

# 采用几何复杂度的室外场景图像分割和深度生成

任艳楠<sup>1</sup> 刘 琚<sup>1,2</sup> 元 辉<sup>1,2</sup> 顾凌晨<sup>1</sup>

(1. 山东大学信息科学与工程学院, 山东济南 250100;

2. 山东大学苏州研究院, 江苏苏州 215021)

**摘 要:** 本文提出一种采用几何复杂度的室外场景图像几何分割和深度生成算法。该算法首先通过图像中主要线段的角度统计分布将室外场景图像的几何结构规划为四种类型; 然后, 利用 meanshift 分割算法将输入图像分割成若干小区域, 依据该图像的场景几何结构将这些小的区域逐步融合成为三个大的区域, 每个区域具有一致的深度分布特点, 由此实现输入图像的几何分割; 最后, 根据几何类型定义标准的深度图, 结合输入图像的几何分割结果获得图像的深度图。实验结果表明可以通过简单的线段角度统计分布实现图像的几何分割, 并进一步获得图像的深度图, 与已有算法相比, 提出的算法可以更好地保持深度图细节, 更接近场景的真实的深度信息。

**关键词:** 图像分割; 深度图; 几何; 融合

中图分类号: TN911.73

文献标识码: A

DOI: 10.16798/j.issn.1003-0530.2018.05.004

## Outdoor Image Segmentation and Depth Generation Based on Geometry Complexity

REN Yan-nan<sup>1</sup> LIU Ju<sup>1,2</sup> YUAN Hui<sup>1,2</sup> GU Ling-chen<sup>1</sup>

(1. School of Information Science and Engineering, Shandong University, Jinan, Shandong 250100, China;

2. Suzhou Research Institute of Shandong University, Suzhou, Jiangsu 215021, China)

**Abstract:** The paper presents an image segmentation and depth generation algorithm based on geometry complexity which is applied to outdoor scenes. Firstly, the angle statistical distribution of main lines in the input image is calculated and then the outdoor scenes are classified into four geometric categories. Secondly, the input image is divided into many small regions using the mean-shift segmentation algorithm, and then these regions are merged into three big regions based on the scene geometry category results. Each big region is in coherent depth distribution. Thirdly, a depth map can be generated based on the geometry segmented result and its standard depth map. Experimental results shows that proposed method can obtain an effective image geometric segmentation and meanwhile get a depth map with better details and more close to the true depth information of the scene.

**Key words:** image segmentation; depth generation; geometric; fusion

## 1 引言

近年来,随着三维视频资源需求的快速增长,如何将已有的二维图像重建为三维图像显得尤为重要<sup>[1,19-20]</sup>。重建三维图像首先需要获得场景的深

度信息,因此,从二维图像的结构信息中推测出场景的深度信息是产生三维视频的重要途径。这些二维视频普遍都没有原始的深度数据或者三维信息。然而,人类视觉系统对于实际场景对象之间的距离远近具有相对性的认知特点,通过分析场景图

片中各个对象的特点并结合深度因素信息,然后分配合适的深度数据以获得场景的深度图。这些深度数据虽然不精确,但在二维转换为三维视频图像时,也可以有效呈现场景的三维效果。

当图像具有一定的深度线索时,例如聚焦、去雾、线性透视和遮挡等,就可以从该图像中估算出深度信息<sup>[2-5]</sup>,但该研究只适用于具有强烈深度因素的场景。有些深度估计算法先建立深度图像和对应彩色图像的图像库,然后利用建模优化算法获得深度图<sup>[6,7,9]</sup>,此类型算法有了更为灵活的应用空间。近年来,采用深度学习的算法<sup>[8]</sup>也可以产生不同室外场景结构的深度图,但这些算法都需要有较大的计算量。

在常见的室外人工场景视图中,由于建筑设计时所具有的几何特性及拍摄角度限制,图像中的某些对象区域保留了非常明显的几何特征,依据这些几何特征,也可以估算出场景的深度信息。D. Hoiem<sup>[10]</sup>指出,在忽略场景的细节时,图像的整体呈现出统一的几何结构。在文献[11]中,作者将常见的视频场景归纳为有限类型,并用一些典型的三维几何结构来代表一般场景。文献[12]将室外三维场景结构进一步描述为几何场景分类问题,将常见的几何场景归纳为12类,并设定了每种类型的标准深度分布图。进一步,文献[13]将这个标准深度分布图作为初始深度图,然后结合图像分割术获得了输入图像的深度图。

