

中图法分类号: TP391.4 文献标志码: A 文章编号: 1006-8961(2011)09-1607-08

论文索引信息: 吴士林, 耿佳佳, 朱枫, 于泳. 基于区域的多类目标识别与分割算法研究 [J]. 中国图象图形学报, 2011, 16(9): 1607-1614

基于区域的多类目标识别与分割算法研究

吴士林^{1),2),3),4)}, 耿佳佳^{1),2),3),4)}, 朱枫^{1),3),4)}, 于泳⁵⁾

¹⁾(中国科学院沈阳自动化研究所, 沈阳 110016) ²⁾(中国科学院研究生院, 北京 100049)

³⁾(中国科学院光电信息处理重点实验室, 沈阳 110016) ⁴⁾(辽宁省图像理解与视觉计算重点实验室, 沈阳 110016)

⁵⁾(Broadcom Corporation 200 Brickstone Square, Andover, MA 01810)

摘要: 为了实现复杂自然场景中多类目标的识别与分割, 利用条件随机场(CRF)对目标特征进行建模, 并在此基础上运用过分割算法将图片分为有限个连续区域, 提出一种新的基于区域的 CRF 模型, 即 R-CRF 模型, 并采用 Joint-boost 算法对标注样本进行训练, 研究基于主题的 R-CRF 模型在多类目标识别与分割中的应用。MSRC-21 类数据库的实验结果表明, 该算法在多类目标识别与分割中取得的结果优于国内外其他算法, 尤其对于其他算法中正确率很低的形状多变而样本少的高结构物体的识别和分割取得了很好的结果。

关键词: 目标识别; 图像分割; 多类; 条件随机场模型; Joint-boost 算法

Region-based multi-class object recognition and segmentation

Wu Shilin^{1),2),3),4)}, Geng Jiajia^{1),2),3),4)}, Zhu Feng^{1),3),4)}, Yu Yong⁵⁾

¹⁾(Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016 China)

²⁾(Graduate School of Chinese Academy of Sciences, Beijing 100049 China)

³⁾(Key Laboratory of Optical-Electronics Information Processing, Chinese Academy of Sciences, Shenyang 110016 China)

⁴⁾(Key Laboratory of Image Understanding and Computer Vision, Shenyang 110016 China)

⁵⁾(Broadcom Corporation 200 Brickstone Square, Andover MA 01810 America)

Abstract: A conditional random field (CRF) model is used to incorporate different feature potentials of objects for multi-class object recognition and segmentation in natural images. By using an over-segmentation algorithm, we propose a new region based CRF model called R-CRF model. We train our model on annotated samples by using Joint-boost algorithm and investigate the performance of the theme based R-CRF model for class based pixel-wise segmentation of images. We compare our results with recent published results on the MSRC 21-class database. The result shows that our theme based R-CRF model significantly outperforms the current state-of-the-art. Especially, by introducing theme and regions, our model obtains greatly improved accuracy of structured classes with high visual variability and fewer training examples.

Keywords: object recognition; image segmentation; multi-class; CRF model; joint-boost

收稿日期: 2010-07-08; 修回日期: 2010-11-10

基金项目: 国家自然科学基金项目(60705011)。

第一作者简介: 吴士林(1981—), 男。中国科学院沈阳自动化研究所模式识别与智能系统专业博士研究生, 主要研究方向为计算机视觉、机器学习。E-mail: wushilin@sia.cn。

0 引言

对图像中多类目标同时进行识别与分割是计算机视觉领域的核心问题之一,受到越来越多的关注^[1-5]。不同于识别某一类目标^[6-11],多类目标识别在对类间的区别特征进行建模的同时保持对类内特征差异的鲁棒性;与传统的分割方法相比,多类目标识别与分割在对图像进行分割的同时实现识别,使得分割后的各个区域具有类别上的语义。目前,对单一类别的目标进行同步识别与分割已经取得不错的成果,文献[10]提出一种基于 Radon 变换和解析 Fourier-Mellin 变换的尺度与旋转不变的单一目标的识别算法。文献[11]提出一种基于贝叶斯模型的相机间人群目标识别。上述各种单一目标的识别方法在处理多视角或严重遮挡情形时遇到困难。几种多类目标模型^[1-2]解决了这个问题,并能很好地识别与分割高纹理特征的大片区域,但是在处理高结构特征的物体时收效甚微,特别是当这些物体的训练样本相对较少且变化较大时正确率很低。

为了实现多类目标的同步识别与分割,在对各类目标特征(如纹理、位置)进行建模的同时,还需要对各类间的相互约束关系进行建模,如天空一般在草地的上面,船的周围一般是水等。这些特征既有全局性的也有区域性的,并且具有很大的不确定性。本文利用条件随机场(CRF)^[12]来实现这些特征的融合。CRF 模型是一种基于无向图的概率模型,用来对序列图像进行标识。该模型可以利用彼此重叠的特征而不需要很强的独立性假设,因此,越来越多地被用来实现前后文相关数据的识别,其在计算机视觉领域的应用也得到了很大发展^[1,13],主要用来融合多重图像特征以实现复杂的视觉任务。

