

一种基于直方图特征和 AdaBoost 的图像中的文字定位算法

李 闯 丁晓青 吴佑寿

(清华大学电子工程系, 智能技术与系统国家重点实验室, 北京 100084)

摘 要 图像中的文字自动定位是计算机视觉领域中的一个新兴研究热点。为了使得定位算法能够适应不同类型的图像和文字, 根据文字所具有的特殊纹理属性, 提出了一种具有普适能力的基于直方图特征和 AdaBoost 的文字定位算法。该算法首先通过提取对文字具有较强鉴别能力的直方图特征和引入 AdaBoost 算法来设计级联结构的纹理分类器; 然后用该分类器的概率输出来生成文字概率图; 在此基础上再通过 CAMSHIFT 算法得到最终的定位结果。实验结果表明, 该算法具有较强的鲁棒性, 能够适应文字在语种、字体、尺度等方面的变化, 在不同类型的图像中都能得到较好的定位结果。

关键词 文字定位 直方图特征 自适应提升 文字概率图 连续自适应均值平移

中图分类号: TP391.43 文献标识码: A 文章编号: 1006-8961(2006)03-0325-07

An Algorithm for Text Location in Images Based on Histogram Features and AdaBoost

LI Chuang, DING Xiao-qing, WU You-shou

(State Key Laboratory of Intelligent Technology and Systems, Department of Electronic Engineering, Tsinghua University, Beijing 100084)

Abstract Automatic text location in images plays an important role in image content understanding, and draws attentions of researchers in the domain of computer vision. Current text location algorithms are mostly adaptive to specific applications; they are sensitive to the variation of text or images and lack robustness. This paper presents a universal approach for text location based on histogram features and AdaBoost. The new algorithm extracts histogram features, which have strong discriminabilities for text and non-text. AdaBoost algorithm with cascade structure is introduced to design the classifier for text texture. The algorithm transfers the binary output of the texture classifier into probability form and generates corresponding text probability image. CAMSHIFT algorithm is used to search for the final location result in the text probability image. The experimental results demonstrate the robustness of the proposed algorithm, which is adaptive to the text of different languages, fonts or scales, and gets promising location results in variant types of images.

Keywords text location, histogram features, AdaBoost, text probability image, continuously adaptive mean shift (CAMSHIFT)

1 引 言

图像中的文字往往携带有非常重要的信息, 它能够被用来描述该图像的内容。自动地获取图像中的文字信息可以帮助人们更有效地理解图像和对图像

进行存储、压缩、检索等进一步处理。在很多具体应用中, 例如车牌定位、视频字幕提取等, 最终的目标就是文字信息的获取。基于上述的原因, 图像中文字信息的自动提取已经成为计算机视觉领域中一个重要的研究方向, 近年来引起越来越多的研究者的关注。

对图像中文字信息的提取包括对文字的定位、

基金项目: 国家自然科学基金项目(60241005)

收稿日期: 2005-03-29; 改回日期: 2005-05-24

第一作者简介: 李闯(1975 ~), 男, 1998 年获清华大学学士学位, 现为清华大学博士研究生。主要研究方向为计算机视觉, 模式识别等。E-Mail: lich@ocrserv.ee.tsinghua.edu.cn

抽取、切分和识别,其中,文字定位是整个过程的第 1 步,它的性能直接决定了文字信息提取系统所能具备的最好性能。与算法比较成熟的文字切分和文字识别相比,图像中的文字定位还没有通用有效的解决方案。现有的算法中往往使用了大量的先验知识和假设条件,缺乏足够的鲁棒性,只有在特定的应用环境中才能获得令人满意的定位准确率。

目前的文字定位算法可以被分为基于连通域和基于纹理的两类方法,其中,基于连通域的方法^[1,2]是首先利用颜色或灰度等属性从图像中提取出连通域;然后根据几何特征来对它们进行鉴别;最后将通过鉴别的连通域合并成文字定位的结果。由于基于连通域的方法易于实现,因而被广泛地应用,然而,当图像的背景比较复杂或者质量较差时,此类算法很难抽取准确的连通域;另外,此类算法中所使用的规则和门限阈值是根据特定的图像集确定的,这使得它们局限于具体的应用,缺乏鲁棒性,很难进行推广。

