

基于笔画和 Adaboost 的两层视频文字定位算法

程豪¹ 黄磊^{1,2} 刘昌平^{1,2} 谭怒涛^{1,2}

摘要 在定位和验证的两级框架下提出了一种新的视频文字定位算法. 在定位模块中, 充分利用字符的笔画属性, 引入对字符区域有很强的响应的笔画算子; 经笔画提取, 密度过滤, 区域分解得候选文本行. 在验证模块中, 提取对文字有较强鉴别能力的边缘方向直方图特征, 使用 Adaboost 算法训练的分类器对候选文本行进行筛选. 实验结果表明, 该算法具有较强的鲁棒性, 在不同类型的视频帧中都能得到较好的定位结果.

关键词 文字定位, 笔画提取, 边缘方向直方图, Adaboost, 视频 OCR
中图分类号 TP391.4

A Two-Level Video Text Localization Algorithm Based on Strokes and Adaboost

CHENG Hao¹ HUANG Lei^{1,2} LIU Chang-Ping^{1,2} TAN Nu-Tao^{1,2}

Abstract This paper proposes a new video text localization algorithm in a localization-to-verification framework. In the localization module, to take full advantage of character stroke attribute, the algorithm introduces a stroke operator which has a strong response to text regions; subsequently, it performs strokes extraction, stroke density filtration and region decomposition to obtain candidate text boxes. In verification module, the algorithm extracts edge oriental histogram features, which have strong discriminabilities for text and non-text, then the Adaboost classifier is used to verify candidate text boxes. Experimental results demonstrate that the proposed algorithm has strong robustness and is capable of obtaining relatively good localization results in various types of video frames.

Key words Text localization, stroke extraction, edge oriental histogram, Adaboost, video OCR

视频数据的快速增长对基于内容的高效检索系统的需求越来越迫切. 视频中的文字作为一种高级语义特征, 对视频内容的理解, 索引具有重要作用. 视频文字定位作为视频中文字信息抽取系统的第一步, 直接影响着系统性能的好坏. 由于视频图像中的文本往往存在于复杂的背景中, 同时文本的颜色、亮度、尺寸和排列也不确定, 给文本检测和定位带来了极大的困难.

目前的文字定位算法主要分为基于区域和基于纹理的两类方法. 其中, 基于区域的方法^[1-2] 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. 该类方法的局限性在于图像质量较差时, 很难抽出准确的连通域, 算法中所使用的规则推广性不够.

基于纹理的方法^[3-4] 将文字视为一种特殊的纹理. 文字纹理的分析工具包括 Gabor 滤波器、小波等. 最近基于机器学习的分类算法也被应用于视频

文字定位中, 如神经网络^[5], 支持向量机^[6] 等. 基于纹理的方法比基于区域的方法更具鲁棒性, 但计算代价也更大.

本文提出了一个由定位到验证的两级视频文字定位新算法. 算法的总框架如图 1 所示, 由定位和验证两大模块构成. 定位模块充分利用文字的笔画信息, 经笔画提取, 密度过滤, 区域分解几个步骤获得候选文本行. 验证模块采用基于机器学习的方法, 提取对文字有较强鉴别能力的边缘方向直方图特征和级联的 Adaboost 分类器对候选文本行进行验证.

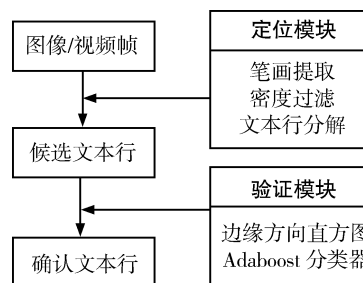


图 1 算法总框架

Fig. 1 Overview of the proposed scheme

本文其余部分组织如下: 第 1 节阐述了基于笔画的方法定位候选文本行的详细过程; 第 2 节描述了基于边缘方向直方图特征和级联的 Adaboost 分类器对文字行定位结果的筛选; 第 3 节给出了实验结果和分析; 第 4 节进行了总结.