在此基础上,通过引入过分割算法提出一种新的基于区域的 CRF 模型(R-CRF 模型),为了降低误分割的影响,在增大其他特征项权重的基础上,通过引入图片主题来约束各类条件分布。MSRC-21 类数据库的实验结果表明,基于主题的 R-CRF 模型能很好地对图片中的多类目标同时进行识别和分割,这些图片包括从任意视角、包含遮挡的高结构特征和高纹理特征的目标。MSRC-21 类数据库是目前最为复杂和全面的被标注的数据

库之一,由 591 幅图片组成,标注有 21 类物体。图 1 给出了数据库的示例图片及其标注图。标注图中不同颜色代表不同的物体类别,黑色部分代表该点不属于 21 类中的任何一类,在训练、调试和测试过程中都被忽略。在本文中,将 MSRC-21 类数据库随机地分成 3 部分,45% 的训练集,10% 的调试集和 45% 的测试集。

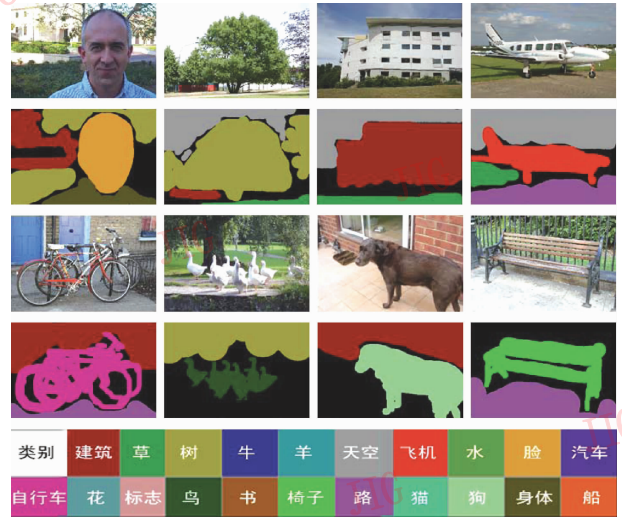


图 1 MSRC-21 类数据库的示例图片及其标注图
Fig. 1 Example images with corresponding groundtruth from MSRC-21 class database

1 基于 CRF 的特征融合模型

由于各类目标的非规范表现(如牛有时躺着,有时立着)以及类与类之间约束关系的不确定性(如鸟有时在空中,有时在水面上),很难对属于某个类别的特征进行精确建模,而只能建立其概率模型。由于存在多个特征类型,如纹理分布、位置等,需要一种模型来实现它们的融合。在此,我们设计了多个特征函数,并用 CRF 模型来实现它们的融合。

1.1 CRF 模型

CRF 模型^[12]是一种基于无向图的概率模型,用来对序列数据进行标识,具有很强的概率推理能力。与经典概率模型隐马尔可夫模型(HMM)相比,CRF 模型主要有两个优点:1)对观测序列没有独立性要求;2)可以对长间隔序列的复杂上下文关系进行描述。因此,CRF 模型能够适用于目标特征间有重叠、像素上下文关系复杂等情形的视觉任务。关于

CRF 模型的详细内容请参阅文献[12],这里简述其基本框架。对于输入观测序列 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, 对其进行标注 $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}$ 的条件概率为

$$P(\mathbf{Y} | \mathbf{X}; \boldsymbol{\theta}) \propto \exp\left(\sum_t f(y_t, y_{t-1}, x, t; \boldsymbol{\theta})\right) \quad (1)$$

式中, f 为特征函数。标识的任务是求取 \mathbf{Y}^* , 满足

$$\mathbf{Y}^* = \arg \max_{\mathbf{Y}} P(\mathbf{Y} | \mathbf{X}) \quad (2)$$

权重参数 $\boldsymbol{\theta}$ 通过在训练样本 $\{x_m, y_m\}_{m=1,2,\dots,M}$ 上最大化下列似然函数得到

$$L_{\theta} = \sum_m \sum_t f(y_t^m, y_{t-1}^m, x^m, t; \boldsymbol{\theta}) \quad (3)$$

1.2 基于 CRF 的特征融合

为了实现多类目标的识别与分割,我们首先构建一个 CRF 模型来学习不同类别的条件分布。对于给定图片 \mathbf{x} 中的某个点 i , 其属于类别 c_i 的条件概率定义为

$$P_1(c_i | \mathbf{x}; \boldsymbol{\theta}) \propto \exp\{f^{\text{texture}}(c_i, \mathbf{x}; \boldsymbol{\theta}_0) + f^{\text{texture-env}}(c_i, \mathbf{x}; \boldsymbol{\theta}_1) + f^{\text{location}}(c_i, i; \boldsymbol{\theta}_2)\} \quad (4)$$

$\boldsymbol{\theta} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ 是模型参数,将在下文中逐一介绍。为简明起见,下面的讨论中提到各特征函数 f 时将省略其参数。