基于纹理的方法^[3,4]认为,图像中的文字具有特殊的纹理属性,可以采用纹理分析的方法来对文字区域进行定位。文字纹理分析的工具包括 Gabor 滤波器^[3]、小波^[4]等。基于纹理的方法通过考察像素的邻域,能够有效地克服背景噪声干扰和图像质量下降给定位带来的困难。同时,部分纹理分析的方法由于采用了机器学习的算法,从而比基于连通域的方法具有更强的鲁棒性。

目前的基于纹理的文字定位算法大多是针对以英文为代表的小字符集的字母文字,在处理中文等象形文字时很难得到同样理想的结果。为了使得定位算法能够适应多语种的变化,需要分析不同类型文字的共性以及它们与非文字类之间的差异,并依此提取鉴别能力更强的特征来进行纹理分类。

本文通过提取直方图特征来区分文字纹理和非文字纹理。不同语种、字体的文字中包含着一定数量的相同的子模式,这种特点是非文字区域所不具备的。直方图特征统计这些模式出现的频率,在对文字和非文字进行分类时具有更强的鉴别能力。

提取直方图特征之后,本文引入 AdaBoost 算法对特征进行选择并设计高性能的分类器。AdaBoost 算法采用自适应的样本训练策略,基于单特征训练弱分类器后将它们集成起来。训练得到的分类器将采用级联的结构,这样一方面能够提高运算的速度,另一方面可以同时获得较高的检测率和较低的虚警率。

级联的 AdaBoost 分类器给出的样本是文字的

概率形式的输出,该输出的概率估计首先被用来生成文字概率图,然后在文字概率图上应用 CAMSHIFT(continuously adaptive mean shift)算法来得到图像中文字区域的外接框。

与现有的基于纹理分析的文字定位算法相比,本文提出的算法由于采用了直方图特征和基于 AdaBoost 的级联分类器,因而更符合文字这种特殊纹理的特点。实验的结果证明,该算法在各种背景的图像中都能比较准确地对文字进行定位,同时能够适应文字在语种、字体、尺度等方面的变化,对比其他的算法具有更强的鲁棒性。

2 基于直方图特征和 AdaBoost 的文字定位算法

为了使得定位算法能够对不同大小的文字进行定位,应对原始的灰度图像在不同尺度上取样,首先获得一系列不同分辨率的子图像;然后采用 16×16 的检测窗口在每个子图像上进行扫描,从扫描窗口图像中提取特征,送入级联的 AdaBoost 分类器进行分类;接着根据分类器输出的后验概率估计生成对应每个子图像的文字概率图,并在文字概率图上应用 CAMSHIFT 算法得到候选文字区域的外接框;最后将不同分辨率上的候选文字区域融合起来,即可得到文字定位的结果。整个算法的流程示意图见图 1。

2.1 直方图特征

与其他目标检测问题的待检测对象不同,文字具有较大的类内散度。不同语种的文字,甚至同一语种的不同文字,其形状会有很大的差异。因此,在文字纹理分析中所采用的特征必须代表文字特有的共性,这样才能够具备较强的鉴别能力。

具备一定知识的观察者能够轻易地将文字从图像中区分出来,即便他并不认识那些字。在这种情况下,观察者做出判断的依据是图像中一些特定的模式在文字区域出现的频度远远高于非文字区域。这些模式包括水平、垂直方向的线条或者拐点等等。在实际的样本集上进行的实验能够证实这一点。实验集由大小为 16×16 的各 10 000 幅文字和非文字图像组成,图 2 给出了部分实验样本的例子。图 3(a)是实验所要考察的模式,它通常伴随着图像中的垂直线条出现;图 3(b)给出了对应的检测算子;如果使用该检测算子对实验集中的样本图像进行检测,并分别统计两类样本的输出,则可得到对应文字样本和非文字样