收稿日期 2007-10-12 收修改稿日期 2008-01-11
Received October 12, 2007; in revised form January 11, 2008
国家高技术研究发展计划 (863 计划) (2007AA01Z174) 资助
Supported by National High Technology Research and Development Program of China (863 Program) (2007AA01Z174)
1. 汉王科技股份有限公司 北京 100193 2. 中国科学院自动化研究所文字识别工程中心 北京 100190
1. Hanwang Technology Co. Ltd., Beijing 100193 2. Character Recognition Engineering Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190
DOI: 10.3724/SP.J.1004.2008.01312

1 基于笔画提取的文本行初定位

无论东方文字还是西方文字, 一个字符均由笔画组成. 通常笔画宽度较小且粗细均匀, 因此笔画是表征字符的重要特征.

在定位模块中充分利用了文字的笔画属性, 首先对视频帧上利用笔画算子提取笔画, 再经笔画密度过滤获得文字候选区域, 最后进行区域分解从而得到候选的文本行.

1.1 笔画提取

视频帧中文字的笔画一般具有下述特性: 1) 为了方便阅读, 文字笔画颜色与背景颜色通常存在较大的色差; 2) 文字笔画可以视为小于某一宽度上限的双边结构; 3) 文字的笔画以 $0, \pi/4, \pi/2, 3\pi/4$ 四个方向为主.

基于文字笔画的上述特性, Ye^[7] 在处理票据文本的二值化中提出笔画算子, 本文将其应用到视频文字定位的笔画提取步骤中. 文字笔画提取过程在灰度视频帧上进行, 如图 2(a). 考虑到文字可能存在两种极性, 即正向文本(亮底暗字)和负向文本(暗底亮字), 而负向文本经过反色可转化成正向文本, 如图 2(b). 下面仅以正向文本为例说明笔画提取的主要步骤.



图 2 笔画提取过程

Fig. 2 Strokes extraction process

在正向文本中, 方向笔画强度的计算公式为

$$DE_d^+(p) = \max_{1 \leq i \leq w-1} \{ \min \{ f_d(p-i), f_d(p+W-i) \} \} - f(p) \quad (1)$$

$$DE_d(p) = \begin{cases} DE_d^+(p), & \text{if } DE_d^+(p) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

式中, $d = 0, 1, 2, 3$ 分别代表笔画最常见的 $0, \pi/4, \pi/2, 3\pi/4$ 四个方向, W 为笔画宽度上限; $f_d(p+i)$ 表示 d 方向上与点 p 距离为 i 的点像素灰度值.

笔画特征图定义为四个方向笔画特征的最大值, 如图 2(c) 所示, 用公式表示为

$$DE(p) = \max_{0 \leq d \leq 3} \{ DE_d(p) \} \quad (3)$$

对笔画特征图的二值化, 采用 Otsu 方法求得最佳阈值, 得笔画二值图如图 2(d).

$$Bina_DE(p) = \begin{cases} 1, & \text{if } DE(p) > T \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

1.2 获取文字候选区域

为了检测到图像中可能的正向文本和负向文本, 分别对原图像及其反色图做笔画提取. 图 2(d) 表示在反色图提取的负笔画图, 图 3(a) 表示在原灰度图像提取的正笔画图.

