

## 面向微光/红外融合彩色夜视的场景解析方法

鲁佳颖, 谷小婧, 顾幸生

(华东理工大学 化工过程先进控制与优化技术教育部重点实验室, 上海 200237)

**摘要:** 彩色夜视技术可以将微光/红外双谱图像融合成一幅适于人眼观察的彩色图像, 而恰当的场景解析方法能够对彩色夜视图像的内容做出自动化分析, 进一步减轻人眼的观测负担。针对彩色夜视场景丰富多变、对算法灵活性要求高的特点, 提出了一种可在线扩展的场景解析方法。该方法基于非参数模型, 预测景物类别时不需要训练过程, 只需要使用数据库中具有语义标记的样本图像, 通过将待解析图像与样本图像进行全局及局部匹配来实现语义标签的传递。而且, 数据库可以根据应用场景的不同随时进行动态扩充。实验结果表明: 该方法在包含城市、乡野等多种场景的夜视图像上, 以及由统计色彩映射、TNO、NRL 等多种融合方法得到的、具有不同色彩表征的彩色夜视图像上都具有令人满意的准确率。

**关键词:** 彩色夜视; 场景解析; 非参数模型; 超像素特征; 马尔科夫随机场

**中图分类号:** TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA201746.0804002

## Scene parsing method toward low-light-level/infrared color night vision

Lu Jiaying, Gu Xiaojing, Gu Xingsheng

(Key Laboratory of Advanced Control and Optimization for Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China)

**Abstract:** Color night vision technology is able to fuse the dual-spectral image of low-light-level and infrared image into a color one suited to human observation. Furthermore, a appropriate scene parsing method on the color night vision image could additionally facilitate human observation by providing automatic content analysis. An online scalable scene parsing method was proposed aiming at the rich and changeable color night vision in practice which required algorithms with high flexibility. The proposed method was based on a non-parametric model that needed no training process when predicting scene categories. It matched the query image and the sample images in database using both global and local features, and then transfered semantic labels of the best-match samples to the query image. Moreover, the database can be dynamically expanded according to different usage scenarios. The experimental results show that proposed method achieves satisfactory accuracies on color night vision images that obtained by a variety of color night vision methods, including the statistical color mapping,

收稿日期: 2016-12-05; 修订日期: 2017-01-10

基金项目: 国家自然科学基金(61205017, 61502293, 61573144); 中央高校基本科研业务费专项

作者简介: 鲁佳颖(1992-), 女, 硕士生, 主要从事彩色夜视图像处理及分析方面的研究。Email: jiayinglu92@gmail.com

导师简介: 顾幸生(1960-), 男, 教授, 博士生导师, 主要从事计划与生产调度、复杂工业过程建模、控制与优化方面的研究。

Email: xsqu@ecust.edu.cn

TNO, and NRL, throughout diverse scenes, including cities, countryside and others.

**Key words:** color night vision; scene parsing; nonparametric model; superpixels feature; Markov random field

## 0 引言

夜视技术分为微光夜视和红外夜视两类。微光夜视生成光谱细节丰富、视觉纹理清晰,但易受环境干扰,容易丢失目标。红外夜视具备良好的热物体探测能力,但图像对比度低,细节表现能力差。随着微光与红外成像技术的发展,彩色融合两者的图像信息,有助于场景理解。在军用方面,有利于人在隐藏、伪装的背景下更快更精确地探测目标,进行态势感知和情报收集<sup>[1-3]</sup>。在民用方面,能弥补单一传感器在农业监控、遥感应用、医学检测<sup>[4-6]</sup>等领域所表现的不足,具有极其广泛的应用前景。

图像场景解析是计算机视觉领域的一个研究热点。目前已被广泛应用于机器人导航、自动控制、环境监测、医疗诊断等领域。从处理单元角度,可分为:基于像素点的解析<sup>[7-8]</sup>,基于分割区域的解析<sup>[9-10]</sup>和基于物体边界的解析<sup>[11-12]</sup>。这些解析方法采用马尔科夫随机场或条件随机场构建解析模型,需训练大量参数。近几年,随着神经网络的发展<sup>[13-14]</sup>,大量工作将卷积神经网络用于图像的场景解析,有效避免人工设计特征的局限性,特征描述能力强,可显著改善图像场景解析的性能。

