

基于 HAB 优化算法的图像语义 目标对象提取研究

陈久军^{1),2)} 肖刚¹⁾ 高飞¹⁾ 高济²⁾ 张元鸣¹⁾

¹⁾(浙江工业大学信息学院, 杭州 310014) ²⁾(浙江大学计算机学院, 杭州 310027)

摘要 提出了一种基于优化 Adaboost 算法(HAB 优化算法)的半监督图像语义目标对象获取方法。在分析 Adaboost 算法评估函数不足的基础上,设计并实现 HAB 优化算法。对比实验结果表明,HAB 优化算法在训练误差与抗干扰能力方面具有更好的性能。在此基础上,研究应用 HAB 优化算法的图像语义目标对象获取方法,从图像对象特征预处理、对象识别器训练、语义对象获取 3 个方面进行论述。通过实验分析,该方法具有良好的图像目标对象获取性能。

关键词 目标对象获取 HAB 优化算法 图像分析 图像语义

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2007)08-1359-08

Research on Image Semantic Object Extracting Method Based on HAB Optimized Algorithm

CHEN Jiu-jun^{1),2)}, XIAO Gang¹⁾, GAO Fei¹⁾, GAO Ji²⁾, ZHANG Yuan-ming¹⁾

¹⁾(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014)

²⁾(College of Computer Science and Technology, Zhejiang University, Hangzhou 310027)

Abstract The paper presents a novel approach to extract a semantic image object based on an optimized Harmonious Adaboost algorithm, shortly HAB, which produces less generalization error and high performance compared to the Gentle Adaboost Algorithm. Some key techniques in the proposed schema, including the pre-processing of image character, the training of object detector and the extracting of semantic image object, are discussed. The experiment shows that the recurrent training process improves the performance of the object detector, and the extracting results demonstrate the availability of the work.

Keywords semantic object extraction, optimized HAB algorithm, image analysis, image semantic

1 引言

图像“语义鸿沟”^[1]问题是图像分析与理解研究中的重要问题。其研究核心在于,如何从图像中提取出可以映射到相关语义的目标对象。

图像语义目标对象是与人类理解相关的图像相连区域,是视觉属性与语义属性的结合体。常用的

获取方法包括:(1) 基于场景方法^[2-4],这种方法将整个图像作为图像特征的潜在模式,以全局视觉属性来描述图像内容。它的优点是不需要引入图像分割技术。但由于整个图像所对应的视觉特征,不能很好地区分不同的图像语义概念,因此解释效果不是很好。(2) 基于图像区域的方法^[5-7],这种方法将图像中具有相关颜色与纹理特征的相连区域作为图像潜在对象模式。不足之处是根据颜色和纹理所

基金项目:国家“973”项目(2003CB317000);浙江省科技厅项目(2005C31010)

收稿日期:2007-01-25;改回日期:2007-03-16

第一作者简介:陈久军(1977 -),男,讲师。2006年获浙江大学博士学位。主要研究方向为图像分析、机器视觉。E-mail: rackycj@

zjut.edu.cn

得到的图像区域并不能很好地直接对应于语义对象,如果两者的吻合程度不好,就不能获取完整的图像语义信息,那么该区域所具有的视觉特征也就不能很好地用来区分不同的语义概念。(3) 基于注释的方法^[8,9],图像对象语义内容通过人工标注方式与图像区域建立映射,使得底层区域视觉特征可以较好地解释对应的语义对象。不足之处是,对大量的图像进行语义标注是一个十分繁琐的过程。(4) 基于对象的方法^[10-12],该方法将人类所理解的语义对象作为图像模式,它的关键是自动获取和语义相关的图像空间对象。其核心工作是从图像背景中区分出具体的对象。目前,存在两方面的挑战:混合场景中多对象的识别^[13]与多样性对象的通用模式获取。SVM^[14]、Boosting^[15-18]是图像空间对象获取的常用技术。其中,Boosting 算法能有效地从复杂图像场景中发现个别的对象类别^[15]。

如何设计更有效的训练方法,以提高图像语义对象的获取的准确性和高效性,是本文所要研究的重点。文中研究了一种基于优化 AdaBoost 算法的半监督图像语义目标对象获取方法,并分析该方法所具有的良好性能。

2 优化 AdaBoost 算法设计

2.1 Boosting 算法的基本思想

Boosting 算法是将弱学习算法提升为强学习算法的有效途径。基本思想是在训练过程中,不断调整训练样本的权重值,提高学习算法的性能。Friedman^[16]已从理论层次证明了 Boosting 算法具有良好近似性能。Boosting 算法目前已有很多改进方案,如 AdaBoost、Gentle AdaBoost 等。这些算法在定义评估函数时,忽略了对已被正确分类的训练样本的性能描述因子,容易出现过拟合现象,影响训练效果。

