基于连通分量特征的文本检测与分割

蒋人杰 戚飞虎 徐 立 吴国荣
(上海交通大学计算机科学与工程系,上海 200240)

摘 要 自然背景中的文本识别具有巨大的应用价值,但其应用却一直受到文本检测和分割技术的限制。为了更有效地进行文本检测与分割,提出了一种基于连通分量特征的自然场景中文本检测分割算法。该算法首先将原始图片通过 Niblack 方法分解为许多连通分量;接着,用一个级联分类器和一个 SVM 组成的两阶段分类模块来验证这些连通分量的文本特征。由于文本连通分量和非文本连通分量在特征上存在差异,大多数非文本会被级联分类器丢弃,而 SVM 则能在此结果上做进一步的验证,因此最终输出只有文本的二值图像。最后用该算法在测试数据上进行了评估实验,评估结果表明,检测精度超过 90%,响应超过 93%。

关键词 级联分类器 两阶段分类 文本检测 文本特征
中图法分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2006)11-1653-04

Using Connected-Components' Features to Detect and Segment Text

JIANG Ren-jie, QI Fei-hu, XU Li, WU Guo-rong
(*Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai* 200240)

Abstract Text recognition in natural scenes has a promising future, but its application is limited by the technique of text detection and segmentation. To detect and segment text effectively, this paper proposes an approach for detecting and segmenting text from scene images by using Connected-Components' features. First, the image is decomposed into a list of Connected-Components(CCs) by Niblack algorithm. Then all the CCs' features are verified by 2-stage classification module which is composed by a cascade classifier and a SVM. Most of non-text CCs are filtered out by cascade classifier and the remaining CCs are further verified by SVM. The final outputs are binary images containing texts only. Experiments have been taken on lots of images, the precision is more than 90% and recall is more than 93%.

Keywords cascade classifier, 2-stage classification, text detection, text feature

1 引 言

文本是日常生活中不可或缺的信息来源,其往往出现于路牌、布告、店名上。如果有一种算法能自动地对自然场景下的文本进行检测,并分割出来,那么在加上识别模块后,就可在很多方面有巨大的应用价值。

近年来,对于文本检测方向已经进行了许多研究,现有方法大致可以划分为基于纹理的方法和基于区域的方法两类。其中,基于纹理的方法是把文本视作一种特殊的纹理,通过检测这种纹理的响应来定位文本并常用基于局部特征[1]、快速傅里叶变换[2]、Gabor[3]、小波[4]等时域、频域方法来达到检测文本区域纹理强响应的目的;而基于区域的方法则是先把图像分解为许多小对象,而且依据小对象的类型,还可以进一步分为基于连通分量(connected-component,简称 CC)的方法[5]和基于边缘的方法[6],再通过验证合并这些对象来定位文本区域,其中启发式规则[5]、神经网络[1,2]、SVM(support vector machine,支撑向量机)[6]都是常见的验证方法,Zhu 提出的级联分类器给了本文启发[7]。

收稿日期:2006-06-01;**改回日期**:2006-07-30

第一作者简介:蒋人杰(1982~),男。2004 年获上海交通大学学士学位,现为上海交通大学计算机科学与技术系在读硕士研究生。主要研究方向为计算机视觉与模式识别。E-mail:blizard1982@sjtu.edu.cn

本文提出了一种基于连通分量特征的算法。该算法首先将输入的原始图像通过 Niblack 方法分解为许多连通分量;接着将所有连通分量都输入到由一个级联分类器和一个 SVM 组成的两阶段验证模块。由于级联分类器由一系列弱分类器串联组成,每个弱分类器关注连通分量的一个特征,并以很短的时间否决多数较为明显的非文本连通分量。而 SVM 则关注级联分类器输出的中间结果,用于进行更深入的验证。只有那些被级联分类器和 SVM 都接受的连通分量才会被认为是真正的文本连通分量,并最终输出到结果图像中。这样一种弱分类器和强分类器的结合,就保证了本文算法的有效性和高效性。