大部分的室外场景包含天空、地面以及建筑物三个部分。从已经获得的深度图来分析,天空区域距离最远,其深度值可以视为一个恒定值;地面的深度则是由近及远变化;而对于垂直于地面的建筑物,它们的深度值可以通过参考建筑物与地面交点的深度值获得。由于天空、地面、建筑物这三者具有完全不同的深度特征,所以可以将此类图像分割为三个不同的区域,每个区域内部的深度值变化规律保持一致。

由此,本文提出了一种基于几何复杂度的图像分割和深度生成算法。首先定义4种场景类型,通过分析图像的几何结构,即用图像中主要线段的角度统计分布直方图,来判断输入图像的场景类型;然后依据场景类型,将输入的单幅图像分割成有限的几个区域,进而估计出场景的深度信息。

## 2 场景分类及判断

文献[12]考虑到了人工建筑的几何特性,忽略了场景中的人物等细节对象,提出了两种深度轮廓模型,即横坐标深度轮廓建模(Abscissa Depth Profile Model, ADPM)和纵坐标深度轮廓建模(Ordinate Depth Profile Model, ODPM),前者可以得到水平、左、右和中四种几何结构,后者得到上、中、下三种结构类型,综合可以将场景几何类型进一步细分为12种。本文着重于研究室外人工建筑物场景,采用了ADPM的四种结构类型,如图1所示。

在图1中,前两行是场景实例,包含了天空、地面和人工建筑物。在这四种几何类型中,天空和地面区域深度特征一致,几何特性也一样,而建筑物的几何特征有所区别。从图中可以看到,(a)列的建筑物处于地面的远端,建筑物整体深度值恒定不变;(b)和(c)中的建筑物沿着地面向左或者向右延伸,此时建筑物整体的深度值跟随着地面发生一致变化,(d)中建筑物的左右两侧分别沿着地面聚焦于中间区域,左右两侧的深度沿着地面向中间区域发生一致变化。由此可以看出,建筑物在场景中的几何结构特征决定了整个场景的几何类型,因而在图1的第三行的场景的几何结构图中只显示了建筑物的几何特点。

一般来说,图片中的人工建筑存在很多线段。众多的线段会呈现出了不同的方向角度特点。因此,场景中主要线段的方向角度可以有效地体现出场景的几何特性和深度特点。

对于输入图像 $I$ ,采用Hough变换<sup>[14]</sup>获得 $I$ 中的有效线段 $l_m \in S_l$ ,其中 $m$ 是线段的编号, $S_l$ 是 $I$ 中线段的集合,其中 $m \leq M$ , $M$ 是 $I$ 中线段的个数;线段的方向角度为 $\theta_m \in S_\theta$ ,分布在 $[-90^\circ, 90^\circ]$ 内。为了便于说明,将这些线段重投影于 $[0^\circ, 180^\circ]$ 间,并定义为 $\theta'_m$ ,并将 $[0^\circ, 180^\circ]$ 分成18个区间。令 $k$ 为角度区间编号, $1 \leq k \leq 18$ ,即其中每个区间定义为 $[(k-1) \times 10^\circ, (k-1) \times 10^\circ + 9^\circ]$ ,则分别为 $[0^\circ, 9^\circ]$ ,  $[10^\circ, 19^\circ]$ ,  $[20^\circ, 29^\circ]$ , ...,  $[160^\circ, 169^\circ]$ ,  $[170^\circ, 179^\circ]$ ,则 $\theta'_m$ 在第 $k$ 区间上出现的数量为 $N_k$ ,若定义 $\theta'_m$ 总数为 $N$ ,则由公式(1)可得 $\theta'_m$ 在 $k$ 区间的频率 $f_k$