本文定义的优化 AdaBoost 算法(命名为 Harmonious AdaBoost,简称“HAB”)综合权衡了训练过程中被正确分类的样本集合($CreExample_m$)和错误分类的样本集合($WroExample_m$),用以获取一个综合协调的权重更新值:(1)对 $WroExample_m$ 中具有较低边界值的样本赋予较高的权重值,使其在下一训练次中,对这些错误分类的训练样本进行重点学习;(2)对 $CreExample_m$ 中的部分样本,如果其边界值很大,说明其被正确分类的可信度很高,为了避

免出现过拟合现象,应调整其权重值。

2.2 Boosting 算法的基本思想

HAB 优化算法从样本权重分布概率的角度,定义了权重和逆权重的概念,对权重更新值的幅度给出了综合的定义,使所获取的新的权重更为合理,确保分类器在新的循环训练中,将学习重点集中在那些最需要训练的样本上。从而降低算法的训练误差,提高抗干扰能力。

令 (x_i, y_i) 表示训练集合, $i \in (1, N)$, N 表示训练样本数目。 $y_i \in (+1, -1)$, $i \in (1, N)$ 表示训练样本的类别集合。利用 HAB 优化算法将弱分类器 $h(x)$ 提升为强分类器 $H(x)$ 。

定义 1 边界值 边界值描述了弱分类器对训练样本所得到的分类结果及其可信程度。

令 $margin(x, y)$ 表示样本 (x, y) 的边界值, δ 表示极大值:

若满足 $margin(x, y) > 0$, 且 $margin(x, y) \rightarrow \delta$, 表示弱分类器对样本 (x, y) 做出了可信的正确分类;

若满足 $margin(x, y) < 0$, 且 $margin(x, y) \rightarrow \delta$, 表示弱分类器对样本 (x, y) 做出了可信的错误分类。

定义 2 正确分类的样本集 定义 $CreExample_m$ 表示在进行第 m 次循环训练之前,已经被正确分类的训练样本的集合。 $CreExample_m$ 中的训练样本满足:

$$margin(CreExample_m(i), y) > 0 \quad i \in (1, l) \quad (1)$$

定义 3 错误分类的样本集合 定义 $WroExample_m$ 表示在进行第 m 次循环训练之前,被错误分类的训练样本的集合。 $WroExample_m$ 中的训练样本满足:

$$margin(WroExample_m(i), y) < 0 \quad i \in (1, k) \quad (2)$$

定义 2 与定义 3 中 l, k 分别表示各自集合中包含的样本数目,并且满足: $l + k = N$ 。

定义 4 权重 ω_m 定义权重满足

$$\omega_{i+1} = \omega_i e^{-\gamma h_m(x_i)} \quad (3)$$

其中, $h_m(x_i)$ 为弱分类器。

权重 ω_i 将根据前一次循环的分类结果调整:

(1) 对于 $WroExample_m$ 中的样本 (x_i, y_i) , 由于 $y_i h(x_i) < 0$, 因此新的权重值将会增大;

(2) 同理, 对 $CreExample_m$ 中的样本, 由于 $y_i h(x_i) > 0$, 其权重将被减小。

定义 5 逆权重 $\bar{\omega}_m$ 逆权重将大的权重值赋给

CreExample_m 中的样本,逆权重与权重的关系如下式:

$$\bar{\omega}_m(i) = (1 - \omega_m(i))\bar{Z}_m \quad (4)$$

其中,ω_m(i)表示在第 m 次循环中,第 i 的训练样本的权重值;ω̄_m(i)表示第 m 次循环中,第 i 的训练样本的逆权重;Z̄_m表示标准化参数,满足使 ω̄_m(i) 满足下式的条件:

$$\sum_{i=1}^N \bar{\omega}_m(i) = 1 \quad (5)$$

基于权重 ω_m 与逆权重 ω̄_m 的定义,对每一次循环得到的 h_m(x),分别计算 h_m(x) 与 y 在权重分布(ω_m)与逆权重分布(ω̄_m)上的条件概率,用来衡量当前循环中弱分类器在 CreExample_m 样本与 WroExample_m 样本上的分类性能。以下两式分别给出了对 P_m⁺¹(x) 与 P_m⁻¹(x) 的定义。

$$P_m^{+1}(x) = P_{\omega_m}(y = +1 | h_m(x)) \quad (6)$$

$$P_m^{-1}(x) = P_{\omega_m}(y = -1 | h_m(x)) \quad (7)$$

P_m⁺¹(x) 表示了在第 m 次循环后得到的 h_m(x) 对分类标记为 y = +1 的训练样本错误分类的概率值;P_m⁻¹(x) 表示了在第 m 次循环后得到的 h_m(x) 对分类标记为 y = -1 的训练样本错误分类的概率值。