1.2.1 纹理特征

首先对图像进行纹理编码,利用 17 维的多通道多尺度高斯型滤波器组对训练集中的所有图片进行滤波^[14],从而图片中每个像素点对应一个 17 维向量。这些滤波器组包括不同尺度和通道下的基本高斯型、 x 和 y 方向的一阶偏导型以及拉普拉斯型,因此,每个向量中包含了相应像素点的区域纹理信息。然后在欧氏空间中利用 K 均值算法对滤波结果进行聚类。对每幅图片中所有像素点赋予其最近邻聚类中心的编号值,得到相应图片的纹理图。给定图片 \mathbf{x} 中纹理编号为 t 的某个点 i 的纹理特征的定义为

$$f^{\text{texture}}(c_i, \mathbf{x}) = \omega_0 \ln \frac{N_{c_i,t} + \alpha_0}{N_t + \alpha_0} \quad (5)$$

其参数为 $\boldsymbol{\theta}_0 = \{\omega_0, \alpha_0\}$ 。 $N_{c_i,t}$ 是训练集里所有图片中纹理为 t , 类别为 c_i 的点的个数,而 N_t 是这些图片中纹理为 t 的所有点的个数。图 2 给出了几幅图片的纹理图及其对应的原始图片,纹理图中不同颜色代表不同的纹理编号(值)。从图中可以看出,纹理特征包含了颜色、边缘等信息。我们相

信,纹理特征包含了足够的信息来进行基于点的分割,因而本模型不再引入其他模型中常用的颜色、边缘等特征。



图 2 纹理图的示例图片及其原始图片

Fig. 2 Example texture maps with original images

1.2.2 纹理环境特征

一个像素属于某个类别的概率不仅仅与其本身的纹理信息有关,还与其周围像素的纹理分布有关,如某个像素周围具有草特征的纹理分布,则其是草或牛的概率将较大。为此,我们建立纹理环境特征函数项 $f^{\text{texture-env}}(c_i, \mathbf{x})$ 。该特征函数反映了点 i 的纹理环境信息,既包括同一物体不同纹理的上下文信息,也包含该点与其他物体纹理的上下文信息,其定义为

$$f^{\text{texture-env}}(c_i, \mathbf{x}) = \ln P_i(c_i | \mathbf{x}) \quad (6)$$

与文献[1]中的形状纹理特征函数 $\psi_i(c_i, \mathbf{x}; \boldsymbol{\theta}_{\psi})$ 相同。其中, $P_i(c_i | \mathbf{x}) = \frac{\exp H(c_i)}{\sum_{c'_i} \exp H(c'_i)} \propto \exp H(c_i)$ 。

分类器 $H(c_i)$ 及参数 $\boldsymbol{\theta}_1$ 采用文献[1]中介绍的特征选取方法获得,该方法将在 1.2.4 节中详细介绍。

1.2.3 位置特征

不同类别的目标在图片中的绝对位置一般不同,如天空一般在图片的上方,草地一般在图片的下方。本文中,不同类别在图片中绝对位置信息用位置特征来表示,即

$$f^{\text{location}}(c_i, i) = \omega_2 \ln \left(\frac{N_{c_i,i} + \alpha_2}{N_i + \alpha_2} \right) \quad (7)$$

其参数为 $\boldsymbol{\theta}_2 = \{\omega_2, \alpha_2\}$ 。此处 i 是指点在归一化的图片中的位置。 $N_{c_i,i}$ 表示训练集里所有图片中位于 i 类别为 c_i 的点的个数,而 N_i 是这些图片中位于 i 的所有点的个数。

1.2.4 算法

为了提高效率,本文中各个特征函数通过带有权值的分块训练方法^[15]分别进行训练,除纹理环境

特征,其余都已在各特征函数中说明。纹理环境特征中的归一化分布函数 $P_i(c_i | \mathbf{x})$ 中的分类器 $H(c_i)$ 通过以下方式获得:采用矩形滤波器来捕捉基于类别的纹理分布信息,每个矩形滤波器为一个纹理分布特征产生种子,其定义为一个五元组 $\nu = \nu(l, w, x, y, t)$ 。其中 l 为矩形的长度, w 为宽度, x, y 为矩形左上角相对于像素 i 的位置, t 为纹理编号。像素点 i 的矩形滤波器的响应值为

$$\nu_{(l,w,x,y,t)}(i) = \frac{\sum_p [T(p) = t]}{\text{area}(\nu)} \quad (8)$$

式中, $T(p)$ 表示点 p 的纹理编号, $\text{area}(\nu) = l \times w$ 表示矩形的面积, $[x]$ indicator 函数, x 为真时取 1, 否则取 0。在实际运行中,矩形滤波器在特定范围 (200×200) 内随机产生。矩形滤波器参数 $(\theta_\nu = \{l, w, x, y, t\})$ 的优化选取通过 Joint-boosting 算法^[16]实现。最终得到一个强分类器 $H(c)$, 它是多个带权值的弱分类器的和。

$$H(c) = \sum_{m=1}^M h_m(c, \nu) \quad (9)$$

式中,每个弱分类器是一个决策函数,

$$h_m(c, \nu) = \begin{cases} a[\nu_{\theta_\nu} > \delta] + b & c \in \mathbf{C} \\ k^\nu & c \notin \mathbf{C} \end{cases} \quad (10)$$