2 图像分解

考虑到自然场景中文本具有对比度低,背景复杂等问题,因此需要一种鲁棒的分解方法。本文采用 Niblack 算法来分解连通分量[8],即

$$J(x,y)=\begin{cases}+1, & I(x,y)>T_+(x,y)\\ -1, & I(x,y)<T_-(x,y)\\ 0, & \text{其他}\end{cases} \quad (1)$$

$$T_\pm(x,y)=\mu(x,y,W)\pm k\cdot\sigma(x,y,W) \quad (2)$$

其中,$\mu(x,y,W)$ 和 $\sigma(x,y,W)$ 分别表示像素点 (x,y) 邻域内的灰度均值和灰度标准差;W 为邻域尺度,在邻域尺度为 40×40Pixels 时,取得最佳效果;k 为经验常数,通常取 0.185。式(1)表示算法是将原始图像 $\boldsymbol{I}$ 变换为 Niblack 三值图像 $\boldsymbol{J}$,其包含白、灰、黑 3 个颜色层,分别对应 +1、0、-1 等 3 个值。其中白色和黑色为前景层,它们都有可能包含文本,而灰色为背景层,不作考虑。因此,提取连通分量的过程可分别在白色层和黑色层进行,其共同组成原始图像的候选连通分量集合。

在本文算法中,每个连通分量保存属于该连通分量的所有像素点,以及其中心色彩 (r_c,g_c,b_c):

$$(r_c,g_c,b_c)=\frac{1}{n}\sum_{i=1}^{n}(r_i,g_i,b_i) \quad (3)$$

中心色彩 (r_c,g_c,b_c) 就是属于该连通分量所有像素点的平均颜色。图 1 显示了 Niblack 分解的结果,图 1(a)经式(1)计算,即得到 Niblack 图像(图 1(b)),其包含白、灰、黑 3 个颜色层及黑白二层共有数千个连通分量,图 1(c)则是用连通分量的中心色彩来表示属于各自的像素点而得到的彩色连通分量图。

(a)原始图像 (b)Niblack 的连通分量图 (c)彩色连通分量图

图 1 Niblack 分解结果

Fig. 1 Niblack decomposition result

3 分类算法细节

本文的分类算法分为:由级联分类器实现的粗分类和由 SVM 实现的细分类。本节将依次描述文本特征的提取,以及级联分类器和 SVM 的使用。

3.1 文本特征

众所周知,由于分类的效果依赖于特征的有效性,因此特征的选择对于分类算法具有非常重要的意义。本文一共提出了 17 个特征用来区分文本连通分量与非文本连通分量。所有这些特征可以被归为:几何特征、形状规则化特征、边缘特征、笔画特征和空间相关特征 5 类。

几何特征用于测量连通分量的尺寸、宽度、高度、位置、长宽比、占空比等基本特征,由于大多数非文本连通分量较之文本连通分量来说,在几何特征上有明显区别,因此可以用这些特征来否决许多明显的非文本连通分量。

形状规则化特征用于衡量连通分量形状上的规则程度,以进一步探索文本连通分量和非文本连通分量之间的区别,这类特征包括:连通分量内部的空洞数量、周长面积比、边界光滑度等等。

边缘特征是字符内在的本质特征。本文提出以下两个基于边缘的特征:边缘对比度特征 $F_{\text{EdgeContrast}}$ 和边缘角度分布特征 $F_{\text{EdgeAngle}}$。其中,边缘对比度特征 $F_{\text{EdgeContrast}}$ 用于衡量字符与背景的对比度(见式(4))。C 是当前连通分量,$\boldsymbol{B}_C$ 是连通分量 C 的边界像素点集合,$\boldsymbol{E}$ 是输入图像的 Canny 边缘点集合。另一方面,通过统计处于文本连通分量边缘上的像素点发现,它们的梯度方向分布呈现极大的对称性[1];而对于非文本连通分量,却不存在这样的对称性。因此,本文算法使用特征 $F_{\text{EdgeAngle}}$ 来衡量这种角度分布的对称性(见式(5))。其中 $A_C(\theta)$ 是连通分量 C 中梯度方向为 θ 的边缘像素点数量。边缘对比度特征和边缘角度特征为

$$F_{\text{EdgeContrast}} = \frac{\boldsymbol{B}_C \cap \boldsymbol{E}}{\boldsymbol{B}_C} \tag{4}$$

$$F_{\text{EdgeAngle}} = \sum_{\theta=0}^{\pi} |A_C(\theta) - A_C(\theta+\pi)| \tag{5}$$

