

基于 Adaboost 的视频文本定位

尹芳^{1,2}, 郑亮¹, 陈田田¹

(1. 哈尔滨理工大学 计算机科学与技术学院, 黑龙江 哈尔滨 150080;
2. 哈尔滨理工大学 仪器科学与技术博士后科研流动站, 黑龙江 哈尔滨, 150080)

摘要:提出了一种基于 Adaboost 的视频文本定位的新方法。首先我们提取视频图像中的连通域, 经过对视频文本区域分析, 提取了视频文本的 5 类特征, 然后利用这 5 类特征经过分类与回归决策树构造了 Adaboost 强分类器, 最后将候选文本区域送入强分类器, 得到正确的文本区域。实验结果表明本方法不仅对视频帧图像中字体、大小和颜色多变的文本具有很好的定位效果而且还实现了视频文本定位要求的快速性和准确性的特点。

关键词:文本定位; 文本识别; 连通域; 强分类器; 分类与回归决策树

DOI: 10.15938/j.jhust.2017.01.018

中图分类号: TP391 **文献标志码:** A **文章编号:** 1007-2683(2017)01-0103-06

Text in Video Location Based on Adaboost

YIN Fang^{1,2}, ZHENG Liang¹, CHEN Tian-tian¹

(1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China;
2. Instrument Science and Technology Postdoctoral Research Station, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: This paper proposes a new method of video text location based on Adaboost. First extracting the connected domains from the video image, we get five classes features of text after analyzing the text areas. Then we construct a strong classifier of Adaboost with CART (Classification And Regression Tree) using the five classes features. Finally, we send the candidate text regions into the strong classifier to get the correct text areas. The experimental results show that not only can this method achieve a good effect on the text location in the video images including the text of various fonts, sizes and colors but also realize rapidity and precision that the video text location requires.

Keywords: text location; text recognition; connected domain; strong classifier; decision or classification tree

0 引言

随着多媒体网络技术的发展,大量的视频进入了人们的学习、生活和工作当中。然而视频中的文

本信息作为一种重要的语义信息,对视频的理解、分析和检索有着重要的作用。视频文本识别主要包括两步。第一步,对视频帧中的文本进行定位;第二步,获取文本定位区域中的文本信息。然后将获取的文本信息送入到 OCR 或利用其他的文本识别技

收稿日期: 2015-09-14

基金项目: 黑龙江省教育厅科学技术研究项目(12541119)。

作者简介: 尹芳(1978—),女,博士,副教授;

郑亮(1989—),男,硕士研究生;

陈田田(1992—),女,硕士研究生, E-mail: 13936421412@163.com.

术进行文本识别。

最近几年,不少视频文本提取的相关算法应运而生。Li等^[1]通过利用混合小波神经网络对 16×16 像素块的视频帧进行检测;Zhong^[2]和Crandall^[3]利用文本图像在垂直和水平方向亮度的变化规律在 8×8 的DCT压缩域直接实现了文本检测。但是基于块的方法在文本定位不够精确。数学形态学,角点检测和边缘检测也是大多学者经常采用的方法。工作人员在制作视频时,对于具有文本的区域,往往使其前景与背景具有鲜明的对比度,在频域上这种对比度表现为高频特征,而在空域上可以通过角点和边缘点来表现,因此角点和边缘检测经常是文本定位采用的方法。例如,Hua^[4]等人先利用SUSAN角点检测算子获得视频帧的角点图,然后对角点按照一定的规律进行合并和过滤最终定位文本区域。这种方法应用起来比较简单,速度较快,但是对噪声比较敏感,只能处理背景简单的视频文本。

近几年基于Adaboost的学习方法也被大量应用于文本识别,并且取得了良好的效果。例如,Chen和Yuille^[5]利用文本图像的X和Y偏导数、灰度直方图和边缘连接特征进行Adaboost学习实现文本定位。