定义 CreExample_m 中样本的性能描述因子 P̄_m⁺¹(x) 与 P̄_m⁻¹(x):

$$\bar{P}_m^{+1}(x) = P_{\bar{\omega}_m}(y = +1 | h_m(x)) \quad (8)$$

$$\bar{P}_m^{-1}(x) = P_{\bar{\omega}_m}(y = -1 | h_m(x)) \quad (9)$$

P̄_m⁺¹(x) 表示了在第 m 次循环后得到的 h_m(x) 对分类标记为 y = +1 的训练样本正确分类的概率值;P̄_m⁻¹(x) 表示了在第 m 次循环后得到的 h_m(x) 对分类标记为 y = -1 的训练样本正确分类的概率值。P̄_m⁺¹(x) 与 P̄_m⁻¹(x) 衡量了分类器 h_m(x) 在 CreExample_m 上的分类性能。

综合式(6)~(9)建立 HAB 优化算法的性能评估函数

$$f_m(x) = (P_m^{+1}(1 - \bar{P}_m^{+1}) - P_m^{-1}(1 - \bar{P}_m^{-1}))(x) \quad (10)$$

相对于 Gentle AdaBoost 算法的评估函数,式(10)综合考虑了当前循环下弱分类器 h_m(x) 在 CreExample_m 样本与 WroExample_m 的性能描述因子,更全面地描述 h_m(x) 当前循环的训练情况。

将式(10)所定义的评估函数代入式(3),可以获取优化权重更新:

$$\omega_{m+1}(i) = \frac{\omega_m(i) \exp(-y_i f_m(x_i))}{Z_m} \quad (11)$$

其中,ω_{m+1}(i)表示权重更新后的第 i 个样本权重

值;ω_m(i)为第 m 次循环中的第 i 个样本权重;y_i ∈ (+1, -1), i ∈ (1, N) 为训练样本类别标识。

分析式(10)与式(11),当 P̄_m⁺¹ = 1 或 P̄_m⁻¹ = 1 时, f_m(x) = 0, 此时 ω_{m+1}(i) = ω_m(i), 表示权重值不再变化。P̄_m⁺¹ = 1 与 P̄_m⁻¹ = 1 表示当前弱识别器对所有分类标识为 y = ±1 的训练样本进行了正确的分类,循环训练自动结束。

综合以上分析,设计图像特征训练的 HAB 优化算法,具体步骤如下:

输入:训练图像特征: v^t, k ∈ (1, numTrain); 训练图像二元标识: c^t = {c^t}, c^t ∈ (+1, -1)

循环次数: N;

输出:强识别器: H(v^t)

初始化:初始权重 ω₀(i) = 1/numTrain;

For m = 1, ..., N and f_m ≠ 0 do

(1) 训练弱分类器: h_m(v) = a(v^t > θ) + b

(a) 获取 h_m = arg min_Δ {J_{wcc}(i)}, 其中

$$J_{wcc}(i) = \sum_{i=1}^{numTrain} \omega_i (c^t - (a(v^t > \theta) + b))^2 \quad i \in (1, numTrain),$$

获取满足 min(J_{wcc}(i)) 时的 θ̂, â, b̂ 值:

$$\hat{b} = \frac{\sum_i \omega_m(i) z_i (v^t \leq \hat{\theta})}{\sum_i \omega_m(i) (v^t \leq \hat{\theta})}$$

$$\hat{a} = \frac{\sum_i \omega_m(i) z_i (v^t > \hat{\theta})}{\sum_i \omega_m(i) (v^t > \hat{\theta})} - b$$

(b) 更新弱分类器参数 h_m(v)

$$h_m(v) = \hat{a}(v^t > \hat{\theta}) + \hat{b}$$

(2) 定义评估参数:

$$f_m(v) = (P_m^{+1}(1 - \bar{P}_m^{+1}) - (P_m^{-1}(1 - \bar{P}_m^{-1}))) (v)$$

(3) 更新权重值:

$$\omega_{m+1}(i) = \frac{\omega_m(i) \exp(-c^t f_m(v^t))}{Z_m}$$

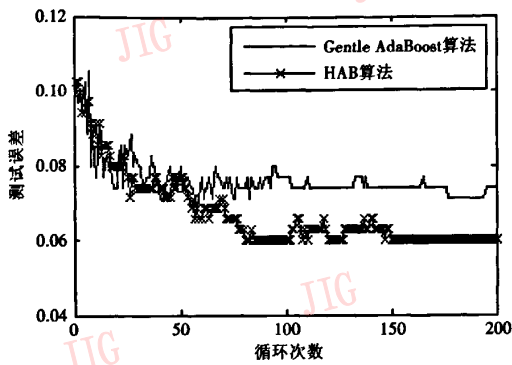
End for

获取优化分类器 H(v^t) = sign [∑_{i=1}^N f_m(h_m(v))]

