

DOI: 10.7652/xjtuxb201712010

采用多层图模型推理的道路场景分割算法

邓燕子, 卢朝阳, 李静, 刘阳

(西安电子科技大学综合业务网理论及关键技术国家重点实验室, 710071, 西安)

摘要: 针对传统图模型分割算法提取的物体边缘不够精细、难以适应复杂道路场景布局的问题,提出了一种基于多层图模型推理的道路场景分割(HGI)算法。该算法先将图像过分割为同质的超像素块,再采用随机森林模型训练超像素块的多类别回归器和相邻超像素的一致性回归器;然后用 2 种回归值计算马尔科夫随机场(MRF)模型的能量项,通过推理得到初始分割;最后为了解决超像素块包含多类别带来的分类混淆,在初始分割基础上构建像素级的全连接条件随机场模型,进行优化得到精细的分割结果。实验结果表明,采用 HGI 算法对人工标注数据库和真实拍摄的场景图像处理能够得到精细的分割边缘,能够解决超像素推理中的类别混淆问题,与传统的 MRF 图模型分割方法相比,在总体精度和平均召回率 2 个指标上分别提高了 2% 和 3%。

关键词: 道路场景分割;多类别图像标记;随机森林;马尔科夫随机场

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 0253-987X(2017)12-0062-06

A Segmentation Algorithm for Road Scenes Using Hierarchical Graph-Based Inference

DENG Yanzi, LU Zhaoyang, LI Jing, LIU Yang

(State Key Laboratory of Integrated Service Network, Xidian University, Xi'an 710071, China)

Abstract: A novel segmentation algorithm for road scenes based on hierarchical graph-based inference (HGI) is proposed to solve the problem that object boundaries extracted by existing graph-based segmentation algorithms are not fine enough and it is hard to adapt to complex road scene layouts. The algorithm first over-segments an image into small homogeneous regions called superpixels, and then the random forest model is used to train a multi-class regressor and a consistency regressor of superpixels. Regression results are then used to calculate the energy terms in a Markov random field (MRF) energy function. An initial segmentation of the image is obtained by using the superpixel MRF inference. A pixel-level labeling based on fully connected conditional random fields is constructed to avoid the label confusion caused by the superpixels and to get a fine segmentation finally. Experimental results show that the proposed algorithm solves the label confusion in the superpixel inference and gets fine segmentation boundaries for both the images in manually labeled datasets and the real road scenes. A comparison with the traditional MRF graph-based inference methods shows that the HGI algorithm provides improvements of 2% and 3% on the overall precision and per-class average metrics, respectively.

Keywords: road scene segmentation; multi-class image labeling; random forest; markov random field

收稿日期: 2017-04-06。 作者简介: 邓燕子(1983—),女,博士生;卢朝阳(通信作者),男,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(61502364)。

网络出版时间: 2017-09-15 网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20170915.2051.002.html>

道路场景分割是计算机视觉领域中的热点问题之一,其目的是分割图像中的主要物体并识别它们的类别,包括道路、道路平面之上的垂直物和天空^[1]。精确的道路分割能够为自动驾驶系统提供导航信息,具有重要的实用价值,但由于道路场景的视觉外观的复杂性,现有的方法很难得到精细的分割结果,因此难以得到实际应用。

传统的道路场景分割采用对图像逐个像素进行分类标注的方法^[2-4],难以在整个图像上获取全局结构信息,并且对高分辨率图像运算比较耗时。Hoiem 等人提出了基于超像素块分类的方法^[1],该方法将图像划分为同质的超像素块,对超像素块进行分类标注,能够提高分割的效率。文献[5-6]采用了更有效的超像素特征进一步提高了分割的准确性。传统的分类方法依赖人工设计的特征,很难描述多变的场景内容,而近年来一些基于深度学习的算法^[7-8]则能更好地进行特征表示。然而,以上基于物体分类的方法均依赖少量的数据库训练样本,当测试数据和训练数据差别较大时会严重影响分类的精确度。由于道路场景存在光线不均匀、物体之间相互遮挡和阴影等问题,仅通过分类很难得到准确的分割结果,因此现有的方法一般采用马尔科夫随机场(MRF)或者条件随机场(CRF)这类图模型推理对分类结果进行后处理^[7]。超像素的 MRF 推理能够使用高效的图割算法^[9]来进行求解,但它只能依赖当前邻域内的节点信息,因此分割结果不够精细。相比之下,CRF 模型在图像像素标记任务上的能力更强。传统的 CRF 模型采用相邻的像素或者图像块构建成对势能函数^[3],它不能描述相邻较远的连接关系,通常会导致目标边缘的过度平滑。采用全连接 CRF 模型可以避免这一缺点^[10],但计算更加耗时。文献[11]提出了一种高效的全连接 CRF 推理算法,可以大幅提升这种模型的运算效率。