以上方法虽然可以实现文本定位,但是效果并不理想。由于视频文本本身的低分辨率,复杂的背景,文本的颜色,字体以及大小多变性等因素使视频文本定位存在不小的难度。

1 Adaboost 算法

AdaBoost算法通过将大量分类能力一般的弱分类器按照一定方法组合起来,构成一个分类能力很强的强分类器。Keams和Valiant已经证明,只要有足够的分类能力比随机猜测好的分类器,就能将组合成误分率趋于0的强分类器。Adaboost算法根据每次训练集中各样本的分类是否正确以及上一次的总体分类准确率来确定每个样本的权值,将修改权值的新数据送给下层分类器进行训练,然后将每次训练得到的分类器融合起来,作为最后的决策分类器,实现对目标的搜索检测。

假设输入 n 个训练样本: $\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_n, y_n\}$,其中 $y_i \in \{0, 1\}, i = 1, 2, \dots, n$; 0表示假样本,1表示真样本。已知训练样本中有 a 个假样本和 b 个真样本,则第 j 个生成的简单分类器如式(1)所示。

$$h_j = \begin{cases} 1 & p_j f_j < p_j \theta_j \\ 0 & \text{others} \end{cases} \quad (1)$$

其中: h_j 表示简单分类器的值; θ_j 表示阈值; p_j 表示不等号的方向,只能取 ± 1 ; f_j 表示特征值, $j = 1, 2, \dots, n$ 。

步骤1:初始化误差权值,对于 $y_i = 0$ 的假样本, $W_{1,i} = 1/2a$;对于 $y_i = 1$ 的真样本, $W_{1,i} = 1/2b$;

步骤2: For $t = 1, 2, \dots, T$, 其中 T 为训练的次数;

1) 权重归一化, $\frac{w_{t,i}}{\sum_{j=1}^m w_{t,j}} \rightarrow w_{t,i}$;

2) 对每个特征 j ,循环生成相应的弱分类器 h_j ,计算相对于当前权重的误差:

$$\varepsilon_j = \sum_i w_{t,i} |h_j(x_i) - y_i| \quad (2)$$

3) 选择具有最小误差 ε_j 的弱分类器 h_t 加入到强分类器中去;

4) 更新每个样本所对应的权重;

$$w_{t+1,i} = w_{t,i} \beta_i^{1-e_i} \quad (3)$$

其中,如果第 i 个样本 x_i 被正确分类,则 $e_i = 0$,反之 $e_i = 1$; 其中 $\beta_i = \frac{\varepsilon_i}{1 - \varepsilon_i}$ 。

步骤3:形成强分类器为

$$h_j(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t < \frac{1}{2} \alpha_t \\ 0 & \text{others} \end{cases} \quad (4)$$

式中 $\alpha_t = \lg \frac{1}{\beta_t}$ 。

Adaboost算法相比于其他机器学习算法所具备的优点:第一,构造一个弱分类器比直接构造一个高精度的强分类器简单的多;第二,不会过度拟合,不会随着迭代次数的增加而降低其泛化能力。

2 候选区域检测

候选区域生成的主要任务是提取图像中的连通区域。其中包括文本区域和非文本区域。然后将这些连通域送入利用Adaboost生成的强分类器。从而将连通区域中的非文本区域剔除,最终得到准确的文本区域。

2.1 灰度化处理

文[6]进行边缘检测,首先将彩色的视频帧图像灰度化,以减少后续图像处理计算量。

灰度化处理的方法有,最大值法、平均值法和加权平均值法,最大值法产生的灰度图像亮度过高,平

均值法产生的灰度图像过于柔和,所以本文采用加权平均值法将视频帧图像灰度化,首先如式(5)给每个像素的 R 、 G 、 B 分量赋予权值,然后取得3个分量的加权平均值

$$R = G = B = (W_R \times R + W_G \times G + W_B \times B) / 3 \quad (5)$$

其中 W_R 、 W_G 、 W_B 分别是 R 、 G 、 B 的权值,根据先验知识,其中 $W_R = 0.30$ 、 $W_G = 0.59$ 、 $W_B = 0.11$ 得到的灰度图像较为合理,即:

$$\begin{aligned} V_{gray} &= R = G = B \\ &= (0.30R + 0.59G + 0.11B) / 3 \end{aligned} \quad (6)$$