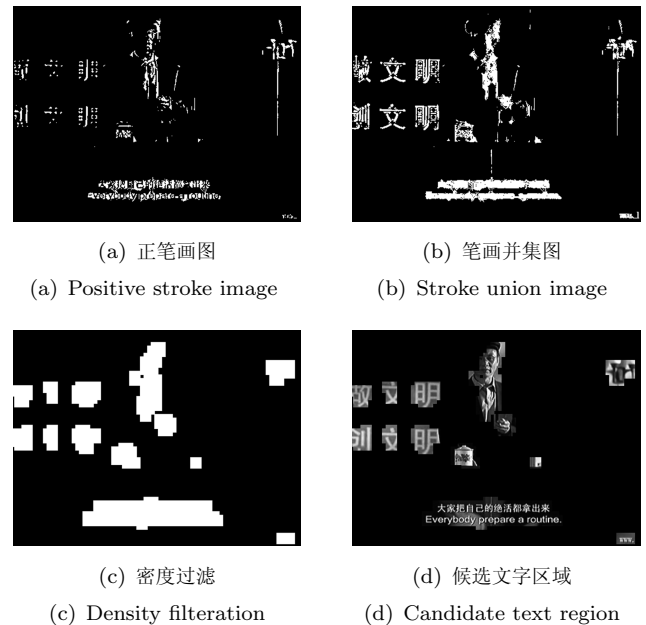


图 3 密度过滤过程

Fig. 3 Density filtration process

笔画图表示了图像中使用笔画算子检测到的宽度小于 W 的细连通体. 通常文字的笔画以及笔画的间隙富含这种特性的连通体, 正负笔画相互补充可使文字区域获得更高的笔画密度. 取正负笔画图的并集, 生成笔画并集图如图 3(b).

定义窗口的笔画密度公式为

$$Den = \frac{m}{w \times h} \quad (5)$$

其中, Den 为笔画密度, m 为窗口中笔画点的个数, W 为窗口宽度, h 为窗口高度.

显然, 文字区域一般都具有很高的笔画密度, 由此使用笔画密度分析的方法来滤除大部分非文本区域.

使用一个 16×16 , 步长为 8 的滑动窗口扫描笔画并集图, 笔画密度高于阈值的窗口置成全白, 视为候选文本区域, 其余部分全部置黑. 笔画密度的经验阈值选择为 0.4.

经过笔画密度筛选后得到的候选文字区域如图 3(c), 图 3(d) 是将该二值图作为掩码, 叠加在原始图像上的结果. 可以看到文字区域均保留, 大部分背景区域被去除. 通过笔画密度过滤获得的候选文字区域的方法快速且有效.

1.3 文本行区域分解

经过前面的步骤, 得到候选文字区域. 视频中的文字行通常按照水平和垂直方向排列, 文本行区域分解就是在候选文字区域中获得精确的単行文本定位框.

下面以水平负向文本行的提取来说明区域分解的主要过程, 这部分由候选文字区域的二值化和连通体分析两大步骤组成.

1) 候选文字区域的二值化

a) 在原始灰度视频帧上进行膨胀操作,

$$I_D^2 = I \oplus S_D \oplus S_D \quad (6)$$

式中, “ \oplus ” 表示形态学中的膨胀运算, I 为灰度视频帧, S_D 为 1×7 的结构元素, I_D^2 表示使用结构元素 S_D 对图像 I 进行两次膨胀的结果图像, 如图 4(a). 膨胀操作使得小号文字连成一片, 大号文字内部更加充实.

b) 采用自适应阈值的方法对膨胀图像作二值化, 域值 $T(p) = m(p) + k * \sigma(p)$, 这里 $m(p), \sigma(p)$ 表示点 p 邻域的平均值和方差. 二值化后的图像为 B

$$B(p) = \begin{cases} 1, & \text{if } I_D^2(p) > T(p) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

c) 在二值图上进行一次开运算

$$B_o^1 = B(p) \circ S_D \quad (8)$$

式中, “ \circ ” 表示形态学中的开运算, S_D 为 1×7 的结构元素. 开运算能消除图像中孤立点, 毛刺和小桥, 使图像中各区域变成边缘相对光滑、简单的结构.

d) 将候选文字区域的掩码叠加在 B_o^1 上得候选文字区域二值图 C , 如图 4(b) 所示.

$$C(p) = \begin{cases} B_o^1, & \text{if } mask(p) \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

2) 连通体分析

a) 在候选文字区域二值图上提取连通体, 如图 4(c) 表示各连通分量的外接矩形框;

b) 合并水平位置同属性连通体, 合并后保存其公共最小外接矩形框;

c) 由文本行的几何属性输出尺寸位置合理的文本行矩形框.