$$f_k = \frac{N_k}{N} \quad (1)$$

由此,可以采用极坐标的形式给出图像中线段的角

度统计分布,如图 2 所示。其中,第一行的图像分别对应了四种不同类型的几何场景,第二行是每幅图像中的线段角度统计分布图,其中蓝色线为角度统计分布。

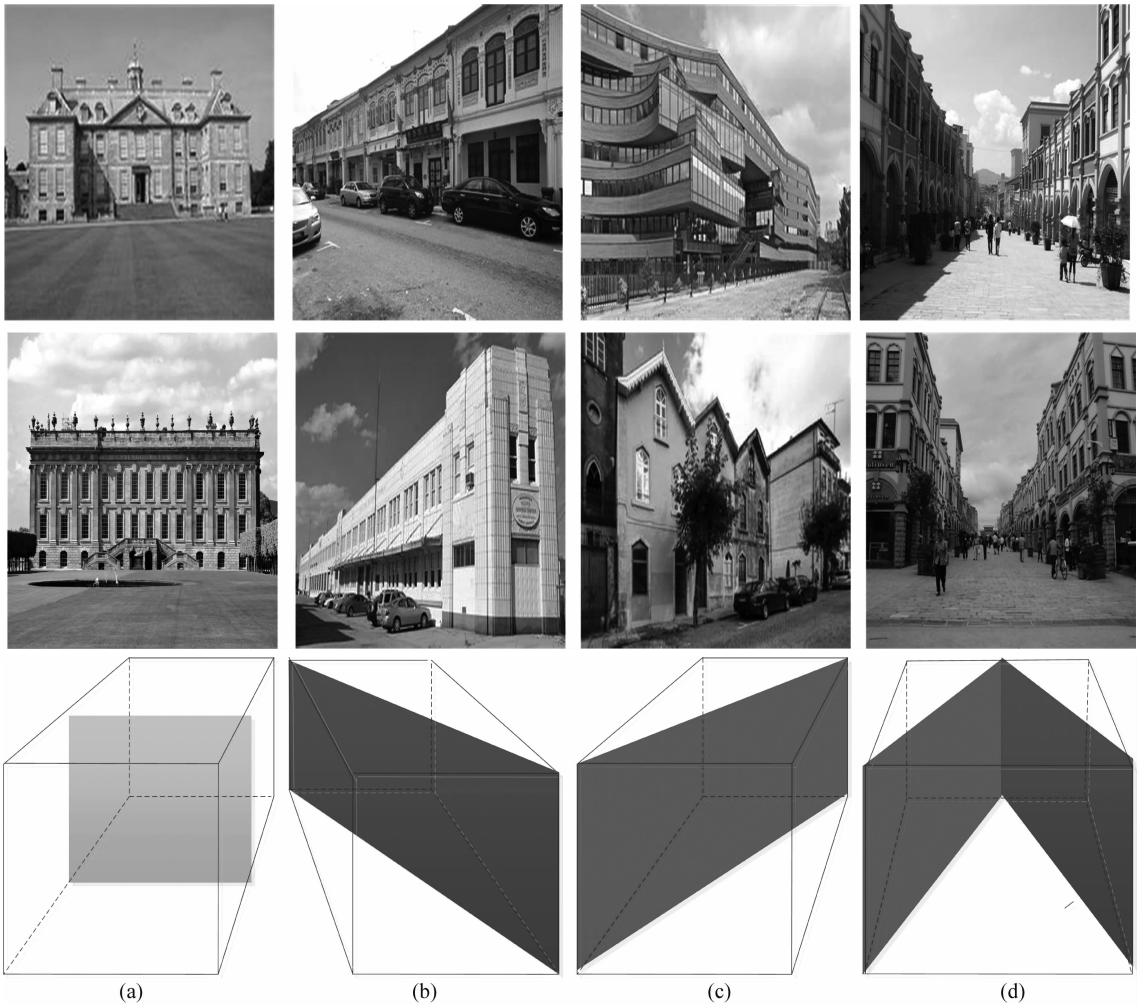


图 1 常见场景几何结构

Fig. 1 Geometric structures of main outdoor scenes

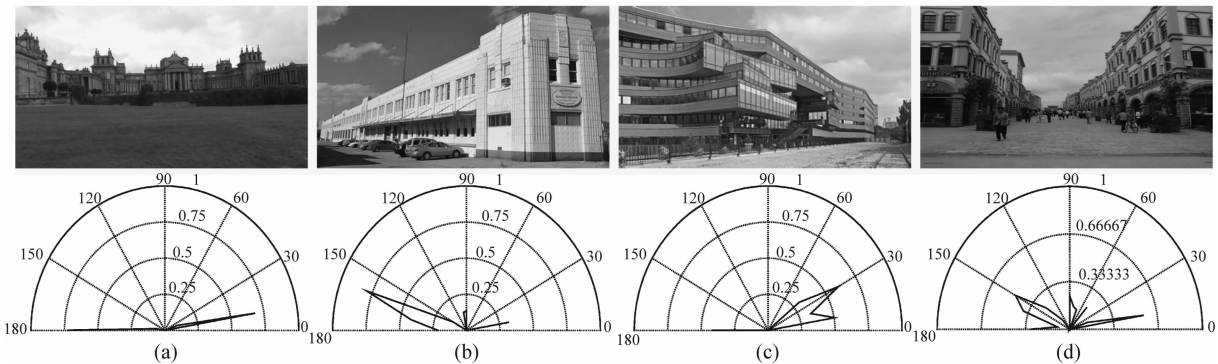


图 2 角度统计分布图

Fig. 2 Angle statistical distribution maps

从图2可以看出,具有第一类型特点的场景,图中提取的线段主要是水平线,角度主要分布趋近于 $0^\circ$ 和 $179^\circ$ ;具有第二类型特点的场景,图中提取的线段角度集中在 $[90^\circ, 179^\circ]$ ;相应地,具有第三类型特点的场景,线段的角度分布主要集中在 $[0^\circ, 90^\circ]$ ;具有第四类型特点的场景,其线段的角分布在上述两个区间都有。由此可以看到,四种不同的深度类型图所具有的线段角度统计分布也是不一样的,因此可以有效地表现出图像的场景结构。

### 3 基于几何复杂度的图像分割和深度生成

在确定了输入图像的几何类型后,我们提出基于几何复杂度的图像分割,即依据图像的几何类型将输入图像分割成天空、建筑物和地面三个区域,每个区域具有一致的深度分布;然后再将这个结果应用于深度生成算法中,结合标准的场景深度图,进而获得高质量深度图。

#### 3.1 基于几何复杂度的图像分割

图像分割技术指将目标图像分割成一些子区域的算法。起初,图像分割方法通常将图像分成许多小的区域,区域内的像素点包含相似的颜色特征,而区域间存在较大的颜色差异,代表算法有meanshift聚类<sup>[15]</sup>,分水岭算法<sup>[16]</sup>,水平集<sup>[17]</sup>等;后来借助于半监督方法<sup>[18]</sup>或者建模识别算法,这些分割的小区域通过区域融合技术,被进一步有目的地融合在一起,将输入图像分割成有限的几个区域,每个区域代表了有实际意义的语义目标对象;与[18]相似,本文在确定了输入图像的几何类型后,将这些分割的小区域进一步融合成三个大的区域,分别对应天空、建筑物和地面三个部分,由此获得基于几何复杂度的图像分割结果。