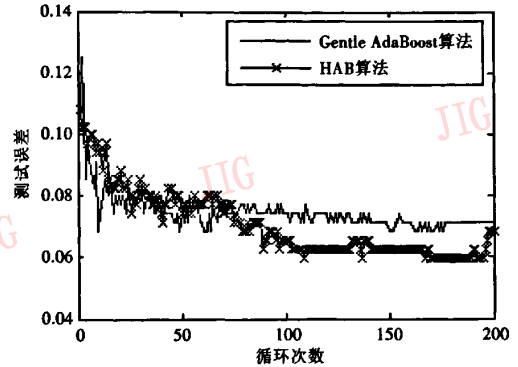
2.3 HAB 优化算法性能比较

基于交通图像特征训练库,定义 Stumps 作为弱分类器,进行算法比较实验。

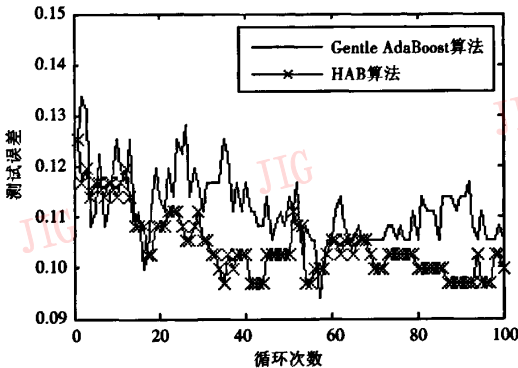
图 1(a)与图 1(b)表示了 K 次迭代交叉验证(K = 5, 9)下, HAB 优化算法与 Gentle AdaBoost 算法的训练误差比较。从图中可以发现, HAB 优化算



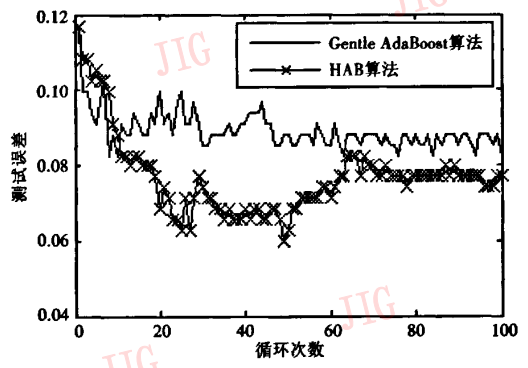
(a) 5次迭代实验误差比较



(b) 9次迭代实验误差比较



(c) 20%噪声干扰性能比较



(d) 30%噪声干扰性能比较

图 1 HAB 测试误差比较实验

Fig. 1 Testing error ratio of HAB

法在经过多次的循环训练后,具有较小的训练误差,其训练准确程度更高。

图 1(c)与图 1(d)比较了两者的抗干扰能力。在图像特征样本数据中增加 20% 和 30% 的噪声数据。根据训练误差值,可以发现,HAB 算法具有更好的抗干扰能力,说明其鲁棒性能更好。

3 应用 HAB 算法的图像目标对象获取

基于 HAB 优化算法的图像目标对象获取是一种半监督的图像学习方法,如图 2 所示。假设需要获取的图像目标语义对象 $Object_i$,通过以下步骤训练:

(1) 图像特征预处理,获取训练图像“特征池”。从已经标注的图像库中,选取部分训练图像,用来提取图像子块特征 $Patches$, $Patches$ 包含了用于描述 $Object_i$ 的不同实例下的相关底层特征;用 $Patches$ 的特征描述与所有训练图像进行特征匹配,

根据是否满足匹配条件,将训练图像分为正样本与负样本两类,用 (\pm) 进行标识。训练图像库“特征池”就包含了两种特征信息:底层特征描述、关于 $Object_i$ 的正负样本标识。

(2) 图像目标对象识别器训练。将“特征池”中的信息作为输入,应用 HAB 算法,训练优化图像二元分类器。

(3) 应用训练得到的强识别器完成对未标注图像的空间对象获取。

该方法主要的优点在于:(1) 利用 HAB 算法所具有的高准确性与高抗干扰能力,确保训练得到的语义对象识别器具有良好性能。(2) 对训练样本特征进行“两级处理”,降低了训练样本的复杂程度:第 1 级,用可变大小“图像子块”描述同一语义对象在不同实例下所具有的典型特征;第 2 级,将“图像子块”特征描述与训练样本进行特征匹配,从中获取用于语义对象 $Object_i$ 训练的“特征池”。(3) 由