基于此框架, 作适当修改可分别得到水平正向, 竖直正/负向的文本行. 图 4(d) 是初定位的结果, 包含水平文本和竖直文本.

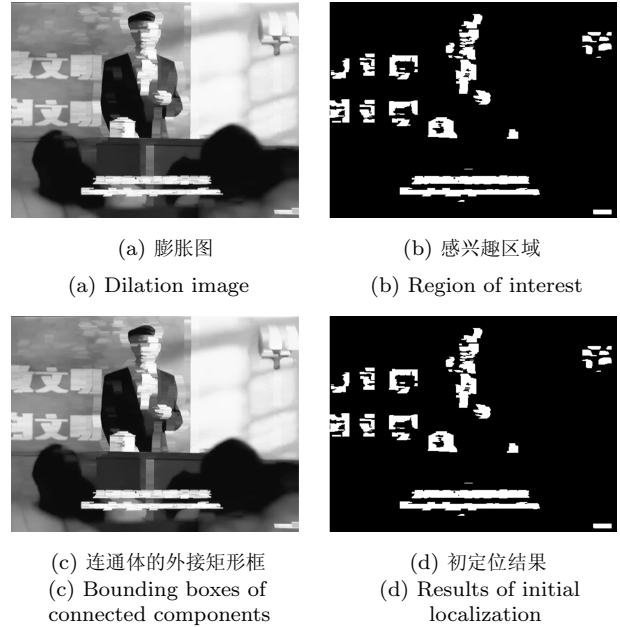


图 4 区域分解过程

Fig. 4 Region decomposition process

2 基于边缘方向直方图特征的文本行验证

在文字定位模块中为了尽量不漏检, 检测结果中还有一定数量的虚警. 在文本框的验证模块中, 使用有监督的学习方法, 通过基于边缘方向直方图特征和 Adaboost 分类器对定位结果进行筛选.

2.1 矩形的边缘方向直方图特征 (EOH)

文字含有丰富的边缘特征, 而文字通常是由一些基本的结构元素构成. 边缘方向直方图是描述这类形状特征的经典方法, 本文引入边缘方向直方图特征来提取文字的局部特征. 该方法首先求得图像边缘, 再计算任意局部区域的边缘直方图分布.

$Sobel$ 算子是简单且有效的计算图像的边缘的方法, 下面将应用 $Sobel$ 算子来提取图像的边缘. 图像 I 中点 p 处的梯度可通过 $Sobel$ 算子得到

$$\begin{aligned} G_X(p) &= Sobel_X * I(p) \\ G_Y(p) &= Sobel_Y * I(p) \end{aligned} \quad (10)$$

式中, $Sobel_X$ 和 $Sobel_Y$ 分别是 X, Y 方向的 $Sobel$ 算子, “*” 表示空间卷积.

点 p 的梯度强度计算如下

$$G(p) = \sqrt{G_X(p)^2 + G_Y(p)^2} \quad (11)$$

为了消除噪声对梯度值的影响, 引入一门限 T 来抑制微小噪声

$$G^T(p) = \begin{cases} G(p), & \text{if } G(p) > T \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

边缘的方向的计算式为

$$\theta(p) = \arctan \frac{G_Y(p)}{G_X(p)} \quad (13)$$

将边缘方向量化成 N 段, 第 k 段的值表示为

$$\Psi_k(p) = \begin{cases} G^T(p), & \text{if } \theta(p) \in bin_k \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

N 值范围在 $4 \sim 8$ 之间有较好的分类效果, 本文取 $N = 4$, 分别代表 $0, \pi/4, \pi/2, 3\pi/4$ 四个主体方向.

经过量化和分类后, 将第 k 方向的点集中在图像 Ψ_k 中, 由此可生成 N 方向分解子图.

