text{same\_label}}(I, J), H_{\text{same\_label}}(I, J) = 1$  表示超像素  $I$  和超像素  $J$  的特征向量比较接近, 属于同一场景结构类别,  $H_{\text{same\_label}}(I, J) = -1$  则表示超像素  $I$  和超像素  $J$  的特征向量差异很大, 不属于同一类场景结构。

“类别一致性”分类器  $H_{\text{same\_label}}(I, J)$  的训练过程如下:

输入:  $P_1, P_2, \dots, P_N$  室外场景图像的训练集合。

输出:  $H_{\text{same\_label}}(I, J)$ 。

1) for  $i = 1, \dots, N$

将  $P_i$  分割成超像素  $S_{ij} (j \in \{1, \dots, M\}, M$  是图像  $P_i$  的超像素数量), 计算每个超像素  $S_{ij}$  的特征向量  $I_{ij}$ 。

end

2) 对属于同一类别的特征向量  $I$  和  $J$ , 计算新的特征向量  $K = |I - J|$ , 标记为正样本。

3) 对不属于同一类别的特征向量  $I'$  和  $J'$ , 计算新的特征向量  $K' = |I' - J'|$ , 标记为负样本。

4) 用正样本和负样本作为训练样本训练  $H_{\text{same\_label}}(I, J)$ 。

## 5 两级分类算法描述

为了提高场景3维结构恢复的正确率, 本文提出了利用场景结构上下文信息的分类算法分两步对超像素的类别进行判定。

### 5.1 第1级分类算法

利用超像素的局部信息提取特征, 使用  $H_{\text{sky}}(I), H_{\text{ground}}(I), H_{\text{vertical}}(I)$  进行初步分类。有些超像素的特征比较明显, 容易分类, 称其为“确定区域”。也有一些超像素特征不明显, 无法确定其场景类别, 这些超像素称为“未知区域”, 暂时不赋予类别标签。

具体来说, 对于图像中的每一个超像素  $I$ , 用  $H_{\text{sky}}(I)$  进行分类得到分类标签  $Label\_Sky \in \{1, -1\}$ ,  $Label\_Sky$  为1说明  $I$  属于“天空”, 为-1说明  $I$  不属于“天空”。同样用  $H_{\text{ground}}(I)$  进行分类得  $Label\_Ground \in \{1, -1\}$ , 用  $H_{\text{vertical}}(I)$  进行分类得到  $Label\_Vertical \in \{1, -1\}$ 。为了保证分类结果的准确性, 只有当3个类别标签中的1个为1, 其他两个为-1的时候, 才能确定该超像素的场景结构类别为分类标签是1的那个场景结构, 这个超像素属于“确定区域”。如果有超过两个类别标签为1, 说明  $I$  可能属于两类不同的结构, 需要在第2级分类算法中利用场景上下文信息进行识别。这些超像素为“未知区域”, 无法仅仅通过区域本身的信息来进行分类。表2是第1级分类算法的判定规则。这一规则背后的思想是: 如果像素属于某一场景结构, 那么他不可能同时属于其他两种场景。在进行分类的时候, 将同时属于多种类别的像素认为是“未知区域”, 暂时无法确定类别。而对那些明确属于某一类别的区域直接进行分类。

表2 第1级分类算法的分类规则  
Table 2 Classification rule of the first level of the algorithm

| Label_Sky | Label_Ground | Label_Vertical | 超像素的类别 |
|-----------|--------------|----------------|--------|
| 1         | -1           | -1             | 天空     |
| -1        | 1            | -1             | 地面     |
| -1        | -1           | 1              | 竖直物体   |
| 其他情况      |              |                | 未知区域   |