由优化选取的一个 $\nu(l, w, x, y, t)$ 特征和一个控制特征值的阈值 δ 组成,并可适用于某几个类 \mathbf{C} 。 k^ν 的作用是减弱各类在样本中分布不均的影响。每轮迭代产生一个弱分类器,通过优化以下目标函数获得

$$h_m = \operatorname{argmin} \sum_{c=1}^{|\mathbf{C}|} \sum_{i=1}^N w_{i,m-1}^c (z_i^c - h(c, \nu))^2 \quad (11)$$

式中, N 为训练集中所有像素点的个数, z_i^c 为训练样本中像素 i 的标注值,当其属于 c 时为 1, 否则为 -1。 $w_{i,m}^c$ 为权重控制项,对以前分错的样本相应加重权值,从而使得整个分类器向好的方向推进。

$$w_{i,m}^c = \exp(-z_i^c H_m(c, \nu_i)) = w_{i,m-1}^c \exp(-z_i^c h_m(c, \nu_i)) \quad (12)$$

以上为 Joint-boosting 算法的基本框架,文献^[1,16]给出了详尽的算法描述。这样,通过 Joint-boosting 的优化最终得到参数 $\{\theta_\nu, \mathbf{C}, a, b, \delta, \{k^c\}_{c \in \mathbf{C}}\}_m$ 。CRF 中的其他特征函数在前文中已经分别说明。现在构造似然函数

$$L_\theta = \sum_n \sum_i \ln P_1(c_i | \mathbf{x}_n; \theta_0, \theta_2) \quad (13)$$

为了防止过拟合,在调试(而非训练)样本中通过最大化 L_θ 求取参数 θ_0, θ_2 。

当 CRF 模型中各个参数都通过学习选定后,测试时通过最大化条件概率(式(4))就可以得到点 i 的类别 c^* 。

2 基于区域的 CRF 模型(R-CRF)

以上为像素级的识别与分割,没有充分考虑到目标的结构信息。本节以上述 CRF 模型的条件概率为初值,通过引入过分割算法,建立基于区域的 CRF 模型,即 R-CRF 模型。

2.1 过分割算法

为了与识别相结合,要求采用的分割算法具有这样的全局特性:不属于同一目标的像素不要分割在同一个区域。Felzenszwalb^[17]于 2004 年提出一种基于图结构的过分割算法,并证明了该算法能够实现具有全局特性的、高效率的分割。其分割算法框架如下。

用一个无向图模型 $G = (\mathbf{V}, \mathbf{E})$ 来表示一幅待分割图片,其中 $\mathbf{V} = \{\nu_i\}$ 为待分割像素点集合, $\mathbf{E} = \{(\nu_i, \nu_j)\}$ 为边集合,每个边对应一个权重 $w(\nu_i, \nu_j)$ 。不同于文献^[17],这里的边权重定义为

$$w(\nu_i, \nu_j) = \sum_{r,g,b} |d(\nu_i, \nu_j)|^2 \quad (14)$$

式中, $d(\nu_i, \nu_j)$ 表示像素点 ν_i, ν_j 各颜色值 (r, g, b) 的差。分割的目的是找到一个区域集合 $\mathbf{S} = \{\mathbf{R}_1, \dots, \mathbf{R}_r\}$ 使得

$$\cup \mathbf{R}_i = \mathbf{V} \text{ 且当 } i \neq j \text{ 时 } \mathbf{R}_i \cap \mathbf{R}_j = \emptyset \quad (15)$$

单个区域的内差异函数定义为

$$\text{Int}(\mathbf{R}) = \max_{(\nu_i, \nu_j) \in \mathbf{E}, \nu_i, \nu_j \in \mathbf{R}} w(\nu_i, \nu_j) \quad (16)$$

两区域的内差异函数定义为

$$\text{MInt}(\mathbf{R}_i, \mathbf{R}_j) = \min(\text{Int}(\mathbf{R}_i) + \tau / |\mathbf{R}_i|, \text{Int}(\mathbf{R}_j) + \tau / |\mathbf{R}_j|) \quad (17)$$

式中, $|\mathbf{R}|$ 为区域 \mathbf{R} 的像素点个数。当 $|\mathbf{R}| = 1$ 时, $\text{Int}(\mathbf{R}) = 0$ 。参数 τ 为一常数, τ 越大,得到的区域越大,本文中取 $\tau = 150$ 。

算法流程

输入:图 $G = (\mathbf{V}, \mathbf{E})$, 有 n 个像素点, m 条边。
输出:将像素点集合 \mathbf{V} 分割成区域集合 $\mathbf{S} = \{\mathbf{R}_1, \dots, \mathbf{R}_r\}$ 。