为了解决单一图模型推理进行类别标记的不足,本文根据道路场景的空间布局特点提出了多层图模型推理的分割算法。先将图像进行过分割将图像划分为同质的超像素块,通过超像素的 MRF 推理快速得到初始分割,为了避免超像素块内存在多类物体造成分类混淆,在初始分割基础上构建像素级的全连接 CRF 模型,通过模型优化得到精细的分割结果。实验部分采用 2 个基准数据库和真实拍摄的数据对算法进行定性和定量的评价,并和现有的方法进行对比,验证了算法的有效性和稳定性,能得到精细的分割边缘。

1 图像超像素分类

1.1 超像素特征

场景分割就是为图像的每个像素点标记几何属性的类别。对于尺寸较大的图像,逐点进行分类效率太低。由于图像中大部分区域都是同质的,将图像划分为同质的区域再进行分类可以提高图像分割的效率,因此本文通过对超像素分类进行场景分割。使用文献[12]的方法获取超像素块,可以将图像划分为均匀的小块区域,且保留物体的边缘和轮廓等重要信息。为了获得区分性较好的视觉特征,需要使用多种特征集合来描述超像素块^[1],为此本文提取了以下 4 类特征。

(1)颜色特征。道路场景图像的颜色信息非常丰富,一般天空会呈现天蓝色、灰白色或者白色,道路则是褐色或者灰褐色,只有垂直物的颜色变化比较大。本文在 RGB(红、绿、蓝)和 HSV(色调、饱和度、亮度)2 种颜色空间提取统计信息,包括超像素 3 个通道的均值、色度直方图和饱和度直方图。

(2)纹理特征。道路场景的纹理特征采用文献[13]中的滤波器组来表示,包括滤波器的平均绝对输出和滤波器最大响应方向的统计直方图。

(3)结构信息。包括尺度不变特征变换(scale-invariant feature transform, SIFT)和方向梯度直方图(histogram of oriented gradient, HOG)。本文采用稠密方式提取这 2 种特征,首先在超像素内计算每个像素的特征向量,然后计算超像素块内所有向量的几何平均作为其特征向量。

(4)位置和形状特征。道路场景中物体的坐标和形状能够为分割提供很强的几何布局线索。本文采用文献[1]中的方法得到位置和形状特征,包括归一化的像素位置,归一化的超像素面积值。

1.2 超像素分类

将图像划分为 3 种几何类别,传统方法一般是为每个类别训练分类器,最后综合得到超像素属于某一类的概率^[1],这种方法的精度和效率都不高。为了更高效地得到超像素类别的置信度图,本文使用随机森林回归器^[14]训练多类别回归模型。将超像素 s_i 的特征向量记为 f_i ,类标集合表示为 $l \in \{1, 2, 3\}$,其中 1、2、3 分别对应道路、垂直物、天空 3 种类别。多类别回归器 H_1 使用训练集 $\{f_i, l_i\}$ 学习得到。在测试过程中,对于一个输入特征,多类回归器 H_1 的输出是 1~3 之间的连续值。