## 5.2 第2级分类算法

第2级分类算法利用场景上下文信息对“未知区域”分类。“未知区域”是那些特征不明显,在第一级分类算法中无法明确分类的区域。但是因为场景结构的分布具有连续性和缓慢变化性,可以通过其周围区域的分类结果判断该区域的场景结构。与“未知区域”邻近,并且特征相似的区域和其属于同一类场景结构的可能性较大,对决定“未知区域”类别的影响也比较大,而那些与“未知区域”距离较远,或者特征相差很大的区域则不能作为“未知区域”分类的参考。特征相似与否通过分类器  $H_{\text{same\_label}}(I, J)$  来判断。找出与“未知区域” $I$  距离最近的  $N$  个“确定区域” $J_k (k \in \{1, \dots, N\})$ , 其类别是  $Label_{J_k}$ , 用  $H_{\text{same\_label}}(I, J)$  判断  $I$  和  $J_k$  是否属于同一类, 如果属于同一类, 则“未知区域” $I$  就可能属于类别  $Label_{J_k}$ , 这等价  $J_k$  给“未知区域属于  $Label_{J_k}$ ”这一论断投了一票, 如果  $I$  和  $J_k$  不属于同一类, 则  $J_k$  不能提供  $I$  的类别信息, 没有投票权力。统计对于“未知区域” $I$ , 3 种类别分别获得的投票数目, 投票数最高的场景结构类别就是  $I$  的类别。算法描述如下:

输入:  $H_{\text{same\_label}}(I, J)$ ;

$N$ : “未知区域”的邻域大小;

$I$ : “未知区域”的特征向量。

输出:  $I$  的类别  $Label$ 。

1) 取和“位置区域”的距离最近的  $N$  个“确定区域”。

2) for  $k = 1, \dots, N$

对于得到的每一个“确定区域” $J$ , 其类别为  $Label_J$ 。

用  $H_{\text{same\_label}}(I, J)$  对  $I$  和  $J$  分类。

if  $I$  和  $J$  属于同一类场景结构类别  $Label_i$  的投票数目加 1。

end

end

3) 统计 3 种场景的得票数, 得票最多的为  $I$  的类别  $Label$ 。

## 6 场景结构恢复的应用

判断出所有像素的场景结构之后, 可以构建基于纹理贴图的简单 3 维模型。去掉天空, 恢复出建筑物和树木等竖直物体的轮廓, 如图 2(b) 所示。由场景的“地面”结构也可以通过直线拟合检测出地平线, 如图 2(a)(b) 中白线所示。最后, 沿地平线折叠图像让“竖直物体”站立起来, 产生 3 维效果。图 1(e)(f) 为从两个不同角度观察到的 3 维模型。图 3 为对图像进行场景结构恢复并创建 3 维模型的结果。

## 7 实验结果与分析

对本文算法在 Hoiem<sup>[11]</sup> 数据集上进行实验。实验计算机的配置为 4 G 内存, 2.70 GHz 奔腾 E5400 CPU, Windows 7 操作系统。实验比较了该算法与 Hoiem<sup>[11]</sup> 算法的识别正确率, 并研究了不同特征对分类的作用。实验结果表明本文提出的算法优于 Hoiem<sup>[11]</sup> 提出的基于多重分割的识别算法。

### 7.1 两级分类算法的识别正确率

用 Hoiem<sup>[11]</sup> 提供的图像训练分类器并测试识别算法, 训练集有 83 张图片, 提取出 16 600 个经过标记的超像素作为训练样本。测试样本有 63 张图片。识别正确率定义为正确恢复场景结构的像素占有所有像素的百分比。实验结果见表 3。图 4 选取两组实验作为比较。在第 1 张图像中, Hoiem 算法将一部分建筑物误判为天空, 在第 2 张图像中, Hoiem 算法错误地将一部分天空分类为“垂直物体”, 而本文算法成功地恢复了这些像素的场景结构类型。

实验结果表明, 两级分类算法在识别正确率上优于 Hoiem 算法, 因为本文算法不仅利用了图像的局部信息, 也利用了场景的全局信息。其次, 比较第 1 级分类算法和第 2 类分类算法可以发现, 将近 20% 的像素由于考虑了场景上下文信息而被正确分类, 这表明场景的上下文是很重要的分类信息。