- 1) 将 E 进行递增排序 $\{o_1, \dots, o_m\}$ 。
- 2) 令初始值 S^0 为: 每个像素点 v_i 分别组成一个区域。
- 3) 对于 $q = 1, \dots, m$ 重复步骤 4)。
- 4) 用以下方法从 S^{q-1} 构造 S^q : 设 v_i, v_j 代表第 q 条边所连接的两个像素, 即 $o_q = (v_i, v_j)$, 设 $v_i \in R_i^{q-1}, v_j \in R_j^{q-1}$ 且 $R_i^{q-1}, R_j^{q-1} \in S^{q-1}$ 。当 $R_i^{q-1} \neq R_j^{q-1}$ 且 $w(o_q) \leq \text{MInt}(R_i, R_j)$ 时, 合并 R_i^{q-1} 和 R_j^{q-1} 得到 S^q , 否则令 $S^q = S^{q-1}$ 。

5) 若区域 R_i 小于最小允许面积 S_{\min} , 将 R_i 与 R_j 合并, 直到所有区域均大于 S_{\min} 。最终 $R_j^* = \min_{R_j} \text{MInt}(R_i, R_j)$ 且 R_i 与 R_j 相邻。

如图 3 所示, 在给定图片 x 中, 不同的连续区域用随机产生的不同颜色来代表, 不同的图片中的相同颜色区域并无关联。本文中最小允许面积 S_{\min} 为 80 个像素大小。



图 3 过分割图片及其原始图片

Fig. 3 Over-segmentation results with original images

2.2 R-CRF 模型

对于给定图片 x 中的某个区域 R_i , 定义其属于类别 c_i 的条件概率为

$$P_2(c_i, R_i | x; \theta) \propto \exp\left(\frac{1}{|R_i|} \sum_{i \in R_i} \ln P_1(c_i | x; \theta)\right) \quad (18)$$

由于参数 θ 已知, 通过最大化此条件概率就可以得到区域 R_i 的类别 c^* 。对于区域 R_i 中的每个点都赋予相同的类别 c^* , 即可实现基于点的分类与分割。区域 R_i 中点 i 的条件概率记为

$$P_2(c_i | x; \theta) = P_2(c_i, R_i | x; \theta) \quad (19)$$

$$i \in R_i$$

3 基于主题的 R-CRF 模型

MSRC-21 类数据库中的图片属于不同的场景类别, 不同的场景类别对其中包含的目标类别有一

定的约束关系。如机场中不会出现牛。为了对这种约束关系进行建模, 我们引入图片的主题这个概念, 并研究在原有 R-CRF 模型中如何利用主题这一约束关系对目标的特征进行建模, 从而提高分类分割的正确率和精度。

通过将图片的主题 St 引入 R-CRF 模型, 我们构建了一个新的基于主题的 R-CRF 模型。对于给定图片 x 中的某个点 i , 其所属类别 c_i 的条件概率定义为

$$P_3(c_i | x, St; \theta) \propto \exp\{f^{\text{texture}}(c_i, x, St; \theta_0) + f^{\text{texture-env}}(c_i, x, St; \theta_1, \theta_s) + f^{\text{location}}(c_i, i, St; \theta_2)\} \quad (20)$$

式中, $St \in \mathbf{St}$ 表示给定图片 x 的主题, $\theta = \{\theta_0, \theta_1, \theta_2, \theta_s\}$ 是模型参数, 其中 $\theta_0, \theta_1, \theta_2$ 与前同, θ_s 将在下文中介绍。

图片的主题可从文件名中自动获得, 文件名以“1”到“20”开头。在同一主题的所有图片中, 某些类物体会频繁出现。我们将文件名以“1”、“5”、“9”开头的图片都赋予“草原”的主题, 因为它们的主题很相像。类似地, 文件名以“3”、“17”开头的图片都以“街道”为主题, 而文件名以“18”、“20”开头的图片都以“河流”为主题, 从而, 该数据库中所有图片属于 16 个不同的主题 $\mathbf{St} = \{0, 1, 2, \dots, 15\}$ 。

3.1 纹理特征

引入主题 St , 得到新的纹理特征

$$f^{\text{texture}}(c_i, x, St) = \omega_0 \ln \frac{N_{c_i, t, St} + \alpha_0}{N_{t, St} + \alpha_0} \quad (21)$$

式中, $N_{c_i, t, St}$ 是训练集中主题为 St 的所有图片中纹理为 t , 类别为 c_i 的点的个数, 而 $N_{t, St}$ 是这些图片中纹理为 t 的所有像素点的个数。

3.2 纹理环境特征

引入主题 St , 定义一个新的归一化分布函数

$$P_i(c_i | x, St) = \frac{P_2(c_i | x) P(c_i | St)}{\sum_{c'_i \in \mathbf{St}} P_2(c'_i | x) P(c'_i | St)} \propto P_2(c_i | x) P(c_i | St) \quad (22)$$

式中, $P_2(c_i | x)$ 为 R-CRF 模型得到的条件概率, 其定义见式 (19)。 $c_i \in \mathbf{St}$ 代表类别 c_i 存在于主题为 St 的某些图片中。得到新的纹理环境特征

$$f^{\text{texture-env}}(c_i, x, St) = \ln P_i(c_i | x, St) = \ln P_2(c_i | x) + \ln P(c_i | St) \quad (23)$$

$$c_i \in \mathbf{St}$$

式中

$$\ln P(c_i | S_t) = \omega_s \ln \frac{N_{c_i, S_t} + \alpha_s}{N_{S_t} + \alpha_s} \quad (24)$$