实验结果比较如图1所示。



图1 灰度化处理所得图像

2.2 边缘检测

为了能给读者提供有关视频的信息及帮助理解,视频文本大多是叠加的,而且一般会与背景有较大的对比度。因此本文根据梯度特征来获得文本区域边缘。

依赖于图像梯度的边缘检测方法根据 X 、 Y 方向滤波器计算梯度估计的不同分为不同类型,例如 Roberts 算子、Sobel 算子和 Prewitt 算子等。本文通过实验比较,选择 Sobel 算子^[4]。

Sobel 算子是一种离散性差分算子,与 Roberts 算子相比,它能够在噪声情况下精确定位图像边缘方向的差分。它利用水平卷积因子和垂直卷积因子(8),在横向 x 和纵向 y 方向上对图像 $G(x, y)$ 做平面卷积,得到亮度差分:

$$\begin{aligned} G(x) &= [f(x+1, y-1) + 2f(x+1, y) + \\ &\quad f(x+1, y+1)] - [f(x-1, y-1) + \\ &\quad 2f(x-1, y) + f(x-1, y+1)] \end{aligned} \quad (7)$$

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \text{ 和 } \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad (8)$$

通过式(9)求得该点的灰度值:

$$g(x, y) = |G(x) + G(y)| \quad (9)$$

设定阈值 T ,提取大于阈值 T 的点作为边缘点。

实验结果如图2所示。

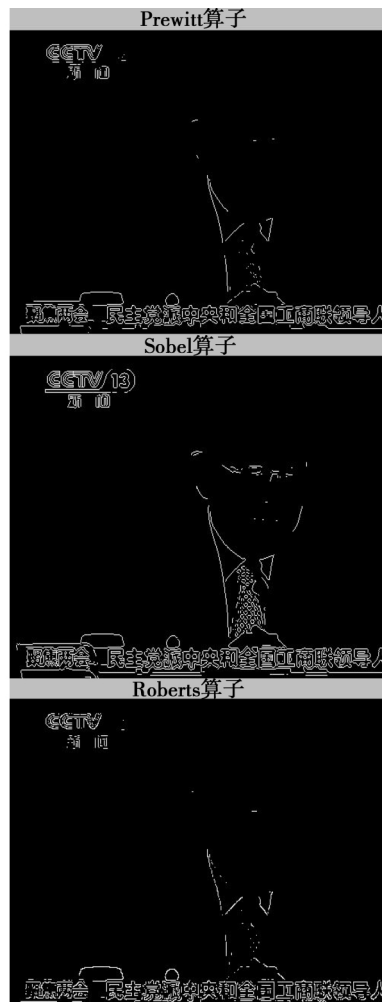


图2 边缘检测比较图

3 基于 Adaboost 的视频文本定位

3.1 总体系统设计

整个系统分为预处理,特征提取和基于 Ada-boost 强分类器3部分,如图3所示。

其中 CART(classification and regression trees)即分类与回归决策树,是运用二元递归分割的过程来分析大数据集。因具有具有分类速度快、精度高、较其他模型更容易理解等优点,被选为本系统 Ada-boost 的基本弱分类器(即 Adaboost with CART)^[4]。

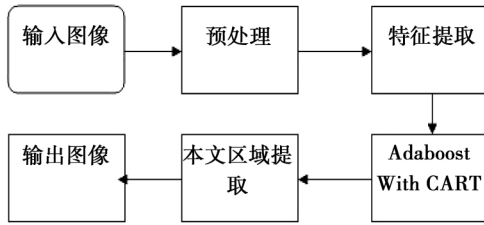


图3 系统总体设计

3.2 预处理

现有视频类型及播放工具多种多样,因此所采集的视频图像也是大小不一。本系统为了将大小不同的数字图像缩放到统一标准,首先进行边界扫描,确定图像的大小,然后利用差值算法,将图像缩放为 64×32 ,为分类做好准备。图像尺寸选择主要考虑图像的变化关系为 2^n , n 为正整数;同时由于文字形状一般为长方形结构,因此在保证形状不失真的情况下尽可能缩小尺寸,以减少计算量,提高处理速度。