在文中,笔者提出了一种基于非参数模型的彩色夜视图像场景解析方法,不同于以往参数化方法,文中方法预测图像景物类别无需训练,只需简单的数据统计,对测试图像和样本图像进行全局及局部特征匹配,将样本的语义标签传递至测试图像。特别适合于彩色夜视图像场景内容丰富、数据库需根据不同场合动态扩展的问题。对典型彩色夜视图像场景解析,结果表明:文中方法在各种场景的彩色夜视图像上都能取得鲁棒的解析结果。

## 1 彩色夜视图像场景解析

由于彩色夜视技术中常用的融合方法包括统计色彩映射、TNO、NRL 等多种。机制不同,各种方法生成的彩色图像的色彩表征差异极大。为不失一般

性,文中数据库中的图像选用带有语义标记的自然彩色图像,通过将待解析的彩色夜视图像与数据库中的自然彩色图像进行全局及局部特征匹配,实现语义标签的传递。算法可分为以下 3 部分,流程如图 1 所示。

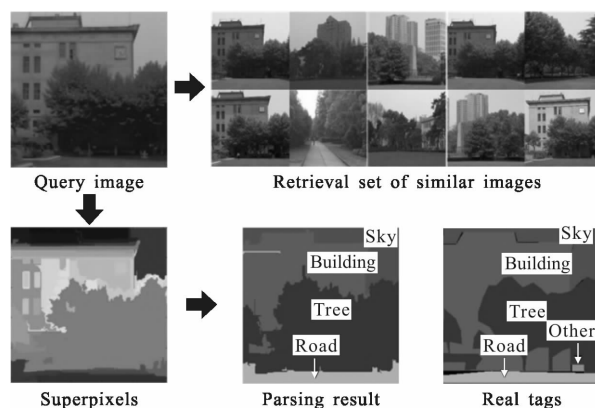


图 1 算法流程图

Fig.1 Flow chart of algorithm

(1) 候选集检索:在数据库中为彩色夜视测试图像寻找一批场景相似的自然彩色图作为候选集。

(2) 超像素语义标签传递:将测试图像和数据库图像超像素分割,采用最近邻算法将候选集超像素上的标签信息传递给测试图超像素。

(3) 平滑修正:利用马尔科夫随机场,结合上下文信息,修正步骤(2)的分类结果。

### 1.1 检索候选集

基于非参模型的彩色夜视图像 (Color Night Image, CNI) 场景解析,第一步就是从自然彩色图像集 NS(Natural Color Image Set) 中检索出与其场景类型、物体种类、空间布局等方面高度相似的自然彩色图像 (Natural Color Image, NCI) 作为标签传递的候选集 (Retrieval Set, RS)。文中采用 Spatial pyramid、Color histogram、Gist 3 个全局特征作为挑选候选集的依据。

首先,计算测试图像与每张数据库图像在 3 种特征上的欧氏距离,并对其进行降序排列。然后,比较数据库图像在 3 种特征上的 3 个排序,取最小排序作为该张数据库图像与测试图像的最短距离。最

后取距离靠前的  $K$  张图像作为候选集。 $K$  值的选取通过实验确定。

### 1.2 超像素标签传递

对彩色夜视图像和自然彩色图像超像素分割,得到超像素  $scn_i, sn_cj$ ,对其提取涵盖纹理、颜色、空间和形状的 20 种图像特征  $f^k(k=1 \cdots 20)$ 。

利用文中得到的候选集  $RS$  和 20 种图像特征  $f^k(k=1 \cdots 20)$ ,采用最近邻算法,对彩色夜视超像素,在候选集上找到与之距离最近的超像素,统计这些超像素多数属于哪个类,就把该彩色夜视超像素分为这个类。定义:

$$L(scn_i, c_i) = \log \frac{P(scn_i | c_i)}{P(scn_i | \bar{c}_i)} = \log \prod_k \frac{P(f_{scn_i}^k | c_i)}{P(f_{scn_i}^k | \bar{c}_i)} = \sum_k \log \frac{P(f_{scn_i}^k | c_i)}{P(f_{scn_i}^k | \bar{c}_i)} \quad (1)$$