其参数为 $\theta_s = \{\omega_s, \alpha_s\}$ 。 N_{c_i, S_t} 是训练集中主题为 S_t 的所有图片中类别为 c_i 的点的个数, 而 N_{S_t} 是这些图片中所有点的个数。在调试过程中, 选取参数 θ_s 的值为最小化数据库中调试集的误差。

3.3 位置特征

引入主题 S_t , 定义新的位置特征

$$f^{\text{location}}(c_i, i, S_t) = \omega_2 \ln \frac{N_{c_i, i, S_t} + \alpha_2}{N_{i, S_t} + \alpha_2} \quad (25)$$

式中, N_{c_i, i, S_t} 是训练集中主题为 S_t 的所有图片中位于归一化位置 i , 类别为 c_i 的点的个数, 而 N_{i, S_t} 是这些图片中位于 i 的所有点的个数。

当基于主题的 R-CRF 模型中各个参数都通过学习选定后, 测试时通过最大化条件概率式(20)就可以得到点 i 的类别 c^* 。

4 实验结果和分析

如图 4 所示, 图片中的绝大部分都被正确地识别与分割, 取得很好的结果。例如, 虽然建筑与飞机、建筑与天空的纹理很接近, 这 3 类仍被很好地识别和分割; 另外, 飞机阴影处的草地与周围草地纹理差异很大, 却被正确地识别成草地。对于基于主题的 R-CRF 模型的识别和分割结果中出现的误差, 图 4 也给出了样图。从图中可以看出, 船的识别存在一定的误差。船的顶部和底部被错分成建筑, 因为此图中船的纹理与外形和数据库中的建筑极为相似。但是船的主体仍然被很好地识别出来。从表 1 中可以看出, 基于主题的 R-CRF 模型在整个数据库中的识别精度达到了 79.2%, 显著优于已有的其他方法。图 5 给出了基于主题的 R-CRF 模型中各个特征项对识别与分割结果的影响。从图 5(c) 可以看出, 由于引进了主题和区域, 仅仅采用纹理环境特征, 基于点的识别和分割精度就取得了显著提高, 从 64.4% 提高到 74.9%。图 5(d) 表明, 纹理特征提高了分割的视觉效果, 基于点的分割正确率提高到 78.5%。整个模型的分割精度为 79.2%, 如图 5(e) 所示。表 2 给出了整个数据库中每一类的基于点的分割精度。表中每行数据代表某一真实的类被分成各类的百分比。表中对角线代表每一类基于点的正确率。

从表 2 可以看出, 21 类中大多数类别的分割精度在 70% 以上。此外, 从表 2 可以看出, 基于主题的 R-CRF 模型很好地解决了以往各方法中存在的问题, 即对变化较大而训练样本较少的高结构特征物体的识别与分割的正确率很低的问题。例如, 对于类别 {标志, 鸟, 椅子, 狗, 船}, 文献 [1] 的精度为 {35.1%, 19.4%, 15.4%, 19.2%, 6.6%}, 文献 [2] 的精度为 {54.8%, 23.0%, 39.6%, 49.6%, 14.0%}。本文中基于主题的 R-CRF 模型将这些类别的分割精度提高到了 {74.0%, 45.6%, 66.3%, 63.7%, 26.4%}。

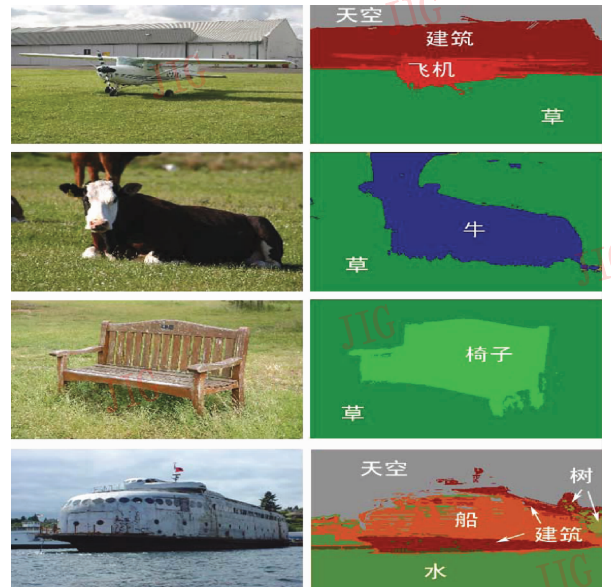


图 4 基于主题的 R-CRF 模型在 MSRC-21 类数据库的实验结果样图

Fig. 4 Example results of theme based R-CRF model on MSRC 21-class database

表 1 基于主题的 R-CRF 模型和其他模型在 MSRC-21 类数据库的结果比较

Tab. 1 Comparison of accuracy of theme based R-CRF model with that of other works on MSRC-21 database

| 算法 | MSRC-21/% |
|----------------|-----------|
| 文献[1] | 72.2 |
| 文献[2] | 75.4 |
| 文献[3] | 71.7 |
| 文献[4] | 76.5 |
| 文献[5] | 76.4 |
| 基于主题的 R-CRF 模型 | 79.2 |