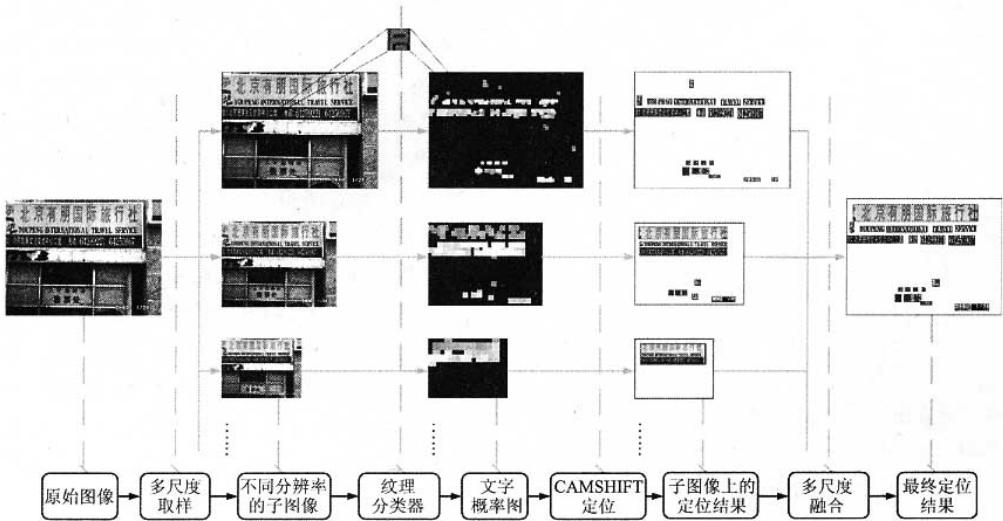
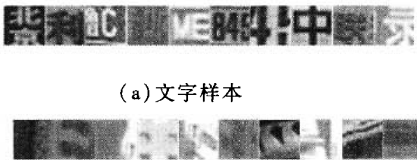


图 1 文字定位算法的流程示意图

Fig.1 Flowchart of the text location algorithm



(a) 文字样本



(b) 非文字样本

图 2 样本示例

Fig.2 Samples

集中在零点附近。这说明了在文字区域中,该模式的出现频度较高,因此可以使用代表该模式分布频度的特征来对图像中的文字进行鉴别。

本文通过直方图特征对图像中的文字区域进行检测。针对每种特定的模式,使用对应的算子在检测窗口图像中进行滤波,滤波的输出值量化后再经过统计生成该检测窗口的直方图。直方图中的统计值就是所要提取的直方图特征。这样获取的直方图特征由于统计了检测窗口内特定模式的分布频度,因此在对文字和非文字分类时具备鉴别的能力。因为直方图特征与模式出现在窗口中的位置无关,所以不同的文字虽然在形状上变化很大,但并不影响利用直方图特征将它们从非文字区域中鉴别出来。图 4 为对应图 3(a)模式的一组直方图特征中的某一维在实验样本集上的特征值分布。可以看出,该特征能够被用来有效地区分文字和非文字。

本文采用的直方图特征包括 3 组,图 5 给出了前两组所对应的模式,其中,图 5(a)~图 5(d)代表文字中常见的水平和垂直线条,图 5(e)~图 5(h)代表不同方向的拐点。同时,为了区分栅栏、门窗等具有与文字类似的上述模式分布的对象,还需要统计检测窗口中不同区域方差的分布直方图,并提取其统计值作为文字检测的第 3 组特征。