具体的过程如图3所示。

在应用几何复杂度的图像分割时,首先采用meanshift算法将输入图像分割成许多小区域,并将所有的区域组成的集合定义为 $S_r$ ,其中的第 $n$ 个区域表示为 $r_n \in S_r$ ,  $n \leq N$ ,而 $N$ 是输入图像所包含的小区域的总数目。

利用前面场景分类算法的结果,采用Hough方

法产生的有效线段和图像中的先验知识,我们可以获得一些区域的初始标记,即基于以下准则获得初始标记图。

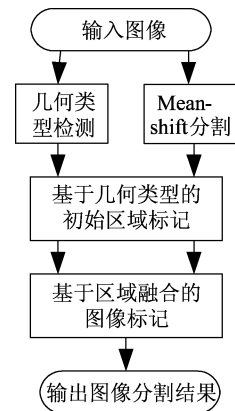


图3 系统流程图

Fig.3 Flowchart of the proposed system

1) 由场景类型选取有效线段,将包含了该线段的区域被标记为建筑物区域;当图像的场景结构被判断为类型1时,则主要提取角度趋近于水平的线段;属于类型2时,则主要提取角度在 $[20^\circ, 70^\circ]$ 的线段;属于类型3时,则主要提取角度在 $[120^\circ, 160^\circ]$ 的线段;属于类型4时,提取角度在两个区间的线段;一般来说,主要提取角度频繁出现的线段;确定所需要的有效线段后,将线段端点所在的区域设定为“建筑物”;

2) 依据以下先验知识来标记出“天空”和“地面”区域,即图像的上部一般都是天空,下部区域一般都是地面,所以设定图像上部边缘包含最多像点的区域标记为“天空”,设定图像下部边缘包含最多像点的区域标记为“地面”。

当融合不同的区域时,定义每个区域的RGB颜色空间的直方图作为其特征向量。像点包含了RGB三个颜色分量,且其灰度值均在 $[0, 255]$ 之间,将其统一量化为16个灰度等级,然后再将三个颜色分量连接起来,则将此三个颜色值融合为一个值,其范围则为 $0 \sim 4096$ ,并将其视为该像点的像素值。此时再计算该区域的像点直方图向量作为该区间的特征向量。