式中:  $\bar{c}_i$  表示除去类  $c_i$  之外的其余类别;  $f_{scn_i}^k$  表示彩色夜视超像素块  $scn_i$  的第  $k$  个特征。更具体地,令  $D$  ( $D = \sum_{j \in NS} sn_cj$ ) 为数据库  $NS$  上所有的超像素集合。 $N_{snc}^k$  ( $N_{snc}^k = \sum_{j \in RS} sn_cj$ ) 为候选集  $RS$  上满足一定条件的超像素  $sn_cj$  集合。该条件是: 候选集上自然彩色超像素  $sn_cj$  第  $k$  个特征  $f_{sn_cj}^k$  与彩色夜视超像素  $scn_i$  第  $k$  个特征  $f_{scn_i}^k$  之间的距离小于阈值  $t_k$ 。那么,有:

$$\frac{P(f_{scn_i}^k | c_i)}{P(f_{scn_i}^k | \bar{c}_i)} = \frac{(n(c_i, N_{snc}^k + \varepsilon) / n(c_i, D))}{(n(\bar{c}_i, N_{snc}^k + \varepsilon) / n(\bar{c}_i, D))} = \frac{n(c_i, N_{snc}^k + \varepsilon)}{n(\bar{c}_i, N_{snc}^k + \varepsilon)} \times \frac{n(c_i, D)}{n(\bar{c}_i, D)} \quad (2)$$

式中:  $n(c_i, S)$  (同理  $n(\bar{c}_i, S)$ ) 表示集合  $S$  中分类为类别  $c_i$  的超像素个数;  $\varepsilon$  为常量,防止出现概率为零的情况。

在整个计算过程中,使用  $l_2$  范数计算特征之间的距离,并且将每个特征上的阈值  $t_k$  设置成:彩色夜视超像素  $scn_i$  特征  $f_{scn_i}^k$  与自然彩色超像素  $sn_cj$  特征  $f_{sn_cj}^k$  之间,前  $T$  个最短距离的中位数。 $T$  值的选取通过实验确定。

最大化公式(1)概率,使得彩色夜视超像素  $scn_i$  通过最近邻算法从自然彩色候选集上找出与之图像

特征最相似的超像素  $sn_cj$ ,并获取其标签,记为  $c_i$ ,遍历所有超像素,完成整张测试图像的初步场景解析。

### 1.3 马尔科夫修正

第 1.2 节已初步得到彩色夜视图像场景解析的结果,但仍然可能出现分类错误的超像素。比如:一个标签为“路”的超像素若被标签为“天空”的超像素包围,是不可信的。为了加强语境制约,定义马尔科夫随机场能量函数:

$$J(c) = \sum_{scn_i \in SCN} E_{data}(scn_i, c_i) + \lambda \sum_{(scn_i, scn_j) \in A} E_{smooth}(c_i, c_j) \quad (3)$$

式中:  $SCN$  表示彩色夜视超像素集合;  $A$  表示邻近超像素对;  $\lambda$  表示平滑常量。该函数由数据项  $E_{data}(scn_i, c_i)$  和平滑项  $E_{smooth}(c_i, c_j)$  两部分组成,数据项部分为彩色夜视超像素  $scn_i$  分类到似然概率最大的类别  $c_i$ ,平滑项部分联系其邻近超像素  $scn_j$  类别  $c_j$ ,修正局部区域的解析结果。

通过上述步骤,测试图像的每个像素点都得到 1 个语义标签结果  $c_i$ ,与真实标签比较,可以计算每个像素点的识别率  $Q_{per-pixel}$ 。由于像素点识别率高低很大程度上取决于能否正确分类图像中的少数大面积物体。考虑小面积物体的影响,文中还计算了各个物体类别的平均识别率  $Q_{mean-class}$ 。 $Q_{per-pixel}$  和  $Q_{mean-class}$  的值越高,说明该算法解析效果越好。

## 2 实验结果与分析

### 2.1 实验设置

数据库图像为白天采集的可见光图像,采用麻省理工人工智能实验室(MIT CSAIL)创建的开放注释工具——LabelMe 对其进行手工标注。主要包含的语义标签有:“天空”、“树木”、“草地”、“路”、“建筑”5类,其余景物统一标注为“其他”。文中对 5 个场景分别拍摄 40 张不同视角的图像,后期通过镜像处理,数据库扩展至 400 张。

测试图像为红外和微光的融合图像,图 2 给出了某一场景的图像概貌,包括 1 幅自然彩色图像,2 幅源图像和 7 幅融合图像。图 2(a)为自然彩色图像;(b)为红外图像;(c)为微光图像。采用 4 种典型彩色夜视融合方法生成:

方法 1: 统计色彩映射法,由 TNO 人力因素研究所的 Toet 等人开发。对红外和微光生成的伪彩色图像,在  $l\alpha\beta$  色彩空间,赋予参考图像的一阶色彩信