在 16×16 的检测窗口尺度范围内,选定的每种模式的长度和宽度都是可伸缩的,这样得到的直方图特征一共有 8 784 维。因为窗口内的像素个数只

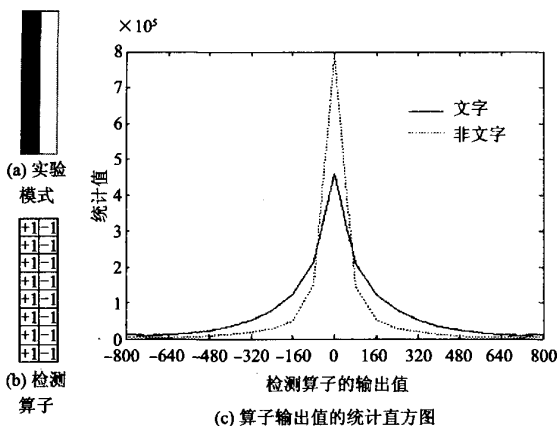


图 3 在样本集上对特定模式的统计实验

Fig.3 Statistic experiment of special mode on sample set

本的统计直方图(图 3(c))。由图 3(c)可以看出,与文字类样本对应的直方图在偏离零点处的统计值要明显大于非文字类样本,而非文字类样本的统计值则

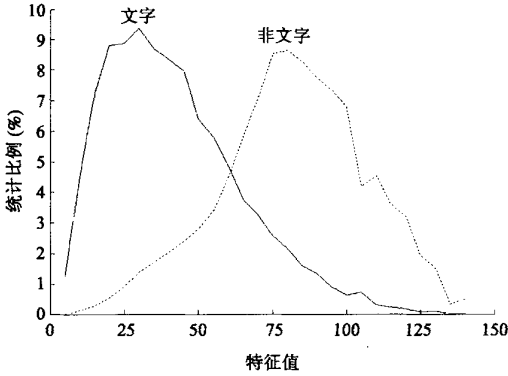


图 4 直方图特征值在实验样本集上的分布

Fig. 4 Distribution of histogram features on experiment sample set

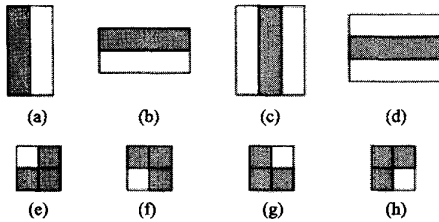


图 5 直方图特征所对应的模式

Fig. 5 Corresponding modes of histogram features

有 256, 所以本文采用的直方图特征是冗余的。

2.2 级联的 AdaBoost 分类器

本文采用 AdaBoost 算法^[5]对提取的高维直方图特征进行选择并构造区分文字和非文字的纹理分类器。

AdaBoost 算法是机器学习领域中的重要算法之一。它通过训练将若干个弱分类器集成为一个强分类器。在训练的过程中, 每个训练样本被赋予一个初始权值, 当一个弱分类器训练完成后, 根据其在训练集上的分类结果对所有的样本权值进行调整, 令下一次训练的弱分类器更关注那些被识别错误的样本。集成后的强分类器的判决结果是所有弱分类器的判决结果的加权和。研究者们已经证明, 使用 AdaBoost 算法能够得到既在训练集上具有低错误率又具备相当泛化能力的分类器。AdaBoost 算法的详细流程可以参见文献[5]。

在本文的算法中, 训练过程中的每个弱分类器都是基于单特征的。针对训练中当前样本权值的分布, 从对应每个直方图特征的弱分类器中选择错误

率最小的一个加入到强分类器中来, 而通过这样的过程在分类器训练的同时就对特征进行了选择。

对文字和非文字进行分类是一个复杂的两类问题, 很难得到一个单一的强分类器同时满足检测率和虚警率的要求。采用级联的分类器能够解决这一矛盾。分类器级联的思路在人脸检测中取得了成功^[6], 证明其在处理目标检测这类问题时是行之有效的, 因而可以利用它来解决文字定位问题。

如图 6 所示, 级联的分类器是指一组串行的分类器。用其分类时, 只有前面一级的分类器判决为正的样本才被送入后面的分类器继续处理, 反之则被认为是负样本直接输出。最后, 只有那些被每一级分类器都判决为正的样本才作为正样本输出。

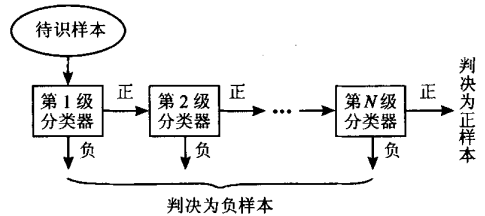


图 6 级联结构的示意图

Fig. 6 Flowchart of cascade structure

在级联结构中, 前面几级的分类器结构比较简单, 且使用的特征数也很少(在本文的算法中, 第 1 级分类器仅使用了 16 个特征), 但已足以滤除大量与正样本差异较大的负样本。后面级次的分类器使用了更多的特征和更复杂的结构, 从而可以区分那些与正样本相似的非样本。在实际检测过程中, 由于图像中的文字区域通常只占很小的比例, 这样大部分的非文字样本在级联结构前端的分类器就会被滤除, 只有少量的样本需要通过所有级的分类器, 因而级联结构大大地降低了运算复杂度。