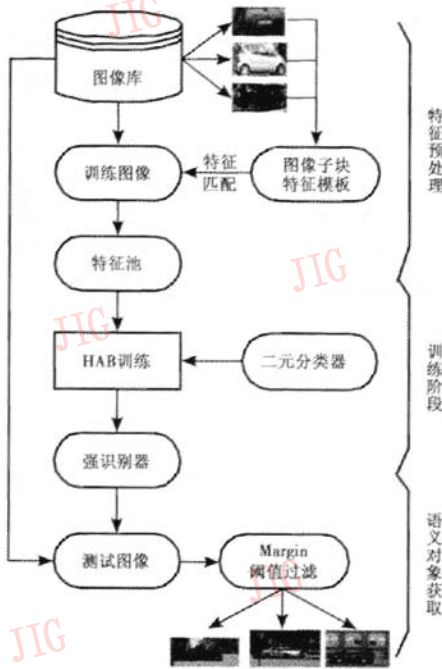


图 2 图像语义目标对象获取
Fig. 2 Image semantic object extraction

于“特征池”中包含了不同实例的典型特征,使训练得到的识别器具有很强的对象获取性能。(4)不需要对语义对象的不同图像姿态实例进行单独离散训练,而是采用联合训练的方式,一方面可以减少需要训练的识别器的数量,降低训练复杂度;另一方面,由于同一语义对象的不同姿态实例在一定程度上都会有共同的特征,通过共享这些特征来提高训练的效率。

3.1 图像特征预处理

定义图像数据库 $D = \{I^{D, Train}, I^{D, Test}\}$, 包含两部分内容:训练图像 ($I_i^{D, Train}$) 和测试图像 ($I_i^{D, Test}$)。训练图像是对目标对象进行语义标注的图像,标注的信息包括图像空间对象的边界多边形以及对应的元语义描述。

训练图像有两个作用:第一,用来提取关于语义对象的图像子块特征描述;第二,获取“特征池”,作为“弱识别器”训练学习的样本数据。测试图像用来测试“强识别器”的性能。

训练图像 $I_k \in I^{D, Train}$, 包含 U_k 个图像语义对象 $Object_i$ $i \in (1, U_k)$ 。 $Object_i$ 定义为一个三元组: $Object_i^k \langle O_{k,i}, F_{k,i}(Patch_{k,i,j}), X_{k,i} \rangle$ 。

其中, $O_{k,i}$ 表示第 k 个图像中,第 i 个对象的语义标识,它通过语义标注的方式获取; F_i 是图像子块的特征描述;

$Patch_{k,i,j}$ $k \in (1, N), i \in (1, U_k), j \in (1, M)$ 表示第 k 个图像中第 i 个语义对象的第 j 个子块;

$X_{k,i}$ 是对象的空间几何特征,如边界多边形的中心位置 C_i 。

图像相关信息 I_k 表示为

$$I_k: \{Object_i^k \langle O_{k,i}, F_{k,i}(Patch_{k,i,j}), X_{k,i} \rangle \mid i \in (1, U_k), j \in (1, M), k \in (1, N)\}$$

三元组 $Object_i^k \langle O_{k,i}, F_{k,i}(Patch_{k,i,j}), X_{k,i} \rangle$ 中,关于 $F_{k,i}(Patch_{k,i,j})$ 的描述是关键。由于统计方法的引入,对图像特征的描述,已经从简单的图像空间特性或频率特性等基本特征描述,逐步转变为对图像纹理内在要素分布特性的统计表达。

要有效地描述图像潜在的纹理统计特征,需要选取合适的图像子空间,本文方法采样多个可变大小的图像子块来描述纹理统计特征。

下式描述了从图像 I_k “特征致密区”与“特征稀疏区”采样,获取 m 个图像子块。

$$Patches = \{Patch_i^K\}_{j=1}^m, K = \{dense, sparse\} \quad (12)$$

图像子块 $Patches$ 的特征描述有两方面用途:

第一,作为图像语义对象 $Object_i$ 的特征模板 $Template_{i,j}$; $Patches$ 在“特征致密区”内进行采样,满足

$$Template_{i,j} = \{Patch_j^{dense}\}_{j=1}^m \quad (13)$$

将采样区集中在 $Object_i$ 的“特征致密区”,可以确保所获取的图像子块准确、集中的描述图像语义对象 $Object_i$ 所具有的典型特征;

第二,描述训练图像的特征信息: $Patches$ 分别从“特征致密区”与“特征稀疏区”进行特征采样,满足式(12)。

(1) 图像语义对象特征模板 $Template_{i,j}$ 的获取

① 图像标准化处理:定义 σ 为图像缩放比例,将图像标准化处理,表示为 I_σ 。相关大小尺寸如表 1 所示。

表 1 图像尺寸标准化处理
Tab. 1 Image size standardization

训练图像	图像子块	标准化对象	测试图像
128 × 200	8 × 8 - 25 × 25	48 × 128	256 × 256

② 选取训练样本:选取经过标准化处理,并且包含语义标识 O_i 的 M_i 个训练图像,作为提语义对

象 $Object_i$ 的训练样本:

$$\{I_{k\sigma}; I_k, O_i = O_i, I_k \subset I^{p, \text{Train}}\}_{k=1}^{M_p} \quad (14)$$

③ 图像子块获取

将标准化处理的图像与预定义的微分算子进行卷积操作 $|I_{128 \times 200} * f_l|$, 其中, $f_l, l \in (1, 3)$ 为预定义的微分算子。