$E_k(R)$ 表示为矩形 R 的 k 方向特征

$$E_k(R) = \sum_{(x,y) \in R} \Psi_k(p) \quad (15)$$

归一化为

$$E'_k(R) = \frac{E_k(R)}{\sum_{k=1}^N E_k(p)} \quad (16)$$

对特定矩形 R , 由此得到的特征为 4 维.

2.2 特征提取

Adaboost 是机器学习领域的重要算法之一, 分类器的训练需要大量的正负样本. 正样本是经人工标注的单字符字块统一缩放到 16×16 的图像, 负样本是相同大小的非字符图像. 训练过程在这些样本图像块上提取特征, 进行有监督的学习. 因此本算法的特征提取均在 16×16 的图像块上进行.

任一矩形可以用四元组来表示: $R = (x, y, w, h)$, 其中 x 和 y 是该矩形左上角点的坐标, w 和 h 是该矩形的宽和高.

在 16×16 的图像块内, 考虑到当矩形过小, 统计直方图没有意义. 任意位置且面积不小于 36 像素矩形表示为 $\{R | 1 \leq w, h \leq 16; w \times h \geq 36;$

$0 \leq x \leq 16 - w; 0 \leq y \leq 16 - h\}$, 计算 4 维方向特征. 这样计算得到的特征维数为 $4749 \times 4 = 18996$. 为了加速运算, 引入积分图的计算方法:

$$SAT_k(x, y) = \sum_{0 \leq x' \leq x, 0 \leq y' \leq y} \Psi_k(x', y') \quad (17)$$

$$E_k(R) = SAR_k(x-1, y-1) + SAT_k(x+w-1, y+h-1) - SAT_k(x-1, y+h-1) - SAT_k(x+w-1, y-1) \quad (18)$$

利用积分图的计算方法可以快速地由 Ψ_k 计算出 $E_k(R)$, 计算复杂度减小至 $O(n^2)$.

2.3 特征选择

上节计算的方向直方图特征集个数远远超过了像素点的个数, 具有很大的冗余. 本文采用 Adaboost 算法^[8] 选择少量区分性好的特征组成有效的分类器. Adaboost 算法通过训练将若干个弱分类器集成为一个强分类器. 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, 从对应每个矩形边缘方向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, 这样在挑选特征的同时实现了分类器的训练.

训练使用 30000 个正样本, 60000 个负样本, 其中正样本取自 2000 个视频帧中的单个字符块, 样本字符含中文 (简体、繁体)、英文、韩文和数字. 设定训练分类器的中止条件为: 正样本最小通过率为 0.996, 最大的虚警率 0.5. 即 $P(x|o) \geq 0.996$, $P(x|\bar{o}) \leq 0.5$, 其中 o 表示正样本, \bar{o} 表示负样本; x 表示分类器判别为正. 初次训练, 选择 3 个特征构成分类器即可达到上述指标, 即仅计算 3 个特征就可以排除约 50% 的负样本, 但虚警率为 50% 的分类器并不能满足实际需求.

2.4 级联的 Adaboost 分类器

级联分类器是指一组串联的分类器, 如图 5 (见下页) 所示. 若每一级的性能指标均设为 $P(x_i|o) \leq 0.996$, $P(x_i|\bar{o}) \leq 0.5$, 则 N 级级联分类器最终的正样本通过率 $HR \geq (0.996)^N$, 最终的虚警率降为 $FA \leq (0.5)^N$. 当 $N = 12$ 时, 可保证检测率高于 0.95 的前提下, 虚警率低于 2.44×10^{-4} .

为了让下一级分类器更加关注那些被误识的样本, 在训练得到第一级分类器后, 用它对训练集所有样本进行检测, 从中选择那些被误识为正的负样本作为下一级训练样本的输入. 每一级分类器采用标准的 Adaboost 算法进行训练. 训练所得 12 级分类器, 每级对应的特征数如图 6 所示.