另外,相邻超像素块之间的差别能够辅助对该

超像素的属性进行判别,因此本文训练了回归器用来估计相邻超像素类别的一致性。一致性回归器采用2个超像素的差值进行训练,本文采用文献[1]中的方法计算超像素特征的差值。超像素特征包含了向量和直方图这两类,需要用不同的方法计算差值。对于颜色、纹理、像素位置、SIFT和HOG等向量特征,使用2个向量之间差的绝对值进行计算;对于色调、饱和度和纹理直方图特征的差值,使用Jensen-Shannon散度进行计算。超像素对的一致性标签表示为 $k \in \{-1, 1\}$,这里-1和1分别对应相邻2个超像素不同和相同的分类结果。利用对比特征训练一致性回归器 H_2 ,用回归值作为2个超像素块属于相同类别的隶属度。回归器的输出也是-1~1之间的连续值。

1.3 超像素类标 MRF 推理

对超像素分类之后,本文采用MRF模型推理的方法得到超像素的类标。采用最小化MRF能量函数来解决图像分割问题^[15]是计算机视觉领域比较流行的方法。对一幅输入图像,假设 $S = \{s_i\}$ 是图像的超像素集合,成对的MRF能量函数定义为

$$E(L) = \sum_{s_i \in S} U_i(l_i) + \lambda \sum_{(s_i, s_j) \in P} V_{ij}(l_i, l_j) \quad (1)$$

式中: i 和 j 是超像素索引号; l_i 和 l_j 分别是超像素 i 和 j 的候选标签; P 是所有相邻超像素对的集合; λ 是成对能量项的权值; U_i 和 V_{ij} 分别为数据项和平滑项势能函数,由超像素的回归结果计算得到,其中 U_i 表示分配类标 l_i 到像素块 s_i 的代价值,本文将其定义为 s_i 的类别回归值和类标值之差的绝对值

$$U_i(l_i) = |H_1(f_i) - l_i| \quad (2)$$

超像素对 $\{s_i, s_j\}$ 之间的成对势能 V_{ij} 是惩罚相邻超像素 s_i 和 s_j 取不同类标的代价值。本文采用Potts模型^[16]来表示成对的代价函数,定义如下

$$V_{i,j}(l_i, l_j) = w(l_i, l_j) T(l_i, l_j) \quad (3)$$

$$T(l_i, l_j) = \begin{cases} 1, & l_i \neq l_j \\ 0, & l_i = l_j \end{cases} \quad (4)$$

式中, w 是相邻超像素之间的权值,由类别一致性回归器的输出来计算

$$w(l_i, l_j) = 1 + H_2(f_i, f_j) \quad (5)$$

对一幅输入图像,计算所有超像素和超像素对的代价函数后,本文使用图割算法^[9]来求解式(1)中的MRF能量函数,可以得到超像素的类别。

2 全连接 CRF 像素级分割

2.1 全连接 CRF 模型

对超像素标记后,假设一个超像素内所有像素

都属于超像素本身的类别,此时获得一个初始的分割结果,但很多时候这种假设并不完全成立。超像素是一个图像粗分割过程,当道路场景存在阴影、物体的光照不均匀等物体边界很难区分的情况时,它可能包含2种以上类别的物体,因此需要对初始分割结果进行逐个像素的精细分类处理。本文采用全连接CRF模型来解决像素标记的问题,这种模型由于用到了更多的邻域信息,因此可以获得精细的物体边缘。为了快速的进行全连接CRF模型的推理,本文使用文献[11]中提出的一种高效的求解方法。

给定一幅包含 N 个像素的图像 I ,随机场 X 包含 N 个随机变量 $\{x_1, x_2, \dots, x_n\}$,这些变量代表对应像素的类别,因此取值范围同样也是类标值。标准全连接CRF能量函数定义如下

$$E(X | I) = \sum_i E_u(x_i) + \sum_{i < j} E_p(x_i, x_j) \quad (6)$$

式中: i 和 j 代表变量的序号,取值为1到 N ; E_u 是一元势能项,表示分配标签 l_i 给变量 x_i 的代价值; E_p 是2个相连接变量的成对势能项。

2.2 模型计算和推理

模型的一元势能项由初始标签计算,假设初始类别有50%的概率是正确的,则一元势能的计算公式为

$$E_u(x_i) = \begin{cases} \log(0.5), & x_i = l_i \\ \log(0.5/3), & x_i \neq l_i \end{cases} \quad (7)$$