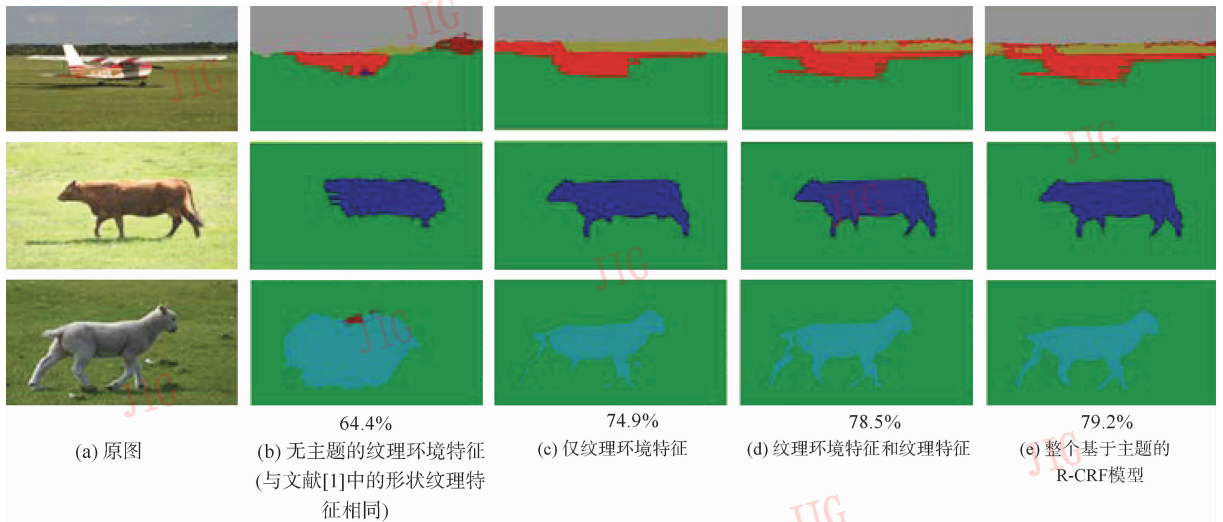


图5 基于主题的 R-CRF 模型中不同特征的分割精度比较

Fig. 5 Average pixel-wise accuracy of different model potentials of theme-based R-CRF model

表2 每一类的基于点的分割精度(对角线)

Tab. 2 Pixel-wise accuracy (across all folds) for each class (rows)

| 类别 | 建筑 | 草 | 树 | 牛 | 羊 | 天空 | 飞机 | 水 | 脸 | 汽车 | 自行车 | 花 | 标志 | 鸟 | 书 | 椅子 | 路 | 猫 | 狗 | 身体 | 船 |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 建筑 | 82.5 | 0.3 | 3.9 | | | 2.0 | 0.1 | 1.0 | 0.3 | 1.6 | 1.8 | | 1.4 | 0.1 | 0.5 | | 4.4 | | | 0.1 | 0.1 |
| 草 | 0.2 | 96.0 | 2.3 | 0.3 | 0.4 | | 0.1 | | | | | | | | | 0.3 | 0.2 | | | 0.1 | |
| 树 | 8.9 | 6.6 | 74.1 | | | 3.8 | 0.1 | 1.6 | | | | | | 0.1 | 0.3 | 1.6 | 0.8 | | 0.1 | 0.3 | 1.7 |
| 牛 | 1.8 | 15.8 | 2.3 | 79.7 | 0.2 | 0.1 | | | | | | | | | | | 0.1 | | | | |
| 羊 | 0.4 | 18.0 | 0.8 | 3.0 | 73.0 | 0.8 | | | | | | | | | | | 4.0 | | | | |
| 天空 | 7.2 | | 0.4 | | | 76.9 | 0.5 | 4.7 | | | | | | 9.8 | | | | | | | 0.3 |
| 飞机 | 11.4 | 4.2 | 0.5 | | | 3.2 | 79.9 | | | | | | | | | | 1.0 | | | | |
| 水 | 20.2 | 5.2 | 1.2 | 0.6 | | 5.0 | | 53.4 | | | | | | 2.9 | | | 7.6 | | 1.0 | 2.4 | 0.5 |
| 脸 | 15.7 | 0.3 | 0.8 | | | | | | 66.9 | | | 0.5 | | | 9.5 | | 0.2 | | | 6.1 | |
| 汽车 | 10.1 | 0.9 | | | | 0.1 | | | | 85.2 | | | | | | | 3.2 | | | 0.4 | |
| 自行车 | 12.7 | 2.1 | 1.9 | | | | | | | 1.0 | 77.6 | | | | | | 4.7 | | | | |
| 花 | 3.1 | 2.3 | 2.7 | | | | | | | | | 91.7 | | | | | | | | | 0.1 |
| 标志 | 17.6 | | 7.6 | | | 0.7 | | | | | | | 74.0 | | | | | | | | |
| 鸟 | 6.8 | 9.8 | | | | 1.1 | | 16.4 | | | | | | 45.6 | | | 19.9 | | | | 0.5 |
| 书 | 3.2 | 0.3 | 0.9 | | | | | | 0.3 | | | | | | 93.7 | | 0.1 | | | 1.5 | |
| 椅子 | | 9.0 | | | | | | | | | | | | | | 66.3 | 21.7 | | | | |
| 路 | 5.7 | 0.2 | 0.8 | | | 0.6 | 0.6 | 0.2 | 0.4 | 1.2 | 1.4 | | | 1.3 | | 1.0 | 84.7 | 0.7 | 0.3 | 1.0 | |
| 猫 | | | | | | | | | | | | | | | | | 18.0 | 82.0 | | | |
| 狗 | 9.9 | 2.7 | 5.3 | | | | | | 2.7 | | | | | | | | 11.5 | | 63.7 | 4.2 | |
| 身体 | 10.7 | 4.1 | 2.9 | | | | | 0.8 | 6.1 | 2.8 | | 2.7 | | | 2.3 | | 4.6 | | | 63.1 | |
| 船 | 51.7 | | 1.1 | | | 3.9 | | 14.5 | | | | | | | | | 1.2 | | | 1.2 | 26.4 |