3.3 特征提取

特征提取是视频定位系统的重要部分,本文为构建分类与回归决策树的 Adaboost 强分类器提供了 5 类特征。

3.3.1 Gabor 特征

视频文本的笔画特征丰富且表现出一定的周期性,本文将视频文本看作一种特殊的纹理,利用 Gabor 滤波器提取文本特征^[8]。本文利用二位 Gabor 滤波器进行分析,其函数如下定义:

$$F(x, y, l, \theta_k) = F_1(x, y) \left[\cos(R) - \exp\left(-\frac{\delta^2}{2}\right) \right] + iF_1(x, y) \sin(R) F_1(x, y) = \frac{\gamma^2}{\delta^2} \exp\left(-\frac{\gamma^2(x^2 + y^2)}{2\delta^2}\right) \quad (10)$$

其中: $R = \gamma(x \cos \theta_k + y \sin \theta_k)$, $\theta_k = \frac{\pi k}{D}$, $\gamma = \frac{2\pi}{l}$, $k = 0, 1, \dots, D-1$; l 代表波长; θ_k 代表方向。

3.3.2 连通性分析

1) 文本尺寸特征。文本区域在整幅图像中通常占有一定的比例,文本区域的大小通常通过文本区的宽高信息来表示。设 $S(cc)$ 表示连通区面积, $S(pic)$ 表示整个图像区域的面积, $w(cc)$ 和 $h(cc)$ 分别表示区域的宽高, $MinWH$ 表示最小长宽, S_Ratio 表示面积比,其中:

$$MinWH = \min\{w(cc), h(cc)\} \quad (11)$$

$$S_Ratio = \frac{S(cc)}{S(pic)} \quad (12)$$

2) 文本区域的边缘密度特征。笔画是构成文本的单位元素,因此文本区域具有很强的边缘密度 ($Edge_Density$)。假设 $P(cc)$ 为边缘像素点的数量,则:

$$Edge_Density = \frac{|P(cc)|}{S(cc)} \quad (13)$$

结合先验知识,实验时阈值定义为 $MinWH > 8$, $0.05 < S_Ratio < 0.5$, $Edge_Density > 0.2$ 。

3.3.3 笔画密度

文本笔画以 $\pi/4, \pi/2, 3\pi/4, \pi$ 4 个方向为主,且前景与背景间存在较大差异^[9]。根据此特性,在亮底暗字的正向文中,利用式(14)计算笔画强度^[10],而对于暗底亮字的负向文本将其进行反色,统一为正向文本进行处理。

$$DE_p^+(q) = \max\{\min_{(1 \leq w \leq 1)} \{f_p(q + wj)\}, f_p(q - 1)\} - f(q) \quad (14)$$

$$DE_p(q) = \begin{cases} DE_p^+(q) & DE_p^+(q) > 0 \\ 0 & \text{others} \end{cases} \quad (15)$$

式中: $p = 0, 1, 2, 3$ 表示 $\pi/4, \pi/2, 3\pi/4, \pi$ 这 4 个方向; W 表示笔画宽度上限; $f_p(q + 1)$ 表示 p 方向上与点 q 距离为 j 的像素点的灰度值。

笔画特征表示 4 个方向上笔画特征的最大值,用式(16)表示:

$$DE_p(q) = \max_{(0 \leq p \leq 3)} \{DE_p(q)\} \quad (16)$$

对笔画特征图做二值化处理,采用 Otsu 求得阈值,即:

$$Bian_DE_p(q) = \begin{cases} 1 & DE_p^+(q) > 0 \\ 0 & \text{others} \end{cases} \quad (17)$$