对于未标定区域 $r_j$ ,其相邻区域集定义为 $SN_i$ 。当其中的某个邻域满足如下公式(2)时,从 $SN_i$ 中

得到与  $r_j$  最为相似的  $r_k$ ,

$$r_k = \arg \max_{r_j \in SN_i} \{ \rho(r_j, r_i) \} = \arg \max_{r_j \in SN_i} \left\{ \sum_{m=1}^{4096} \sqrt{v_j^m \cdot v_i^m} \right\} \quad (2)$$

其中  $\rho(r_j, r_i)$  为巴氏系数<sup>[21]</sup>,  $v_j$  和  $v_i$  分别是  $r_j$  和  $r_i$  的特征向量, 上标  $m$  代表了该特征向量的第  $m$  个分量。巴氏系数定义度量不同区域之间的距离, 该系数数值越大, 说明两者越相似。

然后再将  $r_j$  和  $r_k$  融合成一个新区域, 重新计算该区域的特征向量, 设置该区域的标记, 并重新设置区域集以及区域的邻域集, 重复执行该过程, 一直到所有的区域被标记完毕, 由此获得最后的标记图, 结果如下图4中所示。其中, 第一行是四种类型的输入图像, 第二行是它们所对应的基于几何复杂度的图像分割结果, 其中青绿色区域代表天空部分, 黄色区域代表了图像中的主要建筑物部分, 而深蓝色代表了地面区域。与输入图相比, 基于几何复杂度的分割算法将原图像有效地分割为了三个区域, 即天空、建筑物和地面。

### 3.2 基于几何复杂度的深度生成

在获得了基于几何复杂度的图像分割结果后, 根据每个区域的几何特点来分配合适的深度值, 进而获得体现整体几何特点的深度图, 定义为  $D_g$ 。同时, 考虑到图像归属的几何类型有着标准的场景深度图, 结合颜色分割结果, 为每个小区域分配深度

值, 由此获得体现细节的深度图, 定义为  $D_r$ , 综合两者的结果生成最后的深度图。

在输入图像被自动分割成天空、建筑物和地面三个区域, 每个区域具有一致的深度分布; 此时, 设定标记为天空区域的深度值为固定深度值; 标记地面区域, 采取由近及远的原则设定深度; 对于标记为建筑物的区域, 像点的深度值为在该像点具有同一列坐标的距离该像点最近的标记为地面的像点的深度值, 由此获得深度图  $D_g$ 。

在判断出输入图像的几何类型时, 我们引用了4种ADPM几何场景, 其标准化深度图如下图5所示。

假设输入图像  $I$  属于场景类型  $c$ ,  $c \in \{1, 2, 3, 4\}$ , 则其对应的标准深度图为  $D_c$ , 用 *meanshift* 算法将输入图像分割成的每个区域  $r_n \in S_r$  中, 任一个像点  $p_k \in r_n$ ,  $k \leq N(r_n)$ , 其中  $N(r_n)$  是  $r_n$  中的像素数目, 则可根据公式(3)计算获得  $r_n$  中像素  $p_k$  的深度值, 也就是该区域中每个像素的深度值, 由此可以得到保留细节的深度图  $D_r$ ;