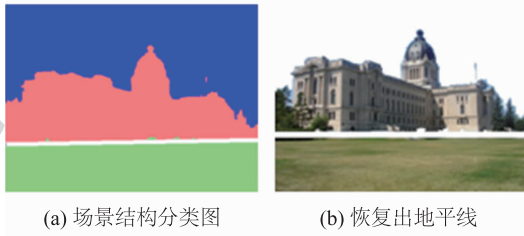


图 2 利用场景结构恢复地平线

Fig. 2 Recovering horizon using scene structure

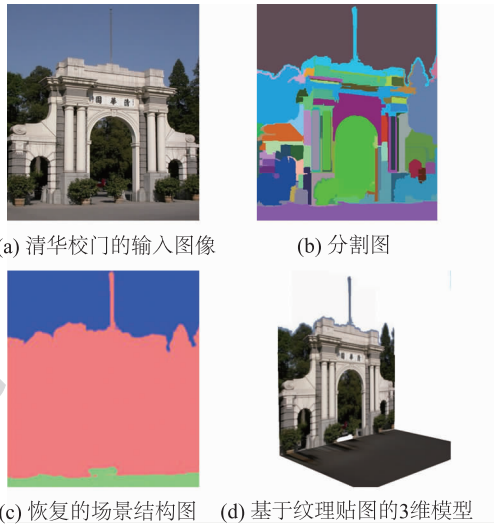


图 3 清华校门的场景结构以及创建的 3 维模型

Fig. 3 Scene structure and the 3D model of Tsinghua gate

表 3 本文算法和 Hoiem 算法的正确率比较

Table 3 Comparison between our method and Hoiem's

|       | 第 1 级分类算法 | 第 2 级分类算法 | Hoiem 算法 |
|-------|-----------|-----------|----------|
| 正确率/% | 70.6      | 92.3      | 88.1     |

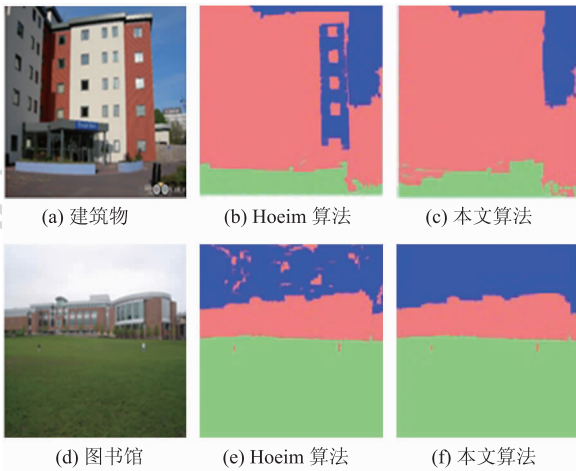


图 4 本文算法和 Hoiem 算法比较

Fig. 4 The comparison of Hoiem's algorithm and our algorithm

### 7.2 图像特征对识别正确率的影响

第 3 节确定了用于分类的 4 种特征,但是并没有用实验数据证明这些特征的有效性以及各个特征如何影响最终分类结果。

进行两组实验,第 1 组每次只使用 1 种特征,第 2 组每次都去掉 1 个特征,用剩余的 3 个特征进行分类。实验结果如表 4 所示。可以发现 4 种特征在场景结构恢复中都有重要作用,而其中位置最重要,尺寸其次,颜色再次,而纹理特征的作用最小,这也印证了第 3 节的讨论。在只用一个特征进行分类时,分类效果也比较理想,这说明每个特征都有着很强的场景结构鉴别能力。实验同时也验证了在问题阐述部分提出的假设:室外环境是高度结构化的,由有限多种场景结构组成,不同的场景结构有很大的差异,可以提取特征进行分类。这也使利用模式识别来解决场景恢复成为可能。