级联结构中每一级的分类器都采用标准的 AdaBoost 算法进行训练, 其训练所使用的正样本和负样本各为 10 000 个。其中, 正样本是从不同类型图像中的文字区域截取的, 包括了中文和英文两个语种。对第 1 级分类器进行训练的负样本则是从训练集图像中的非文字区域随机获取的。在训练得到第 1 级的分类器后, 即用它对训练集图像中的所有检测窗口样本进行分类, 从中选取那些被误识为正的负样本作为下一级分类器训练的负样本。所有级次的训练过程依此类推。

对每级分类器进行训练时, 还需要对 AdaBoost

算法的判决阈值进行调整,以便令这些分类器对正样本的识别准确率很高,即检测率很高,而对虚警率不做要求。这样使得分类器组成级联结构后,仍然能够保持较高的检测率。同时,多级的结构可以保证绝大多数的虚警样本在后面的级次中被正确地识别出来,以满足系统对虚警率的要求。

级联结构的级数可通过在训练集上的测试来确定。本文中,级联结构由 22 级分类器组成,在正负样本各 10 000 个的训练集上,它在保证检测率高于 98% 的前提下,虚警样本的个数不超过 80,即对应的虚警率小于 0.8%。

2.3 文字概率图的生成和 CAMSHIFT 算法

级联结构的分类器的输出是二值判决。由于虚警区域的存在和在文字边界部分分类的可靠性下降,因此在二值输出的基础上很难得到准确的文字区域外接框,这无疑会影响进一步的文字识别。同时,二值输出也令算法缺乏足够的信息用于对定位后的候选文字区域进行鉴别。

基于上述的原因,需要对级联结构的 AdaBoost 分类器的输出进行修正,使其对检测窗口的二值判决转变为概率形式的表述。通过这样的改变,能够生成一系列与不同分辨率的子图像对应的文字概率图,然后在文字概率图上应用 CAMSHIFT 算法就可以得到文字区域的外接框。

2.2 节中已经提到,AdaBoost 算法训练得到的强分类器的判决结果是所有弱分类器的判决结果的加权和。在进行阈值调整后,判决函数如下:

$$H(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i h_i(x) - T \right) \quad (1)$$

其中, N 为弱分类器的个数, T 为调整后的判决阈值, $h_i(x)$ 和 α_i 则分别代表第 i 个弱分类器的判决结果以及对应的加权参数。

式(1)的判决函数给出的是一个二值输出。在文献[7]中,证明了在采用 AdaBoost 算法解决两类分类问题时,若类别空间 $Y = \{-1, +1\}$, 样本 x 对应的后验概率 $p(y=1|x)$ 和 $p(y=-1|x)$ 的估计为

$$p(y=1|x) = \frac{e^{\sum_{i=1}^N \alpha_i h_i(x)}}{e^{\sum_{i=1}^N \alpha_i h_i(x)} + e^{-\sum_{i=1}^N \alpha_i h_i(x)}} \quad (2)$$

$$p(y=-1|x) = \frac{e^{-\sum_{i=1}^N \alpha_i h_i(x)}}{e^{\sum_{i=1}^N \alpha_i h_i(x)} + e^{-\sum_{i=1}^N \alpha_i h_i(x)}}$$

根据式(2),在级联结构的分类器中,对于每一

个被判决为文字的样本 x , 都可以得到它的一组后验概率估计 $\{p_1, \dots, p_i, \dots, p_L\}$, p_i 为第 i 级分类器对样本 x 是正样本的后验概率估计 $p_i(y=1|x)$ 。为了从这样一组后验概率估计值中得到整体级联结构对 x 是文字样本的概率估计,可以把级数为 L 的级联结构视为 L 个分类器的组合。这样根据多分类器组合理论中的和准则^[8],可采用如下的形式来代表整体级联结构输出的后验概率估计 p_c (下角 c 代表 cascade):