$$D_r(p_k) = \frac{\sum_{k=1}^{N(r_n)} D_c(p_k)}{N(r_n)} \quad (3)$$

综合上述两个深度图  $D_r$  和  $D_g$  来获得最后的深度图  $D$  如下式(4)所示, 并采用联合滤波器平滑保留细节。

$$D = \frac{(D_r + D_g)}{2} \quad (4)$$

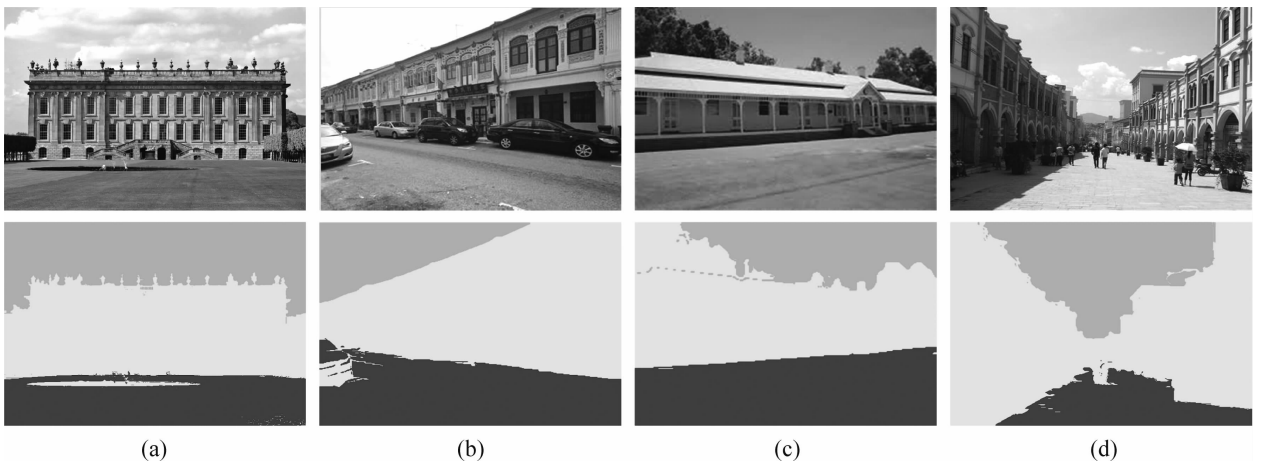


图4 基于几何复杂度的图像分割

Fig. 4 Image segmentation based on geometric complex

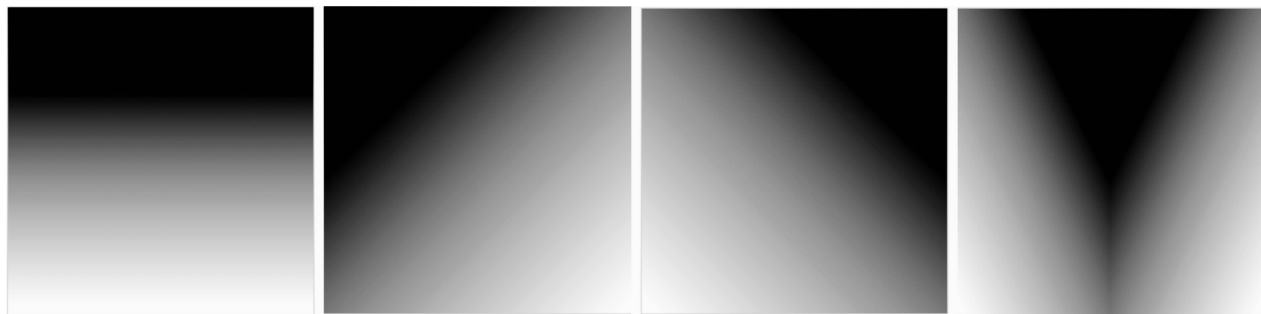


图5 标准深度图

Fig.5 Standard depth maps

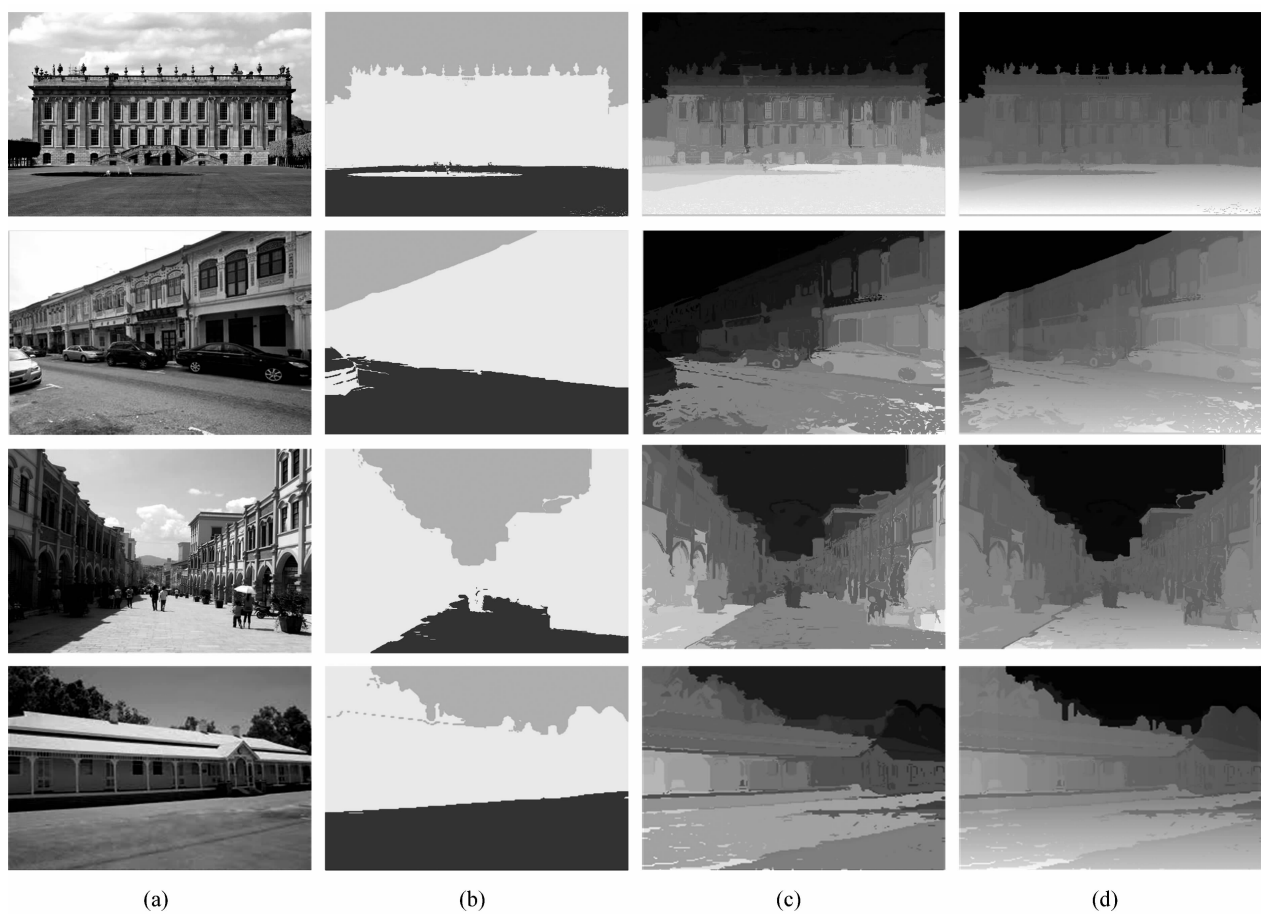


图6 基于几何复杂度的图像分割和深度生成

Fig.6 Image segmentation and depth generation based on geometric complex

### 3.3 实验结果

为说明提出算法的有效性,我们和算法[13]做了比较。两种算法都是全自动深度估计算法,图6中显示了比较结果。其中,(a)列是输入图像,(b)列是图像分割结果,(c)列是依据算法[13]获得的