注:表中数据代表真实的类被分成各类的百分比。

5 结 论

利用 CRF 模型实现多个特征的融合,运用过分割算法将图片分为有限个连续区域,提出一种新的 R-CRF 模型。通过引入区域和主题,构建“基于主题的 R-CRF 模型”,并研究其在多类目标识别与分割中的应用。在 MRSC-21 类数据库的实验结果表明,本文算法在多类目标识别与分割中取得的结果明显优于已有的算法,尤其对于其他算法中正确率很低的形状多变而样本少的高结构物体的识别和分割取得了很好的结果。

参考文献 (References)

- [1] Shotton J, Winn J, Rother C, et al. TextonBoost: joint appearance, shape and context modeling for multi-class object recognition and segmentation [C]//Proceedings of European Conference on Computer Vision. New York, USA: Springer, 2006, 3951: 1-15.
- [2] Yang L, Meer P, Foran D J. Multiple class segmentation using a unified framework over mean-shift patches [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2007: 1-8.
- [3] Schroff F, Criminisi A, Zisserman A. Object class segmentation using random forests [C]//Proceedings of the 19th British Machine Vision Conference. Leeds, England: BMVA, 2008: 1-10.
- [4] Gould S, Rodgers J, Cohen D, et al. Multi-class segmentation with relative location prior [J]. International Journal Computer Vision, 2008, 80(3): 300-316.
- [5] Gould S, Fulton R, Koller D. Decomposing a Scene into geometric and semantically consistent regions [C]//Proceedings of International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2009: 1-8.
- [6] Winn J, Shotton J. The layout consistent random field for recognizing and segmenting partially occluded objects [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2006, 1: 37-44.
- [7] Opelt A, Pinz A, Zisserman A. Incremental learning of object detectors using a visual shape alphabet [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2006, 1: 3-10.
- [8] Winn J, Jovic N. Locus: learning object classes with unsupervised segmentation [C]// Proceedings of International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2005, 1: 756-763.
- [9] Kumar P, Torr P, Zisserman A. Obj cut [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2005, 1: 18-25.
- [10] Wang Xuan, Xiao Bin, Ma Jianfeng. Scaling and rotation invariant analysis approach to object recognition based on radon and analytic Fourier-Mellin transforms [J]. Journal of Image and Graphics, 2008, 13 (1) : 2157-2162. [王暄, 肖斌, 马建峰. 基于 Radon 和解析 Fourier-Mellin 变换的尺度与旋转不变目标识别算法 [J]. 中国图象图形学报, 2008, 13 (1) : 2157-2162.]
- [11] Deng Yingna, Zhu Hong, Liu Wei. Bayesian human recognition across multiple cameras in crowded situations [J]. Journal of Image and Graphics, 2009, 14 (9) : 1751-1755. [邓颖娜, 朱虹, 刘薇. 基于贝叶斯模型的相机间人群目标识别 [J]. 中国图象图形学报, 2009, 14 (9) : 1751-1755.]
- [12] Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data [C]//Proceedings of International Conference on Machine Learning. MA, USA: The International Machine Learning Society, 2001: 282-289.
- [13] He X, Zemel R S, Ray D. Learning and incorporating top-down cues in image segmentation [C]//Proceedings of European Conference on Computer Vision. New York, USA: Springer, 2006, 3951: 338-351.
- [14] Winn J, Criminisi A, Minka T. Object categorization by learned universal visual dictionary [C]//Proceedings of International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2005, 2: 1800-1807.
- [15] Sutton C, McCallum A. Piecewise training of undirected models [C]//Proceedings of 21st Conference on Uncertainty in Artificial Intelligence, Edinburgh, Scotland: AUAI Press, 2005: 568-575.
- [16] Torralba A, Murphy K P, Freeman W T. Sharing visual features for multiclass and multiview object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 19(5): 854-869.
- [17] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation [J]. International Journal of Computer Vision, 2004, 59 (2) : 167-181.