定义图像掩模 $M(3 \times 3)$, 在 $Object_i$ 的“特征致密区”内随机取 n 个采样点 $\{\langle x_j, y_j \rangle\}_{j=1}^n$, 并在 $8 \times 8 \rightarrow 25 \times 25$ 大小范围内随机确定对应于 $\langle x_j, y_j \rangle$ 对象子块大小 $size_j$ 和 f_l 。

给定 M_p (选取 $M_p = 8$) 个训练图像, 对每一个图像分别应用微分算子 f_l 进行操作, 分别从每个图像中获取 20 个图像子块。因此对于给定的 $Object_i$, 总共提取了 $8 \times 3 \times 20 = 480$ 个图像子块。这些图像子块特征共同组成了用于描述图像语义对象 $Object_i$ 的特征模板。

$$Template_{i,j} = \{Patch_j^{\text{dense}} \langle x_j, y_j, size_j, f_l \rangle\}_{j=1}^{480} \quad (15)$$

图 3 是从训练库中提取的部分“Car”对象的图像样本。所标注的语义信息包含了对象边界、相关几何特征、对应语义等。

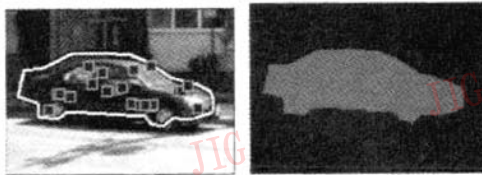


图 3 图像特征子块的选取与对象分割

Fig. 3 Image character blob sampling and segmentation

图 3 中, 左图是“Car”对象的图像子块采样, 右图是对汽车轮廓的图像分割。这些采样的图像子块用于描述“Car”统计纹理特征, 共同构成了“Car”的特征模板。

图 4 是部分图像子块描述, 它们具有不同的大小, 范围在 8×8 到 25×25 之间。通过这种可变大小的图像子块采样, 可以最大程度地描述出不同 $Object_i$ 实例所具有的典型纹理统计特征。

(2) 训练图像的二元标识

训练图像的特征描述同样采用图像子块的方式。根据式(12)的定义, 分别从“特征致密区”和“特征稀疏”获取图像子块。其中从“特征致密区”获取的图像子块 $Patch_j^{\text{dense}}$ 描述了图像所包含的语义对象 $Object_i$ 的特征; 从“特征稀疏区”获取图像子

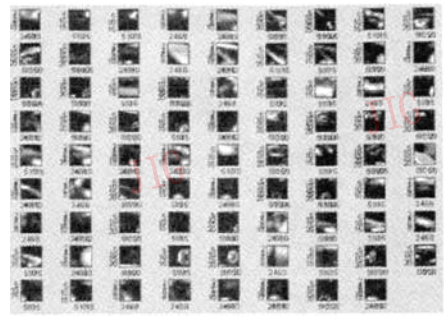


图 4 “car”对象的图像子块

Fig. 4 Object blobs sampling of car

块 $Patch_j^{\text{sparse}}$ 描述图像背景特征。

由于采用了二元分类器作为弱识别器, 因此需要对训练图像样本进行二元标识: 用 $Object_i$ 的特征模板 $Template_{i,j}$ 与训练图像的特征进行匹配, 如果满足匹配条件, 则该训练图像特征为正样本, 标记为 (+1); 如果不满足匹配条件, 则该图像为负样本, 标记为 (-1)。

定义 $F(x, \sigma)$ 为训练图像中, 在 x 位置, 以 σ 比例下的特征向量集合。 $F^h(x, \sigma)$ 表示其中第 h 个特征描述, 将训练图像与特征模板 $Template_{i,j}$ 特征匹配: $F^h(x, \sigma) = (M * |I_\sigma \oplus Template_{i,j}|^p)^{1/p}, j \in (1, 480)$

$$(16)$$

其中, x 是采样点位置, σ 是比例系数, I_σ 是标准化处理的图像; $Template_{i,j}$ 表示语义对象为 $Object_i$ 的特征模板; M 为图像掩模; 系数 p 用来生成不同类型的特征。对式(17)进行简化:

$$F^h(x, \sigma) \approx \max_{x \in S_M} \{|I_\sigma \otimes P_j|\} \quad (17)$$

其中, $S_M(x)$ 表示 x 位置图像掩模的支持度。

将训练图像的特征集合 $F(x, \sigma)$ 进行二元标识: 满足匹配条件的特征标记为 (+1); 不满足的特征标记为 (-1)。

下式定义了训练图像的“特征池”, 包括图像子块特征和对应的二元分类标识 (± 1)。

$$Pool = \{F(\langle F_i^{\text{sparse}}, x_i, y_i \rangle, \langle F_j^{\text{dense}}, x_j, y_j \rangle) \in (-1, +1)\} \quad (18)$$