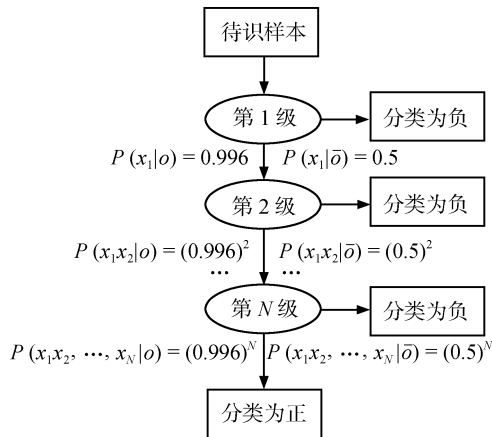


图 5 级联结构示意图

Fig. 5 Flowchart of cascade structure

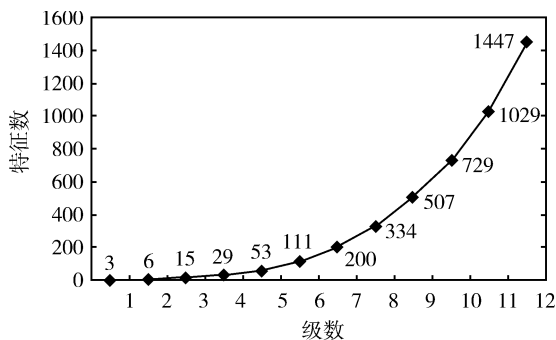


图 6 各级分类器特征数示意图

Fig. 6 Number of features in each stage

在级联结构中, 分类器的特征数随级数的增加而显著增加. 前面几级的分类器使用特征数少, 但可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, 从而可以更好地区分那些与正样本近似的负样本. 在实际检测中, 大部分非文字会在前端的分类器滤除, 因而级联机构降低了运算的平均复杂性.

2.5 分类器识别结果对文本候选区域的验证

级联结构的分类器的输出是对每个检测窗口的二值判决, 由此作为文本行验证的依据. 算法步骤如下:

- 1) 扩大: 定位得到的文本框多为文本行紧密的包围框. 为了保证文字完全在框内, 将水平 (竖直) 文本框四周均扩大 0.5 倍的框高 (框宽);
- 2) 缩放: 将扩大的水平 (竖直) 文本框固定宽高比的缩放到高度 (宽度) 32 像素, 其中字高约为 16 像素;
- 3) 检测: 用 16×16 , 步长为 2 的检测窗口在经过上述处理的文本框上扫描. 对扫描经过的窗口用已训练好的分类器判决该窗口是否属于文字, 判别为正的窗口用方框标注, 本例标注结果如图 7 所示;

4) 判断: 在未扩大区域内, 文本框中标注面积与区域面积之比大于阈值 T , 判别该文本框有效, 否则作为虚警去除. 在本例中 T 取 0.4, 由此图 7 (a)~(e) 确认为文本; 图 7 (f) 被去除.

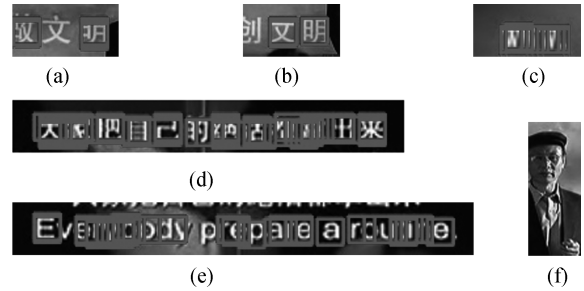


图 7 检测结果标注图

Fig. 7 Detection results

在经过分类器的筛选后, 错误的候选被去除, 验证过程在降低误检率方面是十分有效的. 本例的最终定位结果如图 8 所示.



图 8 定位结果

Fig. 8 Localization results

3 实验结果及分析

3.1 算法准确性的评估

首先参照文献 [6] 中的文本定位性能评估方法, 定义如下的评判准则

$$\delta = \begin{cases} 1, & \text{if } \frac{o}{g} > 0.95 \text{ and } \frac{o}{r} > 0.75 \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

其中 r, g 和 o 分别表示检测到的文本区域, 真实的文本区域和它们的重叠区域. 即当重叠区域大于真实文本区域 95% 并且大于检测区域 75% 时, 认为文本区域被正确检测. 真实文本由人工标注产生.