表 4 使用不同特征时的分类正确率

Table 4 The classification result using different features

| 特征类型   | 分类正确率/% |
|--------|---------|
| 所有特征   | 92.3    |
| 只有颜色特征 | 79.1    |
| 只有尺寸特征 | 80.9    |
| 只有纹理特征 | 74.1    |
| 只有位置特征 | 85.9    |
| 没有颜色特征 | 89.6    |
| 没有尺寸特征 | 87.7    |
| 没有纹理特征 | 90.2    |
| 没有位置特征 | 86.4    |

## 8 结 论

提出一种融合场景上下文信息的两级分类算法,从单幅图像中恢复场景 3 维结构。首先确定具有显著特征的区域场景结构,然后根据已经识别的区域,利用场景分布的连续性和缓慢变化性,最终识别出所有区域的场景结构。最后,利用恢复出来的场景结构,使用基于图形的渲染技术,对 3 维场景进行简单的建模。实验结果表明了这种识别算法相比于已提出的算法,将识别率从 88.1% 提高到了 92.3%。由于目前场景结构类别只

有3种,没有对“垂直物体”进一步分类,所以恢复出来的3维场景结构只是粗略的,如何恢复出更多种类的3维场景结构,是下一步工作的方向。

### 参考文献(References)

- [1] Marr D. Vision [M]. San Francisco: Freeman, 1982: 235-257.
- [2] Hartley, Zisserman A. Multiple View Geometry in Computer Vision[M]. Cambridge, UK: Cambridge University Press, 2004: 56-60.
- [3] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. International Journal of Computer Vision, 2001, 47(1-3):7-42.
- [4] Pollefeys M, Koch R, Van Gool L. Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters[C] // Proceedings of ICCV. Bombay, India: IEEE Computer Society, 1998: 90-96.
- [5] Forsyth D A, Ponce J. Computer Vision: A Modern Approach [M]. London: Prentice Hall, 2003: 123-125.
- [6] Dellaert F, Seitz S, Thorpe C, Thrun S. Structure from motion without correspondence [C] // Proceedings of CVPR. Hilton Head, SC, USA: IEEE Computer Society, 2000:557-564.
- [7] Zhang R, Tsai P, Cryer J, et al. Shape from shading: a survey [J]. IEEE Trans. on PAMI, 1999, 21(8):690-700.
- [8] Criminisi, Reid, Zisserman. Single view metrology[J]. International Journal of Computer Vision, 2000, 40(2): 123-148.
- [9] Horry Y, Anjo, Arai K. Tour into the picture: using a spidery mesh interface to make animation from a single image [C] // ACM SIGGRAPH. Los Angeles, USA: ACM Press, 1997:225-232.
- [10] Hoiem D, Efron A, Hebert M. Automatic photo pop-up [C] // ACM SIGGRAPH. Los Angeles, USA: ACM Press, 2005:577-584.
- [11] Hoiem D, Efron A, Hebert M. Geometric context from a single image [C] // Proceedings of ICCV. Beijing, China: IEEE Computer Society, 2005:654-661.
- [12] Hoiem D, Efron A, Hebert M. Putting objects in perspective [C] // Proceedings of CVPR. New York, USA: IEEE Computer Society, 2006: 2137-2144.
- [13] Felzenszwalb P, Huttenlocher D. Efficient graph-based image segmentation [J]. International Journal of Computer Vision, 2004, 59(2): 167-181.
- [14] Freund Y, Schapire R. Experiments with a new boosting algorithm [C] // Proceedings of ICML. Bari, Italy: IEEE Computer Society, 1996:148-156.
- [15] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting [J]. The Annals of Statistics, 2000, 38(2):337-374.
- [16] Viola P, Jones M. Fast and robust classification, using asymmetric adaBoost and a detector cascade [C] // Proceedings of NIPS. Vancouver, British Columbia, Canada: IEEE Computer Society, 2001:1311-1318.
- [17] Collins M, Schapire R, Singer Y. Logistic regression, adaboost and bregman distances [J]. Machine Learning, 2002, 48(1-3): 253-185.