图 5 中左图表示了训练图像的特征采样, 右图表示了经过特征匹配之后的图像边界值描述。

3.2 目标对象识别器的训练

训练图像经过特征预处理后, 作为输入, 应用 HAB 优化算法, 对弱识别器(二元分类器)进行训练。



图 5 训练图像特征采样
Fig. 5 Training image character sampling

选取 M_i ($M_i = 200$) 个训练图像,为了避免过拟合问题,这些训练图像中不包含用于获取 $Object_i$ 的特征模板 $Template_{i,j}$ 的图像。

$$\{I_{k^*}: I_k, O_i = O_i, I_k \in I^{D, Train}, k \in (1, (M_i - M_p))\}$$

定义四元组 $\{v^h, a, b, \theta\}$, 其中 v^h 表示特征池中的第 h 个特征描述; $\{a, b, \theta\}$ 为“弱识别器”参数, 定义

$$f_m(v) = a(v^h > \theta) + b \quad (19)$$

算法的输入是训练图像特征池中的底层特征与对应的二元分类标识 (± 1)。

利用 HAB 优化算法对训练样本进行循环迭代式的学习。在开始循环之前,设定每一个图像训练样本的权重初值为 $w_0(i) = 1/numTrain$ 。训练的本质是对每一次循环所得到的弱识别器,评估其分类性能 $f_m(v)$, 由此对相应的图像样本的权重值进行调整,确保在下一次循环中,将训练学习的重点放在那些被错误分类的样本上。

当循环完成,或满足 $f_m \neq 0$ 时,训练结束,输出强分类器: $H(v^h) = \text{sign} \left[\sum_{m=1}^N f_m(h_m(v)) \right]$ 。

4 图像语义对象的获取

应用以上方法,可以训练出用来获取多语义目标对象识别器,本实验中训练得到了“car”、“tree”、“mountain”等相关语义对象强识别器。

利用相关识别器,可以对测试图像 $\{I_{k^*}: I_k \in I^{D, Test}\}_{k=1}^{M_i}$ 进行语义目标对象的获取。识别器可以获得测试图像 I_{k^*} 的优化的边界值图,根据所设定的阈值 θ ,标识边界值大于 θ 的图像区域,实现对语义对象的有效获取。

图 6 描述了“car”语义对象获取的两个实验结果。标记为 (a) 的图像列描述了需要进行语义对象获取的原始图像;标记为 (b) 的图像列描述了识别

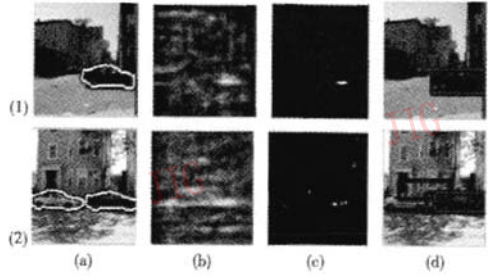


图 6 “car”语义对象的获取
Fig. 6 Car semantic object Extraction

器对原始图像的优化边界值输出;标记为 (c) 的图像列表示阈值过滤;标记为 (d) 的图像列表示语义对象的获取。第 1 个实验结果都较好地获取了相关的语义对象,在第 2 个实验中,对其中一个“car”实例的标记出现了偏差。分析该实例的 (b) 与 (c) 子图像,识别器已经识别出了语义对象的边界值区域,由于阈值设置过低导致标记偏差。可以通过调整阈值来提高获取的准确程度。

为了验证强识别器的性能,定义两个性能评估参数:识别精确度和识别回报率。

定义 6 识别精确度 给定需要获取的语义对象 $Object^k$, 测试图像集合中,包含了 $\{O_i^k\}_{i=1}^n$ 个 $Object^k$ 的实例,通过识别器,正确获取 $\{O_j^k\}_{j=1}^m$ ($m \leq n$) 个 $Object^k$ 的对象实例,则定义两者之间的比值为识别精度:

$$Precision = \frac{|\{O_j^k\}_{j=1}^m| (m \leq n)}{|\{O_i^k\}_{i=1}^n|} \times 100\% \quad (20)$$

定义 7 识别回报率 识别器获取的对象数目 $\{O_j^k\}_{j=1}^m$ ($m \leq n$) 与图像中包含的语义对象总数 $\{O_i^k\}_{i=1}^n$ 的比值:

$$Recall = \frac{|\{O_j^k\}_{j=1}^m| (m \leq n)}{|\{O_i^k\}_{i=1}^n|} \times 100\% \quad (21)$$

图 7 描述了 10 次循环迭代训练与 50 次循环迭代训练的“car”对象识别器的性能比较。

从图中可以发现,50 次循环迭代训练得到的识别器在性能上远远优于 10 次循环训练得到的识别器。说明通过 HAB 算法的训练,图像语义对象识别性能得到了很大的提高。