深度图,(d)列是本算法获得的深度图。在(b)列的图像分割结果中,青绿色区域代表天空部分,黄色区域代表了图像中的主要建筑物部分,而深蓝色代表了地面区域。从(c)列和(d)列可以看到,算法[13]考虑了12个场景分类,虽保留了建筑物等纹

理对象的深度细节,但当该区域占有足够大的图像面积且位于深度一致变化区域,该假设会产生较大的误差,且可能导致相邻区域间的深度值变化不连续性,本算法保留深度图足够细节,同时加强了图像中相邻区域的深度变化的连续性,获得的深度图质量更高。

## 4 结论

本文提出了一种应用于室外场景的基于几何复杂度的图像分割和深度生成算法。文中关注的大部分室外场景主要包含天空、地面以及建筑物三个部分,具有各自不同的深度特点。首先,通过图像中主要线段的角度统计分布将室外人工场景的几何结构规划为四种类型;然后,利用 meanshift 分割算法将输入图像分割成许多小的区域,并依据场景几何结构类型将这些小的区域融合成为天空、人工建筑和地面三个大的区域,最后依据此分割结果得到图像的几何结构深度图。实验结果表明与已有算法相比,提出的方法简化了场景的分类过程;保留足够细节,同时加强了图像中相邻区域的深度变化的连续性,提高了深度图的质量。对于复杂的室外场景或者室内场景,当可以提取主要线段的角度统计分布实现几何结构规划时,亦可采用本算法实现相关场景的基于几何复杂度的图像分割并获得有效的深度图,否则无法适用于本算法。