$$p_c(y=1|x) = \frac{1}{L} \sum_{i=1}^L p_i \quad (3)$$

由于被判决为非文字的样本在级联结构中间的某一级就被滤除,没有通过所有的分类器,因而无法使用式(3)来表达对其后验概率的估计。对于这些样本,可选择其在级联结构中经历的最后一级,即将其判决为负的那一级分类器,将该分类器对样本的后验概率估计作为级联结构的概率形式的输出:

$$p_c(y=1|x) = p_i, i \text{ 满足 } H_i(x) = -1 \quad (4)$$

这样的选择使得在二值判决中被判决为非文字的样本与那些被判决为文字的样本相比,其对应的概率估计 p_c 也较小,保证了概率形式的输出与原有的二值判决并不矛盾。

通过式(3)和式(4)就能够得到整体级联结构分类器的概率形式的输出。尽管这样的输出并不是样本真正的后验概率,但是仍然可以有效地利用它来对文字区域进行定位和鉴别。

定义了分类器概率形式的输出之后,就可以利用它来生成文字概率图。本文使用 16×16 的检测窗口来对不同分辨率的子图像进行扫描,扫描的步长为窗口边长的 $1/2$ 。级联的分类器对检测窗口进行分类后,窗口中每个像素的像素值被分类器的概率形式的输出所替换。如果像素被多个检测窗口同时覆盖,则选择其中较大的输出值。经过这样的变换后,就可以得到一组文字概率图。与一般意义上的图像不同,文字概率图中的像素值为 $[0, 1]$ 中的实数,代表该像素是文字像素的概率估计。图 1 中的文字概率图是将其像素值映射到 $[0, 255]$ 区间后得到的,从图 1 可以看出,在相应的尺度下,原图像中的文字区域在文字概率图中具有较大的像素值。

得到文字概率图之后,就可以利用 CAMSHIFT 算法来寻找文字区域的外接框。CAMSHIFT 算法最初的目的是对视频图像中的人脸进行定位跟踪^[9],之后有研究者将其引入到文字定位中来^[10]。该算法的处理对象是 2 维的概率分布,而文字概率图恰

好满足这样的条件。应用 CAMSHIFT 算法在文字概率图中确定文字区域外接框的搜索流程如下：

(1) 设定起始的搜索窗口 W_0 的中心 (x_0^s, y_0^s) 以及大小 (w_0^s, h_0^s) (上角 s 代表搜索窗口, 即 search window), 本文算法选择文字概率图中像素值大于 0.5, 并且未被前面的搜索遍历到的像素作为起始的搜索中心；

对于第 i 次的搜索窗口 W_i ：

(2) 根据式(5)求取搜索窗口的质心位置, 并循环这一过程, 直到其收敛到新的搜索中心 (x_{i+1}^s, y_{i+1}^s) ；

(3) 根据式(7)求取新的窗口宽度和高度, 并循环这一过程, 直到其收敛到 (w_{i+1}^s, h_{i+1}^s) ，

如果 $|x_{i+1}^s - x_i^s| > \varepsilon_x$ 或 $|y_{i+1}^s - y_i^s| > \varepsilon_y$ 或 $|w_{i+1}^s - w_i^s| > \varepsilon_{width}$ 或 $|h_{i+1}^s - h_i^s| > \varepsilon_{height}$, 则将中心为 (x_{i+1}^s, y_{i+1}^s) , 大小为 (w_{i+1}^s, h_{i+1}^s) 的窗口 W_{i+1} 作为下一次的搜索窗口, 重复步骤 2 和步骤 3, 否则结束搜索过程, 进入下一步；

(4) 利用最后收敛的搜索窗口来确定外接框, 外接框的中心即窗口的中心, 其宽度和高度则根据式(6)重新确定。

下文给出搜索过程中的有关公式。对于第 i 次的搜索窗口, 首先定义其 2 维 $p+q$ 阶矩：

$$M_{p,q}(i) = \sum_x \sum_y x^p y^q I(x, y)$$

其中, $(x, y) \in W_i$, $I(x, y)$ 为文字概率图中像素点 (x, y) 的像素值。

则搜索窗口的质心位置为

$$x_{i+1}^s = \frac{M_{1,0}(i)}{M_{0,0}(i)}, y_{i+1}^s = \frac{M_{0,1}(i)}{M_{0,0}(i)} \quad (5)$$