这种方法对笔画均匀的文字区域具有较好效果。

3.3.4 纹理统计特征

纹理统计是文本定位的常用方法^[11],本文使用 4 种纹理特征来区分文本区域和非文本区域。设 $p(i, d, j, \theta)$ 为特征图像中点 (i, j) 关于距离 d 和方向 θ 分离的联合概率,则能量、熵、对比度和逆差距 4 种纹理统计特征定义如下:

$$\text{能量: } E(d, \theta) = E_{i,j} p^2(d, \theta)$$

$$\text{熵: } H(d, \theta) = E_{i,j} p^2(d, \theta) \log p(d, \theta)$$

$$\text{对比度: } I(d, \theta) = E_{i,j} (i - j)^2 p(d, \theta)$$

$$\text{逆差距: } L(d, \theta) = E_{i,j} \frac{1}{1 + (i - j)^2} p(d, \theta)$$

3.3.5 图像的 X, Y 导数的方差和期望

Chen 和 Yuille^[8]使用 $X - Y$ 导数的平均值和标准差提取文本特征。对于文本图像, X 方向的偏导

数具有极小值,而 Y 方向的偏导数具有极大值,根据 X 偏导数的极大值和极小值来进行区域分割,然后计算被分割区域的方差和期望作为特征。

本文利用这 5 类特征生成弱分类器,通过 Ada-boost 进行结合,生成一个精度较高的强分类器。

3.4 基于 Adaboost 的强分类器构建和实验结果分析

3.4.1 数据库的构建

目前还没有统一标准的视频文本数据库,本文在自建视频文本数据库上进行实验。在网络上共搜集了 300 幅图像,其中包括新闻、电影和广告等视频文本图像,图像中的文本大多数为中文,还有少量的英文和数字,并且这些文本在尺寸,颜色,对比度以及字体方面都有所不同。从 300 幅图像中选出 200 幅图像作为训练样本,其余 100 幅图像作为测试样本测试本文提出的文本定位方法的效果及泛化能力。训练中的正样本是通过手动标记图像中的文本区域获得的,负样本是通过程序对背景图像随机抽取得到的。实验过程中共采集 1 860 个正样本和 8 000 个负样本作为 Adaboost 分类器的训练样本,正、负样本的部分图像如图 4。



(a)部分正样本



(b)部分负样本

图4 部分训练样本图像

3.4.2 基于 Adaboost 的强分类器构建

为了给正负样本分配恰当的权值,我们使用了非对称 Adaboost 的方法^[12],通过改进经典 Adaboost 算法,定义弱分类器为

$$h_j' = h_j(x) - \frac{m}{M} \lg 2 \quad (18)$$

其中 $m = 1 \cdots M$, M 为迭代的次数。设置迭代次数 M 为 100。

从 1 860 个正样本和 8 000 个负样中提取 5 类特征集,并利用这些特征集构建了一个最大深度为 4 的分类与回归决策树。树的每个节点都是 Ada-boost 弱分类器,而节点后继分支表示文本区域和非文本区域属性的可能值。最终结果是由最终的 Ad-

aboost 强分类器输出。

利用正确率 p (提取出的正确文本区域/提取出的区域)、召回率 r (提取出的正确文本区域/总文本区域) 及其两者的调和平均数 $f(正确率 \times 召回率 \times 2 / (正确率 + 召回率))$ 作为评价的性能指标。利用 100 幅检测样本图像,对 5 类特征集的分类性能作了比较,如图 5 所示。

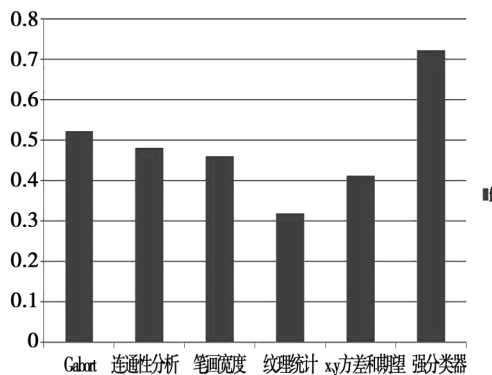


图5 各特征分类效果

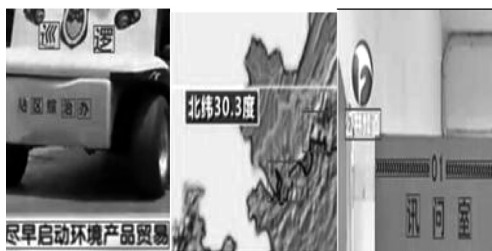
尽管每个分类器的分类性能并不佳,但是经过分类与回归树的 Adaboost 算法合并后,表现出了强大的分类性能。