2个相连接变量间的成对势能定义为一组高斯核函数的线性组合^[11],形式如下

$$E_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^M w^{(m)} k^{(m)}(h_i, h_j) \quad (8)$$

$$k^{(m)}(h_i, h_j) = \exp\left(-\frac{1}{2}(h_i, h_j)^T \mathbf{A}^{(m)}(h_i, h_j)\right) \quad (9)$$

式中: $w^{(m)}$ 是线性组合的权值; $k^{(m)}$ 是一个高斯核函数; h_i 和 h_j 分别对应变量 x_i 和 x_j 的特征,本文采用3个颜色通道的亮度值和位置坐标构成5维的特征向量; $\mathbf{A}^{(m)}$ 是一个用于定义高斯核函数形状的正定矩阵; μ 是类标一致性函数,用于惩罚2个相似的像素赋予不同类标的代价,采用文献[9]中的方法学习得到类标一致性函数值。

在全连接CRF模型中,由于每个变量都与其他变量相连接,对于尺寸大的图像,全连接CRF推理过程会非常耗时。本文采用文献[11]中提出的平均场近似的方法来求解,利用高效的滤波算法^[17]进行加速,从而大幅提高推理过程的运算效率。经过全连接CRF的学习和推理之后,可以得到每个像素点

的类标,并且能够保持不同类别物体间精细的边缘。

2.3 本文算法流程

本文提出的基于多层图模型推理的道路场景分割算法流程如图 1 所示,详细步骤如下。

- 步骤 1 将输入图像划分成同质的超像素块。
- 步骤 2 计算所有超像素块类别回归值,计算所有相邻超像素块属于同类的回归值。
- 步骤 3 采用式(2)~(5)计算 MRF 模型的能量函数,使用图割算法得到超像素类标。
- 步骤 4 采用式(7)~(9)计算全连接 CRF 模型的能量函数,进行函数优化得到像素点的类标,得到最终的分割结果。

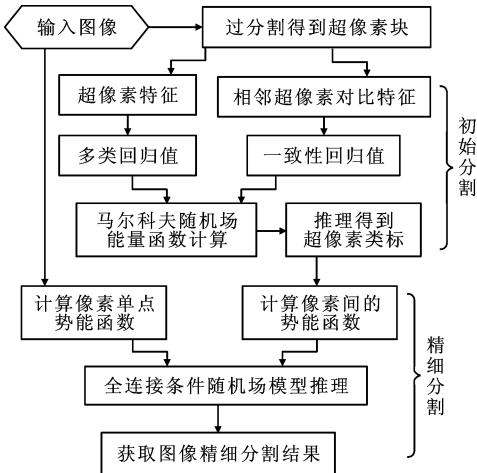


图 1 采用多层图模型推理的道路场景分割算法流程

3 实验结果和分析

为了验证本文方法对道路场景分割的有效性,通过分析算法中不同层次的分割结果来检查对精度的提升,并同文献[1,3,5,18]中的方法进行了对比。其中,文献[3]采用基于邻域 CRF 模型的方法进行图像多类别的标记,文献[18]融合离线学习和在线学习的结果进行分割,文献[5]采用了基于多视觉特征的图像分割。对实验结果的分析采用定性和定量的评价方法。

3.1 实验设计和数据库

实验采用 GC^[1]和 KITTI-Layout^[18]2 个场景分割数据库。GC 数据库包含 300 张户外场景,图像大小包括 300×240 像素和 1 024×768 像素这 2 种尺寸,并且提供了对超像素进行的 3 种几何类别的标记真值。采用同文献[1]相同的训练策略,所有比较方法都进行 5 次交叉验证,每次随机选取数据库一半数据进行训练,剩余进行测试。在本文的方法中,MRF 能量函数中参数 λ 设置为 10。

KITTI-Layout 数据库包含了 323 张街道场景图像,图像大小均为 1 242×375 像素,该数据库提供了每个像素的类别标记。为了方便进行超像素类标回归的训练,需要生成超像素的类标。理想情况下,超像素和包含的像素应被标记为相同的类别。但有时一个超像素可能包含多个类别物体,在这种情况下,取包含最多像素的类别作为超像素的类标。实验同样采用 5 次交叉验证实验,每次随机选取 150 张图像用来训练,其余用来测试,参数的选择与 GC 数据库所使用的相同。