## 参考文献

- [1] Fehn C. Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3d-tv[J]. Proc. SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI. San Jose, CA, USA: SPIE, 2004: 93-104.
- [2] Guo G, Zhang N, Huo L, et al. 2d to 3d conversion based on edge defocus and segmentation[C]//IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA, 2008: 2181-2184.
- [3] He K, Sun J, Tang X. Single Image Haze Removal Using Dark Channel Prior[C]//IEEE International Conf. on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 2009: 1956-1963.
- [4] Palou G, Salembier P. Occlusion-based depth ordering on monocular images with binary partition tree[C]//IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 2011: 1993-1996.
- [5] Palou G, Salembier P. From local occlusion cues to global monocular depth estimation[C]//IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2012: 793-796.
- [6] Saxena A, Sun M, Ng A Y. Make3D: learning 3D scene structure from a single still image[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009, 31(5): 824-840.
- [7] Mohaghegh H, Karimi N, Reza Soroushmehr S M, et al. Single image depth estimation using joint local-global features[C]//IEEE International Conf. on Pattern Recognition (ICPR), Cancun, Mexico, 2016: 27-732.
- [8] Mancini M, Costante G, Valigi P, et al. Toward domain independence for learning-based monocular depth estimation[J]. IEEE Robotics and Automation Letters, 2017, 2(3): 1778-1785.
- [9] Martínez-Usó A, Latorre-Carmona P, Sotoca J M, et al. Depth estimation in integral imaging based on a maximum voting strategy[J]. Journal of Display Technology, 2016, 12(12): 1715-1723.
- [10] Hoiem D, Efros A A, Hebert M. Geometric context from a single image[C]//IEEE International Conf. on Computer Vision (ICCV), Beijing, China, 2005: 654-661.
- [11] Nedovic V, Smeulders A. W. M., Redert A, et al. Depth information by stage classification[C]//IEEE International Conf. on Computer Vision (ICCV), Rio de Janeiro, Brazil, 2007: 1-8.
- [12] Jung C, Kim C. Real-time estimation of 3D scene geometry from a single image[J]. Pattern Recognition, 2012, 45(9): 3256-3269.
- [13] Lee H, Jung C, Kim C. Depth map estimation based on geometric scene categorization[C]//IEEE The 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision, Incheon, South Korea, 2013: 170-173.
- [14] Rafael C, Woods R E. Digital Image Processing[M]. 3rd ed. Prentice Hall, Upper Saddle River, 2008.
- [15] Comaniciu D, Meer P. Mean shift: a robust approach to

ward feature space analysis[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619.

[16] Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(6): 583-598.

[17] Sumengen B. Variational image segmentation and curve evolution on natural images[D]. Barbara Santa; University of California, 2004.

[18] Ning J, Zhang L, Zhang D, et al. Interactive image segmentation by maximal similarity based region merging[J]. Pattern Recognition, 2010, 43: 445-456.

[19] 邹芙蓉, 安平, 李贺建, 等. 实时3DV系统中面向虚拟视绘制的快速深度编码[J]. 信号处理, 2012, 28(4): 565-571.

Wu Fuqiong, An Ping, Li Hejian, et al. Fast Depth Coding for Virtual View Synthesis in Real Time 3DV System[J]. Signal Processing, 2012, 28(4): 565-571. (in Chinese)

[20] 陈坤斌, 刘海旭, 李学明. 构造全局背景的虚拟视点合成算法[J]. 信号处理, 2013, 29(10): 1307-1314.

Chen Kunbin, Liu Haixu, Li Xueming. Virtual View Synthesis Using Generated Global Background[J]. Journal of Signal Processing, 2013, 29(10): 1307-1314. (in Chinese)

[21] Kailath T. The divergence and Bhattacharyya distance measures in signal selection[J]. IEEE Transactions on

Communications Technology, 1967, 15(1): 52-60.

### 作者简介



**任艳楠** 女, 1977年生, 河南新乡人。山东大学信息科学与工程学院博士研究生, 研究方向为图像处理和计算机视觉技术等。

E-mail: yannan\_ren@163.com



**刘 琚** 男, 1965年生, 山东临沂人。山东大学信息科学与工程学院教授、博士生导师, 研究方向为信号处理和技术等。

E-mail: juliu@sdu.edu.cn



**元 辉** 男, 1984年生, 山东威海人。山东大学信息科学与工程学院教授, 研究方向为视频压缩和编码。

E-mail: Huiyuan@sdu.edu.cn



**顾凌晨** 女, 1994年生, 山东临沂人。山东大学信息科学与工程学院硕士研究生。研究方向为多媒体检索、图像处理等。

E-mail: lingchengu@126.com