息。使用 4 种不同的自然光照图像作为参考图像,融合图像如图 2 (d)~(g) 所示, 简记为 SM、SM.2、SM.3、SM.4。特别的,图 2(d)使用真实场景(a)做参考图像。

方法 2: NRL 法, 美国海军研究室开发的假彩色方法。红外的灰度值从红色 LUT 中选择, 微光的灰度值从青色 LUT 中选择。如图 2(h)所示, 简记为 NRL。

方法 3: TNO 法。提取红外和微光的主要成分输

入 RG 两个通道, B 通道置零。如图 2(i)所示, 简记为 TNO。

方法 4: 特有成分调制法。归一化微光图像特有成分, 并调制红外图像。如图 2(j)所示, 简记为 TYTZ。

利用这 4 种融合算法, 可得到 7 种不同色彩表现的彩色融合图像。文中对 5 种场景的图像进行融合, 得到 35 幅彩色夜视融合图像作为该实验的测试图像。

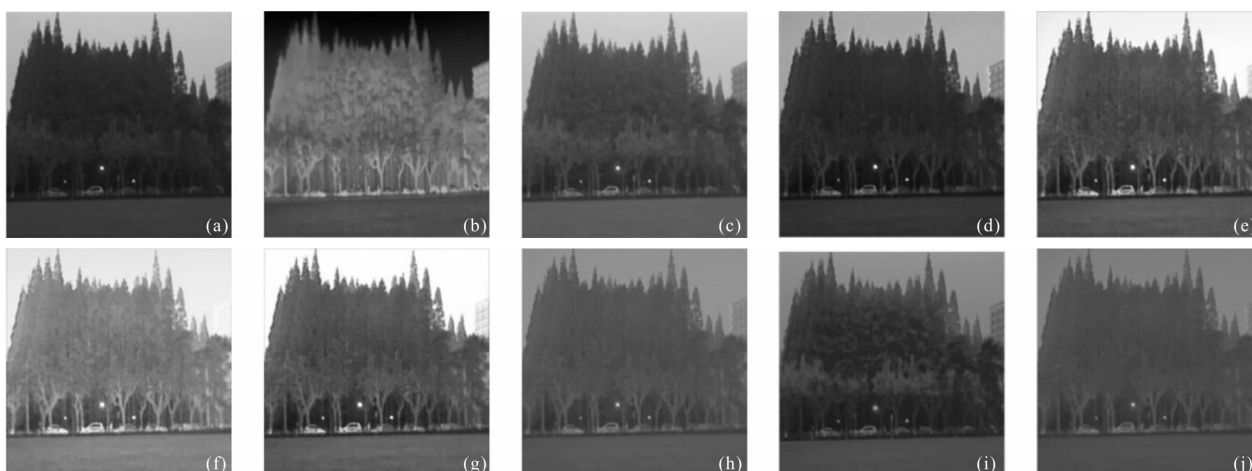


图 2 场景 1

Fig.2 Scene 1

## 2.2 实验结果

对彩色夜视图像进行场景解析, 经多次实验, 选取  $K=200$  为候选集数量,  $T=80$  为设置阈值  $t_k$  时选择的个数。最终得到 5 个场景 35 张测试图像的像素识别率  $Q_{per-pixel}$  和类别平均识别率  $Q_{mean-class}$ , 如表 1~5 所示。

表 1 场景 1 融合图像的各个类别识别率、类别平均识别率和像素识别率

Tab.1 Recognition rate of each class, class average and pixels of fused image in scene 1

Scene1	Tree	Sky	Build	Grass	Road	$Q_{mean-class}$	$Q_{per-pixel}$
SM	98.81	94.53	-	98.61	-	97.32	95.82
SM.2	98.96	90.10	-	98.26	-	95.77	95.02
SM.3	96.69	91.23	-	99.47	-	95.80	94.01
SM.4	98.45	93.64	-	97.73	-	96.61	95.28
NRL	95.53	93.97	-	97.93	-	95.81	93.55
TNO	95.74	78.29	-	99.47	-	91.17	90.98
TYTZ	98.66	76.80	-	99.47	-	91.64	92.59

表 2 场景 2 融合图像的各个类别识别率、类别平均识别率和像素识别率

Tab.2 Recognition rate of each class, class average and pixels of fused image in scene 2