3.2 实验结果评价

对分割的定量评价,采用总体精度(用符号 A_o 表示)和平均召回率(用符号 A_r 表示)2 个常用的指标^[19],计算方法如下:假设 $M \in R^{3 \times 3}$ 是 3 种几何类别的混淆矩阵,列表示类别的真值,行表示预测的类别,则 M_{ij} 表示类别 i 被标记成类别 j 的数量。总体精度和平均召回率的指标分别定义为

$$A_o = \sum_{i=1}^3 M_{ii} / \sum_{i=1}^3 \sum_{j=1}^3 M_{ij} \tag{10}$$

$$A_r = \frac{1}{3} \sum_{i=1}^3 (M_{ii} / \sum_{j=1}^3 M_{ij}) \tag{11}$$

表 1 是在 GC 数据库上 2 个指标的对比结果。其中超像素分割是本文算法中经过超像素 MRF 推理的初始分割结果。从表 1 结果中可以看出,初始分割的结果精度最低,但总体精度和文献[1]方法相差不大,因为这 2 组结果都是对超像素进行标记。经过精细分割后的最终结果要比初始分割结果有显著的提升。总体精度和平均召回率指标分别提高了近 2%和 4%,结果优于所比较的传统方法。

表 1 GC 数据库客观评价实验结果对比

算法	总体精度	平均召回率
文献[1]	0.881 0	0.879 7
文献[3]	0.898 2	0.900 5
文献[5]	0.893 3	0.893 3
超像素分割	0.880 4	0.854 6
本文	0.904 6	0.903 5

注:黑体数据为最优分割结果。

表 2 是在 KITTI-Layout 数据库上 2 个指标的比较结果。由于在训练回归器时本文对超像素进行了标记,因此首先统计了标记结果的可靠性。超像素类标的统计结果都接近 98%,说明为超像素标记的结果能够满足训练回归器的要求。通过分析表 2 的结果,可以得到和 GC 数据库类似的结果。本文

方法的总体精度指标最优,在平均召回率指标上略低于最好的多类别分割方法。通过客观指标的对比验证了本文方法的有效性,并且在整体上优于传统的方法。

表 2 KITTI-Layout 数据库客观评价实验结果对比

算法	总体精度	平均召回率
超像素类标	0.978 2	0.978 6
文献[3]	0.945 6	0.929 6
文献[18]	0.915 6	0.915 5
超像素分割	0.935 4	0.908 9
本文	0.946 8	0.928 4

图 2 和图 3 分别给出 2 个数据集中图像的分割结果。在分割结果中,黑色表示道路区域,灰色表示垂直物区域,白色表示天空。从图中可以看出本文的结果更接近真值。文献[3]的结果存在孤立的小块,且边缘过于平滑,而本文方法能获得在不同类别物体边界上更精确的边缘结构。为了进一步检验算法的适应性,本文对实际拍摄的道路场景进行了实验,结果如图 4 所示。在进行超像素分类时,没有对实际拍摄数据进行重新训练,而是直接采用在 KITTI-