搜索窗口的宽度和高度则为

$$w_i = 1.4 \sqrt{(a+c) + \sqrt{b^2 + (a-c)^2}} \quad (6)$$

$$h_i = 1.4 \sqrt{(a+c) - \sqrt{b^2 + (a-c)^2}}$$

其中, $a = \frac{M_{2,0}}{M_{0,0}} - x_{i+1}^{s^2}$, $b = 2 \left(\frac{M_{1,1}}{M_{0,0}} - x_{i+1}^s y_{i+1}^s \right)$,

$c = \frac{M_{0,2}}{M_{0,0}} - y_{i+1}^{s^2}$ 。

因为在搜索过程中, 搜索窗口可能会被文字区域完全覆盖, 所以其宽度和高度要有一定程度的扩大。设扩大的值为 Δw 和 Δh , 则扩大后搜索窗口的宽度和高度为

$$w_{i+1}^s = w_i + \Delta w, h_{i+1}^s = h_i + \Delta h \quad (7)$$

在采用 CAMSHIFT 算法完成对文字概率图的搜

索后, 就得到一组候选的文字区域外接框。可以利用文字概率图对这些候选区域进行鉴别。对于每个外接框, 求取其在文字概率图中的平均像素值, 即

$$\rho = \frac{M_{0,0}}{w \cdot h} \quad (8)$$

w 和 h 分别为外接框的宽度和高度。那些 ρ 值低于一定阈值的候选区域被认为是虚警区域, 而从定位结果中去除。

最后, 将不同分辨率上的定位结果融合起来, 融合的过程基于下述原则: (1) 当不同分辨率上的定位外接框基本重合时, 则选择较高分辨率上的定位结果, 因为在较高的分辨率上定位更为精确; (2) 当不同分辨率上的定位外接框彼此交叠或者覆盖, 为了尽可能的保证检测率, 则选择尺寸较大的外接框。融合后的外接框按照相应的比例映射到原始图像中, 就组成了最终的文字定位结果。

3 实验结果

级联结构的 AdaBoost 分类器是本文算法的核心, 首先通过实验对该分类器的性能进行验证。实验在级联结构的第 1 级所使用的训练集上进行, 其中正负样本各 10 000 个。本文采用检测率 R_D 和虚警率 R_F 两个指标来考察算法的性能,

$$R_D = \frac{N_D}{N_+} \times 100\%, R_F = \frac{N_F}{N_-} \times 100\%$$

其中, N_D 为分类正确的正样本数, N_F 为被判为正的负样本数, N_+ 和 N_- 则分别为训练集上的所有正样本数和负样本数。

表 1 分别给出了第 1 级分类器和整体级联结构的检测率和虚警率。在第 1 级中, AdaBoost 算法通过数量很少的几个特征训练得到的分类器, 在保证很高的检测率的前提下, 仍然能够拒绝相当数量的负样本。同时整体级联结构与单级的分类器相比, 检测率的下降不多, 而虚警率通过采用多个分类器的级联则有大幅度的改善。

由于文字定位算法缺乏统一的评价标准和测试图像库, 所以很难将本文的算法与其他的研究工作

表 1 级联结构的 AdaBoost 分类器在训练集上的实验结果
Tab. 1 Experiment result of cascade AdaBoost classifier

	第 1 级分类器	级联的分类器
检测率 R_D (%)	99.92	98.80
虚警率 R_F (%)	26.96	0.72

进行比较。但是,已有的算法大多局限于特定类型的图像或特定语种的文字,而本文提出的算法在多种类型的图像中都能比较准确地对文字进行定位,同时能够适应文字在语种、字体、尺度等方面的变化。图7

分别给出了采用本文算法对自然景物图像、车牌图像、杂志封面图像、新闻视频图像进行文字定位的结果。图8则给出了在图像中有多个语种文字或文字字体、尺度变化较大的情况下的算法定位结果。