Scene 2	Tree	Sky	Bsld	Grass	Road	$Q_{mean-class}$	$Q_{per-pixel}$
SM	98.98	97.92	-	4.85	94.07	73.96	86.88
SM.2	99.23	96.89	-	0	98.93	73.76	87.02
SM.3	98.17	98.11	-	0	97.34	73.41	86.21
SM.4	98.69	98.11	-	0	95.71	73.13	86.35
NRL	97.58	98.21	-	0	98.89	73.67	86.01
TNO	98.08	93.66	-	0.89	98.73	72.84	86.06
TYTZ	99.19	87.21	-	0	99.35	71.44	86.28

从表中可以看出, 文中的解析方法在各个场景上都有较高的像素识别率和类别平均识别率, 说明文中算法可鲁棒解析各典型彩色夜视图像。分析发现, 场景 2 中“草地”识别率很低, 接近于 0, 而在其他场景中“草地”的识别率却很高。这主要是因为场景 2 的“草地”在拍摄时生长较稀疏, 不够繁茂, 无法体现草的特

表 3 场景 3 融合图像的各个类别识别率、类别平均识别率和像素识别率

Tab.3 Recognition rate of each class, class average and pixels of fused image in scene 3

Scene 3	Tree	Sky	Build	Grass	Road	$Q_{\text{mean-class}}$	$Q_{\text{per-pixel}}$
SM	96.41	98.63	72.26	-	99.70	91.75	86.63
SM.2	98.29	88.65	65.74	-	19.48	68.04	77.99
SM.3	97.47	76.69	70.23	-	6.25	62.66	77.13
SM.4	97.86	93.58	70.36	-	98.39	90.05	85.68
NRL	97.17	85.10	72.60	-	99.31	88.55	85.34
TNO	97.41	91.67	56.94	-	19.96	66.50	74.43
TYTZ	98.05	78.98	48.38	-	99.77	81.30	74.94

表 4 场景 4 融合图像的各个类别识别率、类别平均识别率和像素识别率

Tab.4 Recognition rate of each class, class average and pixels of fused image in scene 4

Scene 4	Tree	Sky	Build	Grass	Road	$Q_{\text{mean-class}}$	$Q_{\text{per-pixel}}$
SM	99.17	98.53	-	95.17	-	97.62	98.24
SM.2	97.29	98.25	-	98.06	-	97.87	97.66
SM.3	98.10	98.31	-	93.05	-	96.49	97.16
SM.4	95.21	98.60	-	92.73	-	95.51	95.52
NRL	92.65	96.33	-	4.94	-	64.64	76.46
TNO	98.84	95.63	-	97.31	-	97.26	97.80
TYTZ	95.33	95.41	-	99.50	-	96.75	96.16

表 5 场景 5 融合图像的各个类别识别率、类别平均识别率和像素识别率

Tab.5 Recognition rate of each class, class average and pixels of fused image in scene 5

Scene 5	Tree	Sky	Build	Grass	Road	$Q_{\text{mean-class}}$	$Q_{\text{per-pixel}}$
SM	92.98	96.62	93.04	89.84	-	93.12	93.33
SM.2	92.30	94.95	97.62	91.28	-	94.04	93.56
SM.3	93.53	98.87	70.60	98.18	-	90.30	91.24
SM.4	90.12	93.78	95.27	89.10	-	92.07	91.54
NRL	87.47	97.72	94.12	77.82	-	89.28	89.43
TNO	95.70	90.48	42.52	99.96	-	82.17	86.39
TYTZ	94.69	95.40	22.20	99.57	-	77.97	83.36

征。因此,该算法在处理特征不强的物体时,仍有所欠缺,后续选择场景时,需考虑物体的特征是否明显。场景 3 中,“路”这一物体在各测试图像上得到的识别率差距比较大,有明显波动,进一步分析得出,该算法对小面积物体的识别率并不稳定,存在一定解析难度。纵观各场景下那些识别率稳定并显著较高的物体,都占据较大面积,因此,针对小面积物体,进一步提高算法识别率,将具有极大研究价值。

对 7 种融合图像计算所有场景下的像素识别率均值和类别平均识别率均值,如表 6 和表 7 所示。发现 7 种融合图像识别率从高到低的分布,大致符合人眼对彩色夜视图像的主观视觉感知。从这一发现可以看出:在纹理特征占据主导地位的前提下,色彩特征在该算法中是一个具有极大价值的互补信息。因此,笔者认为以场景解析的识别率作为一个综合客观评价指标,用于彩色夜视图像质量评价,具有一定的研究前景。