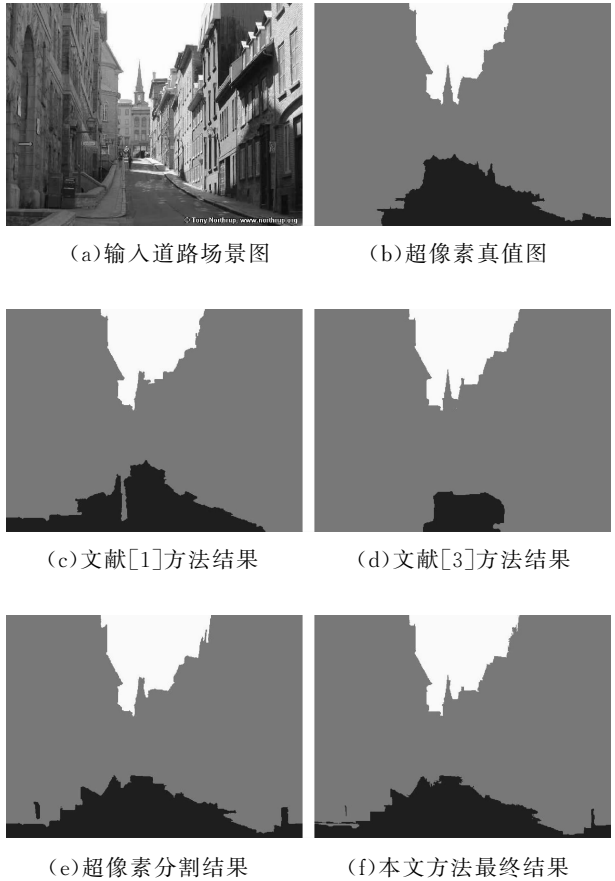


图 2 GC 数据库道路场景分割结果

Layout 数据库上训练的结果。从图 4 可以看到,本文方法能够给出较好的分割结果,并分割出较完整的道路区域,证明了算法具有较好的适应性。

由于本文在计算全连接 CRF 的成对能量函数时使用颜色作为特征向量,因此算法容易受到特别暗的阴影覆盖的影响。另外在垂直物和道路的颜色接近时也较难分割。在未来的工作中将会对这些问题进行进一步的改进。

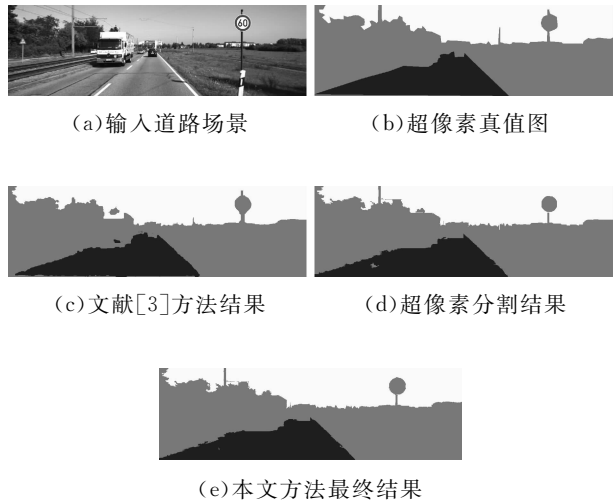


图 3 KITTI-Layout 数据库道路场景分割结果

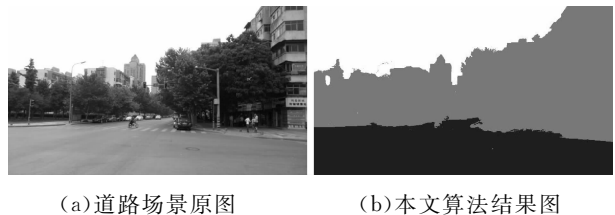


图 4 实际拍摄的道路场景分割结果

4 结 语

针对传统图模型推理的场景分割方法容易受初始分割影响导致精度降低的问题,提出了一种基于多层图模型推理的道路场景分割方法。首先通过训练多几何类别回归器和一致性回归器进行超像素分类;然后使用回归结果来构建 MRF 能量函数,通过超像素 MRF 推理获得图像的初始分割;为了解决一个超像素块包含多类别带来的分类混淆,采用全连接 CRF 的像素级标记获得场景的精细分割。实验部分采用 2 个数据库和真实拍摄图像对算法进行验证,通过定性和定量分析证明了本文方法用于场景分割的准确性和稳定性,并且综合性能优于传统算法。

参考文献:

- [1] HOIEM D, EFROS A, HEBERT M. Recovering surface layout from an image [J]. *International Journal of Computer Vision*, 2007, 75(1): 151-172.
- [2] COSTEA A D, NEDEVSCI S. Fast traffic scene segmentation using multi-range features from multi-resolution filtered and spatial context channels [C]// *IEEE Intelligent Vehicles Symposium Proceedings*. Piscataway, NJ, USA: IEEE, 2016: 328-334.
- [3] GOULD S. Darwin: a framework for machine learning and computer vision research and development [J]. *Journal of Machine Learning Research*, 2012, 13(18): 3533-3537.
- [4] COSTEA A D, NEDEVSCI S. Semantic channels for fast pedestrian detection [C]// *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE, 2016: 2360-2368.
- [5] 邓燕子, 卢朝阳, 李静. 交通场景的多视觉特征图像分割方法 [J]. *西安电子科技大学学报*, 2015, 42(6): 11-16.
DENG Yanzi, LU Zhaoyang, LI Jing. Segmentation of the image with multi-visual features for a traffic scene [J]. *Journal of Xidian University*, 2015, 42(6): 11-16.
- [6] CHACRA D A, ZELEK J. Road segmentation in street view images using texture information [C]// *Proceedings of 2016 IEEE Conference on Computer and Robot Vision*. Piscataway, NJ, USA: IEEE, 2016: 424-431.
- [7] FARABET C, COUPRIE C, NAJMAN L, et al. Learning hierarchical features for scene labeling [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1915-1929.
- [8] 王海, 蔡英凤, 贾允毅, 等. 基于深度卷积神经网络的场景自适应道路分割算法 [J]. *电子与信息学报*, 2017, 39(2): 263-269.
WANG Hai, CAI Yingfeng, JIA Yunyi, et al. Scene adaptive road segmentation algorithm based on deep convolutional neural network [J]. *Journal of Electronics and Information Technology*, 2017, 39(2): 263-269.
- [9] 余森, 胡占义. 高阶马尔科夫随机场及其在场景理解中的应用 [J]. *自动化学报*, 2015, 41(7): 1213-1234.
YU Miao, HU Zhanyi. Higher-order Markov random fields and their applications in scene understanding [J]. *Acta Automatica Sinica*, 2015, 41(7): 1213-1234.
- [10] HORNE L, ALVAREZ J M, SALZMANN M, et al. Efficient scene parsing by sampling unary potentials in a fully-connected CRF [C]// *IEEE Intelligent Vehicles Symposium Proceedings*. Piscataway, NJ, USA: IEEE, 2015: 820-825.
- [11] KRAHENBUHL P, KOLTUN V. Efficient inference in fully connected CRFs with Gaussian edge potentials [C]// *Proceedings of 25th Annual Conference on Neural Information Processing Systems 2011*. Red Hook, NY, USA: Curran Associates Inc., 2011: 109-117.
- [12] FELZENSZWALB P F, HUTTENLOCHER D P. Efficient graph-based image segmentation [J]. *International Journal of Computer Vision*, 2004, 59(2): 167-181.
- [13] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(5): 898-916.
- [14] XIAO L, DAI B, LIU D, et al. Monocular road detection using structured random forest [J]. *International Journal of Advanced Robotic Systems*, 2016, 13(3): 101-114.
- [15] 徐胜军, 韩九强, 赵亮, 等. 用于图像分割的局部区域能量最小化算法 [J]. *西安交通大学学报*, 2011, 45(8): 7-12.
XU Shengjun, HAN Jiuqiang, ZHAO Liang, et al. Algorithm of minimizing local region energy for image segmentation [J]. *Journal of Xi'an Jiaotong University*, 2011, 45(8): 7-12.
- [16] GRIDCHYN I, KOLMOGOROV V. Potts model, parametric maxflow and K-submodular functions [C]// *Proceedings of the IEEE International Conference on Computer Vision*. Piscataway, NJ, USA: IEEE, 2013: 2320-2327.
- [17] ADAMS A, BAEK J, DAVIS M A. Fast high-dimensional filtering using the permutohedral lattice [J]. *Computer Graphics Forum*, 2010, 29(2): 753-762.
- [18] ALVAREZ J M, GEVERS T, LECUN Y, et al. Road scene segmentation from a single image [C]// *Proceedings of the 12th European Conference on Computer Vision*. Berlin, Germany: Springer Verlag, 2012: 376-389.
- [19] LADICKY L, RUSSELL C, KOHLI P, et al. Associative hierarchical random fields [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(6): 1056-1077.

(编辑 刘杨)