对定位算法的准确性的测试, 选择数据集为两组静态图片: a 组为 Adaboost 训练中所使用的 2000 幅视频帧, b 组为未参与训练的 4000 幅视频帧. 视频帧内容包括新闻、访谈、电影、娱乐、广告等; 分辨率从 320×240 到 1024×768 不等; 其文字大小从 $1/25$ 图像高至 $1/3$ 图像高; 文字包含中 (简繁)、英、韩等多语种. 人工标注数据集 a 中文本行

个数为 5 195, 数据集 b 为 10 368. 在此两个数据集上测得算法的召回率和精度如表 1 所示

表 1 算法性能

Table 1 Performance of our algorithm

检测方法	召回率	精度
初始定位	95.3% (a)	70.4% (a)
	95.4% (b)	95.4% (b)
定位 + 验证	95.0% (a)	92.5% (a)
	94.8% (b)	91.3% (b)

$$\text{召回率 (Recall)} = \frac{\text{返回结果中正确结果的数目}}{\text{测试集中正确结果的数目}}$$

$$\text{精度 (Precision)} = \frac{\text{返回结果中正确结果的数目}}{\text{返回结果的数目}}$$

测试结果在训练集和测试集上性能相差不大, 因为初始定位与训练集数据无关, 验证算法由于选择大样本训练对特定训练集的依赖亦不强. 视频图像中约 95% 的文字被成功检测, 只有约 5% 质量很差的文字检测失败. 这说明初始定位中使用的笔画特征对图像中的各式字符均有很好的响应, 同时又能去除绝大部分背景.

验证阶段在保证较高召回率的同时能够将检测精度提高到 90% 以上. 对结果中仍存在的约 10% 的虚警部分可在 Video OCR 系统后续的跟踪和分割阶段被去除. 这充分说明了由机器学习选择出的边缘直方图特征具有很强的文字和非文字区分能力. Adaboost 级联分类器具有很低的错误率. 验证模块使得定位的精度得到了很大的提高.

3.2 算法速度的评估

训练分类器的过程在 AMD 64 位双核 2.01 G CPU 和 2 G 内存的台式机进行, 在大样本集上 (正样本 30 000, 负样本 60 000) 使用 Adaboost 方法训练 12 级分类器耗时 47 天. 由于离线训练耗时并不影响检测算法的执行效率, 所以没有对该部分作过多优化.

初始定位的执行时间和笔画参数选择与视频的分辨率密切相关. 为了解决尺度问题, 在检测前先将较大的视频帧按比例缩小到高度为 360 像素, 实验

测得笔画宽度 W 设为 6 像素时, 对高度小于 1/4 图像高的文字均有较好响应. 初始定位在 Pentium IV 2.8 G CPU, 512 M 内存的计算机上处理一幅图像的时间为 60 ms.

验证过程处理的图像区域较少, 对于包含 3 个文本行的视频帧处理时间为 29 ms.

定位算法在视频流上检测, 平均速度为每帧 90 ms, 即 1 秒约可处理图像 11 帧. 文字通常会在视频中停留 2 秒以上, 在 Video OCR 系统中一般对视频进行跳帧定位, 间隔小于 1 秒就不会带来漏检. 因此算法的运算速度完全满足实时系统的需要. 采用由定位到验证的两层结构, 既获得了较高的性能指标, 又降低了运算的复杂度.

3.3 算法间的比较与定位结果示例

由于视频文字定位算法缺乏公开的数据集, 表 2 列出了一些最新算法分别在自建数据集上自测的性能参数.

不难看出, 本文算法选择更大的数据集上, 测得的准确性与其他最新算法相当; 同时本文算法在速度上的优势明显. 图 9 显示了部分图片的定位结果.