5 结论

提出了基于 HAB 优化算法的图像语义目标对

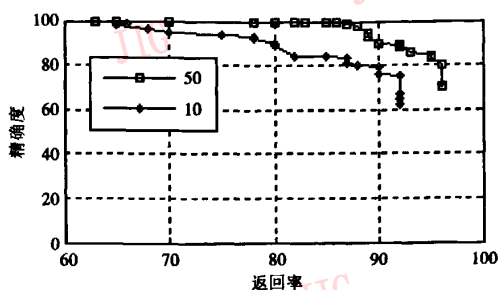


图 7 语义对象识别性能

Fig. 7 Semantic object extracting performance

象获取方法。该方法优点在于:(1) 定义了全面的性能评估函数 $f_m(v)$, 能对当前循环中的弱识别器的分类性能做出综合的评价, 确保循环的有效性;(2) 采用了可变大小的图像子块描述方式, 对同一语义目标对象的不同姿态实例进行的特征采样, 确保了训练样本的完整性;(3) 基于性能评估函数 $f_m(v)$ 的最优化的权重更新模式, 确保了每一次训练都集中在最需要学习的样本上, 尤其是对出现错误分类的样本;(4) 在 $f_m(v)$ 定义中加入了被正确分类的图像样本的性能因子, 确保其具有合理的边界值, 避免出现过拟合现象。实验结果表明, 基于 HAB 算法的图像语义对象获取方法具有良好的识别性能和准确度。

参考文献 (References)

- Smeulders A W M, Worring M, Santini S, *et al.* Content-based image retrieval at the end of the early years[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 22(12): 1349 ~ 1380.
- Chang E, Goh K, Sychay G, *et al.* CBSA: Content-based annotation for multimodal image retrieval using bayes point machines[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, 13(1): 26 ~ 38.
- Vailaya A, Figueiredo M, Jain A K, *et al.* Image classification for content-based indexing[J]. *IEEE Transactions on Image Processing*, 2001, 10(1): 117 ~ 130.
- Smith J R, Li C S. Image classification and querying using composite region templates[J]. *Computer Vision and Image Understanding*, 1999, 75(1-2): 165 ~ 174.
- Carson C, Belongie S, Greenspan H, *et al.* Blobworld: image segmentation using expectation-maximization and its application to image querying[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(8): 1026 ~ 1038.
- Wang J Z, Li J, Wiederhold G. SIMPLcity: semantic sensitive integrated matching for picture libraries[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(9): 947 ~ 963.
- Wang W, Song Y, Zhang A. Semantic-based image retrieval by region saliency[A]. In: *Proceedings of International Conference on Image and Video Retrieval (ICVR 2002)* [C], London, UK, 2002: 29 ~ 37.
- Benitez A B, Chang S F. Semantic knowledge construction from annotated image collections[A]. In: *Proceedings of the 2002 International Conference on Multimedia & Expo (ICME 2002)* [C], Lausanne, Switzerland, 2002: 26 ~ 29.
- Zhu X, Huang T S. Unifying keywords and visual contents in image retrieval[J]. *IEEE Multimedia*, 2002, 9(2): 23 ~ 33.
- Weber M, Welling M, Perona P. Towards automatic discovery of object categories[A]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)* [C], Hilton Head Island, South Carolina, 2000, (2): 101 ~ 108.
- Luo J, Eitz S. A physical model-based approach to detecting sky in photographic images[J]. *IEEE Transactions on Image Processing*, 2002, 11(3): 210 ~ 212.
- Li S, Lv X, Zhang H J. View-based clustering of object appearances based on independent subspace analysis[A]. In: *Proceedings of the International Conference on Computer Vision (ICCV 2001)* [C], Vancouver, Canada, 2001: 295 ~ 300.
- Torralla A, Murphy K P, Freeman W T. Sharing features: efficient boosting procedures for multiclass object detection[A]. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)* [C], San Diego, USA, 2004: 762 ~ 769.
- Papageorgiou C, Poggio T. A trainable system for object detection[J]. *International Journal of Computer Vision*, 2000, 38(1): 15 ~ 33.
- Viola P, Jones M. Robust real-time object detection[J]. *International Journal Computer Vision*, 2004, 57(2): 137 ~ 154.
- Freund Y. Boosting a weak learning algorithm by majority[J]. *Information and Computation*, 1995, 121(2): 256 ~ 285.
- Amores J, Sebe N, Radeva P. Efficient object-class recognition by boosting contextual information[A]. In: *Proceedings of 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2005)* [C], Estoril, Portugal, 2005: 28 ~ 35.
- Yokono J J, Poggio T. Boosting a Biologically Inspired Local Descriptor for Geometry-free Face and Full Multi-view 3D Object Recognition[R]. MIT-CSAIL-TR-2005-046, Cambridge, USA: The MIT Computer Science and Artificial Intelligence Laboratory, 2005.