3.4.3 实验结果分析

为了体现方法的优越性,采用同样的训练样本训练了 SVM(支持向量机)^[13],并利用同组测试样本进行两者分类性能的比较。两者对同一图像的实验结果部分图像如图 6,结果比较如表 1。所用 100 幅测试样本包括视频叠加文本和场景文本,在这 100 个样本中,人工标注 197 个文本行。



(a)基于 Adaboost 分类器的分类效果



(b)基于 SVM 分类器的分类效果

图6 基于 Adaboost 分类器和基于 SVM 分类器的分类效果比较

表 1 基于 Adaboost 分类器和基于 SVM 分类器
实验结果比较

	准确率/ p	召回率/ r
基于 Adaboost 的分类器	0.761	0.740
基于 SVM 的分类器	0.682	0.636

通过表 1 可以看出本文提出的方法不管在准确率还是召回率上都是优于使用基于 SVM 分类器的效果的。并且在实验过程中,在同样的机器上本文方法的执行效率也比基于 SVM 方法的分类器的执行效率高。这也恰恰满足了视频文本定位快速性的特别要求要求。

本文使用 100 幅图像作为测试样本,测试结果准确率达到 73%。由于训练样本有限,通过增加训练的次数,提高文本定位的准确率和召回率。

4 结 论

本文基于 Adaboost 算法将视频文本定位问题转化为二分类问题,将多个弱分类器组合成一个强分类器,实现了对视频文本的定位。并且在视频文本定位的正确率、实时性和泛化能力上取得了良好的效果。但是,在实验过程中发现系统对噪声比较敏感,所以在训练的后期阶段对去噪方面集中了不少注意力,因此我们在选取样本时,应该选取更加全面的样本,以解决这一问题。

参 考 文 献:

- [1] LI H. Automatic Text Detection and Tracking in Digital Video [J]. IEEE Transition Image Proccession, 2000, 9(1): 147 - 156.
- [2] ZHONG Y, ZHANG H J, JAIN A K. Automatic Caption Localization in Compressed Video [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000, 22(4): 385 - 392.
- [3] David Crandall. Extraction of Special Effects Caption Text Events from Digital Video [J]. International Journal on Document Analysis and Recognition, 2003(5): 138 - 157.
- [4] HUA X, CHEN X, WEN Y L, et al. Automatic Location of Text in Video Frames [J]. Proceeding of ACM Multimedia 2001 Workshops: Multimedia Information Retrieval (MIR2001), 2001: 24 - 27.
- [5] CHEN X R, YUILLE A L. Detecting and Reading Text in Natural Scenes [C] // Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004: 366 - 373.
- [6] 吴丹敏. 复杂背景下的视频文本的提取 [D]. 西安: 西安科技大学, 2013.
- [7] 付光远. 一种基于 Sobel 分解算子的图像边缘检测并行算法 [J]. 微电子学与计算机, 2006, 23(9): 132 - 134.
- [8] 尹芳, 陈德运, 吴锐. 改进的 Gabor 小波变换特征提取方法 [J]. 计算机工程, 2012, 38(15): 145 - 147.
- [9] 张建明, 王娟, 张菊, 等. 基于条件笔画密度提取的文本定位方法 [J]. 计算机工程与设计, 2011, 32(10): 3446 - 3449.
- [10] 程豪, 黄磊, 刘昌平. 基于笔画和 Adaboost 的两层视频文字定位算法 [J]. 自动化学报, 2008, 34(10): 1312 - 1318.
- [11] 马海清, 黄剑华, 刘家峰. 于边缘检测和纹理分析的图像定位方法 [J], 计算机应用, 2008(S1): 41 - 44.
- [12] VIOLA P, JONES M. Fast and Robust Classification Using Asymmetric Adaboost and a Detector Cascade [J]. Advances in Neural Information Processing Systems, 2002(14): 1311 - 1318.
- [13] 孙红星, 赵楠楠, 徐心和. 基于小波变换和 SVM 的文本区域定位 [J]. 东北大学学报(自然科学版), 2007, 28(2): 165 - 168.

(编辑: 温泽宇)