图 9 更多定位结果

Fig. 9 More localization results

表 2 算法性能比较

Table 2 Performance comparison of different algorithms

算法	召回率	精度	图像尺寸/ 计算机/耗时	测试集
本文算法	94.8%	91.3%	480 × 360/ P4 2.8 G 512 M/ 0.09 s	4 000 视频帧含 10 368 个文本框
文献 [5] 中算法	92.8%	91%	352 × 240/ Sun Ultra 1 Workstation/ 1 s	75 帧含 153 个文本框
文献 [9] 中算法	94.5%	90.3%	384 × 288/ P4 2.4 G 512 M/ 0.83 s	220 视频帧含 788 个文本框

4 总结

本文针对视频图像中的文字提出了一种新颖的文字定位算法. 由粗到精的框架降低了算法的时间复杂性. 算法综合利用了文字的笔画特征和边缘分布的局部特征, 结合 Adaboost 分类器, 可以快速有效地从视频帧中定位出文本. 通过与人工标注结果的比对, 文本的算法具有很好的性能.

References

- 1 Lyu M R, Song J Q, Cai M. A comprehensive method for multilingual video text detection, localization, and extraction. *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, **15**(2): 243–255
- 2 Li C, Ding X Q, Wu Y S. Automatic text location in natural scene images. In: *Proceedings of the 6th International Conference on Document Analysis and Recognition*. Seattle, USA: IEEE, 2001. 1069–1073
- 3 Yan H, Zhang Y, Hou Z G. Automatic text detection in video frames based on bootstrap artificial neural network and CED. In: *Proceedings of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. Plzen-Bory, Czech Republic: Science Press, 2003. 1–6
- 4 Mao W G, Chung F L, Lam K K M, Siu W C. Hybrid Chinese/English text detection in images and video frames. In: *Proceedings of the 16th International Conference on Pattern Recognition*. Quebec, Canada: IEEE, 2002. 1015–1018
- 5 Li H P, Doermann D, Kia O. Automatic text detection and tracking in digital video. *IEEE Transactions on Image Processing*, 2000, **9**(1): 147–156
- 6 Ye Q X, Huang Q M, Gao W, Zhao D B. Fast and robust text detection in images and video frames. *Image and Vision Computing*, 2005, **23**(6): 565–576
- 7 Ye X Y, Cheriet M, Suen C Y. Stroke-model-based character extraction from gray-level document images. *IEEE Transactions on Image Processing*, 2001, **10**(8): 1152–1161
- 8 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Hawaii, USA: IEEE, 2001. 511–518
- 9 Wang J, Zhou Y H. An unsupervised approach for video text localization. *IEICE Transactions on Information and Systems*, 2006, **89**(4): 1582–1585



程 豪 首都师范大学硕士研究生. 主要研究方向为图像处理、文字识别. 本文通信作者.

E-mail: chenghao@hanwang.com.cn

(**CHENG Hao** Master student at Capital Normal University. Her research interest covers image processing and character recognition. Corresponding author of this paper.)



黄 磊 中国科学院自动化研究所副研究员. 主要研究方向为文字识别, 人脸识别, 智能监控.

E-mail: lei.huang@mail.ia.ac.cn

(**HUANG Lei** Associate professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers character recognition, face

recognition, and intelligent surveillance for security.)



刘昌平 中国科学院自动化研究所研究员. 主要研究方向为文字识别, 人脸识别, 智能监控.

E-mail: changping.liu@mail.ia.ac.cn

(**LIU Chang-Ping** Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers character recognition, face

recognition, and intelligent surveillance for security.)



谭怒涛 中国科学院自动化研究所博士研究生. 研究方向为人脸识别、文字识别. E-mail: nutao.tan@ia.ac.cn

(**TAN Nu-Tao** Ph.D. candidate at the Institute of Automation, Chinese Academy of Sciences. His research interest covers face recognition and character recognition.)