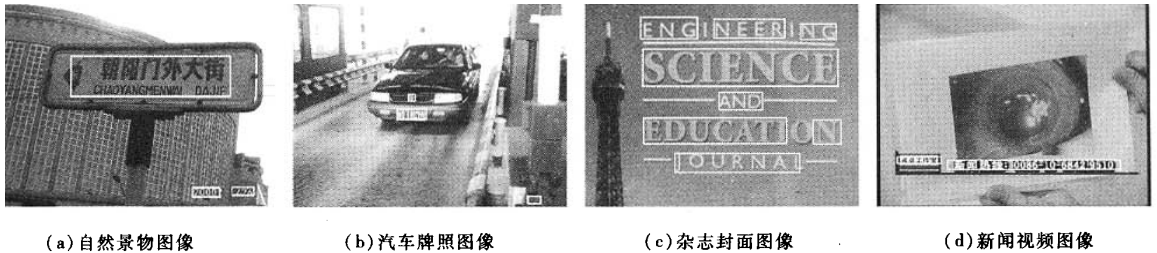


图7 本文算法对不同类型图像的文字定位结果
Fig. 7 Text location results of different types of images

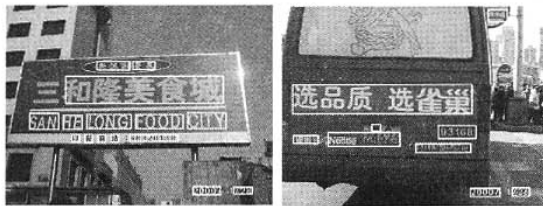


图8 图像中的文字具有不同语种、字体、尺度时
本文算法的定位结果
Fig. 8 Location results of text with different languages,
fonts or scales

4 结 论

本文提出了一种基于直方图特征和 AdaBoost 的文字定位算法。该算法通过提取具备较强鉴别能力的直方图特征来区分文字和非文字,同时引入 AdaBoost 算法来进行特征选择,并对分类器进行训练。为了获得满足要求的检测率和虚警率,并降低运算的复杂度,分类器采用的是级联结构。

新算法中,级联结构的 AdaBoost 分类器的输出不是简单的二值判决,而是概率形式的输出,然后在此基础上生成文字概率图,并在文字概率图上采用 CAMSHIFT 算法得到最终的定位结果。

实验结果表明,新算法在多种类型的图像中都能获得比较好的定位结果,同时能够适应文字在语种、字体、尺度等方面的变化。与其他算法相比,新算法具有更强的鲁棒性。

参考文献 (References)

- Jain A K, YU Bin. Automatic text location in images and video frames[J]. *Pattern Recognition*, 1998, 31(12): 2055 ~ 2076.
- LI Chuang, DING Xiao-qing, WU You-shou. Automatic text location in natural scene images [A]. In: *Proceedings of 6th International Conference on Document Analysis and Recognition [C]*, Seattle, America, 2001: 1069 ~ 1073.
- YAN Hao, ZHANG Yi, HOU Zeng-guang, *et al.* Automatic text detection in video frames based on bootstrap artificial neural network and CED [A]. In: *Journal of 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision [C]*, Plzen-Bory, Czech Republic, 2003.
- MAO Wen-ge, CHUNG Fu-lai, Lam K, *et al.* Hybrid Chinese/English text detection in images and video frames [A]. In: *Proceedings of 16th International Conference on Pattern Recognition [C]*, Quebec, Canada, 2002: 1015 ~ 1018.
- Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. *Journal of Computer and System Sciences*, 1997, 55(1): 119 ~ 139.
- Viola P, Jones M. Robust real time object detection [A]. In: *Proceedings of 2nd IEEE Workshop on Statistical and Computational Theories of Vision [C]*, Vancouver, Canada, 2001.
- Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting[J]. *Annals of Statistics*, 2000, 28(2): 337 ~ 407.
- Kitler J, Hafez M, Duin R, *et al.* On combining classifiers[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(3): 226 ~ 239.
- Bradski G. Computer vision face tracking for use in a perceptual user interface[R]. *International Technology Journal*, 1998: 2nd Quarter.
- Kim Kwang-in, Jung Kee-chul, Kim Jin-hyung. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(12): 1631 ~ 1639.