表 6 所有场景类别平均识别率的均值

Tab.6 Mean value of class average recognition rate of all scenes

$Q_{\text{mean-class}}$	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Mean value	Rank
SM	97.32	73.96	91.75	97.62	93.12	90.75	1
SM.2	95.77	73.76	68.04	97.87	94.04	85.90	3
SM.3	95.8	73.41	62.66	96.49	90.30	83.73	5
SM.4	96.61	73.13	90.05	95.51	92.07	89.47	2
NRL	95.81	73.67	88.55	64.64	89.28	82.39	6
TNO	91.17	72.84	66.5	97.26	82.17	81.99	7
TYTZ	91.64	71.44	81.3	96.75	77.97	83.82	4

表 7 所有场景像素识别率的均值

Tab.7 Mean value of pixels recognition rate of all scenes

$Q_{\text{per-pixel}}$	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Mean value	Rank
SM	95.82	86.88	86.63	98.24	93.33	92.18	1
SM.2	95.02	87.02	77.99	97.66	93.56	90.25	3
SM.3	94.01	86.21	77.13	97.16	91.24	89.15	4
SM.4	95.28	86.35	85.68	95.52	91.54	90.87	2
NRL	93.55	86.01	85.34	76.46	89.43	86.16	7
TNO	90.98	86.06	74.43	97.8	86.39	87.13	5
TYTZ	92.59	86.28	74.94	96.16	83.36	86.67	6

### 3 结论

针对彩色夜视图像场景内容丰富多变,需动态扩充数据库的特点,文中提出了一种基于非参数模型的彩色夜视图像解析算法。通过特征匹配,实现语义标签传递,达到解析场景的目的。对5种场景,7种典型彩色夜视图像进行实验,证明该算法可有效解析场景内容。同时,文中发现识别率的高低与彩色夜视图像的主观评价具有一致性,可作为一种客观评价指标,全局评价彩色夜视图像质量,这是后续着重研究的方向。

#### 参考文献:

- [1] Toet A, Ijspeert J K, Waxman A M, et al. Fusion of visible and thermal imagery improves situational awareness [J]. *Displays*, 1997, 18(2): 85-95.
- [2] Zou X T, Bhanu B. Tracking humans using multi-modal fusion [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
- [3] Michael A, Irvine J M. Information fusion for feature extraction and the development of geospatial information[C]//Proceeding of the 7th International Conference on Information Fusion, 2004: 976-982.
- [4] Bulanon D M, Burks T F, Alchanatis V. Image fusion of visible and thermal images for fruit detection [J]. *Biosystems Engineering*, 2009, 103(1): 12-22.
- [5] Jiang Dong, Zhang Dafang, Huang Yaohuan, et al. Survey of multispectral image fusion techniques in remote sensing applications[C]//Image Fusion and Its Applications, 2011: 1-22.
- [6] Daneshvar S, Ghassemian H. MRI and PET image fusion by combining IHS and retina-inspired models [J]. *Information Fusion*, 2010, 11(2): 114-123.
- [7] He Xuming, Zemel R S, Carreira-Perpinan M A. Multiscale conditional random fields for image labeling[C]//Proceedings IEEE Conference Computer Vision and Pattern Recognition, 2004, 2: 695-702.
- [8] Ladicky L, Sturges P, Alahari K, et al. What, where and how many combining object detectors and CRFs [C]//Proceedings European Conference Computer Vision, 2010: 424-437.
- [9] Galleguillos C, Mcfee B, Belongie S, et al. Multi-class object localization by combining local contextual interactions [C]//Proceedings IEEE Conference Computer Vision and Pattern Recognition, 2010: 113-120.
- [10] Gould S, Fulton R, Koller D. Decomposing a scene into geometric and semantically consistent regions[C]//Proceedings IEEE International Conference Computer Vision, 2009: 1-8.
- [11] Divvala S, Hoiem D, Hays J, et al. An empirical study of context in object detection[C]//Proceedings IEEE Conference Computer Vision and Pattern Recognition, 2009: 1271-1278.
- [12] Felzenszwalb P, Mcallester D, Ramanan D A. Discriminatively trained multiscale deformable part model [C]//Proceedings IEEE Conference Computer Vision and Pattern Recognition, 2008: 1-8.
- [13] 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.
- [14] Rahul M. Deep deconvolutional networks for scene parsing [J]. *Computer Science*, 2014, ArXiv: 1411. 4101.