笔画特征同样揭示了文本的本质。一个字符由笔画组成,通常笔画的宽度较小且粗细均匀。式(6)笔画宽度均值 $F_{\text{StrokeMean}}$ 以及式(7)笔画宽度标准差 $F_{\text{StrokeDev}}$ 分别对应了"小"和"均匀"两个特点。其中,$s(\cdot)$是数学形态学中的 Skeleton 算子;$d_{\min}$代表 Skeleton 框架上的像素点到笔画外部像素点的最小距离;$\mu(\cdot)$和 $\sigma(\cdot)$分别为计算的均值和标准差,

$$F_{\text{StrokeMean}} = \mu(d_{\min}(s(C))) \tag{6}$$

$$F_{\text{StrokeDev}} = \sigma(d_{\min}(s(C)))/\mu(d_{\min}(s(C))) \tag{7}$$

以上所有特征都是基于单个连通分量的。然而,通过多个连通分量直接的空间关系,可以找到更多特征来进一步提高分类算法的性能。空间相关特征正是基于这样一种考虑,这些特征包括相邻连通分量的距离、最小外接连通分量等等。

3.2　级联分类器

在粗分类的阶段中,本文使用一个级联分类器来否决大部分非文本连通分量。该级联分类器由一组弱分类器串联而成,每个弱分类器关注一个 3.1 节中提出的特征,共有 17 个弱分类器。图 2 显示了级联分类器的结构。

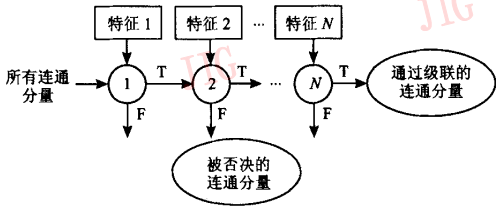

图 2　级联分类器结构

Fig. 2　Structure of cascade classifier

在粗分类算法开始的时候,所有在 Niblack 算法中得到的连通分量都被输入到第 1 个弱分类器。它只负责一个特征,依次为所有连通分量计算该特征,并依据特征值对输入连通分量进行判断:接受(T)与否决(F)。被否决的连通分量被视为非文本并立即丢弃;而对于接受的连通分量,则送入下一级弱分类器,重复相似的过程,直到整个级联判决结束。在级联分类器输出的中间结果中,大约有 90% 的非文本连通分量被丢弃,同时绝大部分的文本连通分量得以保留。图 3 显示了级联分类器的分类效果。

(a) 输入:所有连通分量　　(b) 输出:级联分类器分类后的连通分量

图 3　级联分类器处理结果

Fig. 3　Result of cascade classifier

更为重要的是,所有 SVM 用于精确分类的特征都可以在级联中以较小的时间代价获取。尽管 SVM 有着强大的分类能力,但在 SVM 分类前,还必需为每个输入连通分量计算所有 17 个特征,这将是一个非常耗时的工作,然而由于凭借着级联分类器的优势,没有必要为所有 Niblack 输出的连通分量计算 17 个特征,故可大大节省时间。非文本连通分量的数量在每一级弱分类器之后不断下降,由于多数非文本都在级联的早期就被否决了,因此节约了许多无意义的计算。尽管在输出的中间结果中仍有不少非文本存在,但是后继的 SVM 有能力纠正这些疏漏。Zhu 提出了一种级联分类器的训练算法[7],本文仍然使用该算法来训练级联分类器。

3.3　SVM 分类器

本文的精确分类阶段由 SVM 实现,以便对级联分类器输出的中间结果做进一步验证。本文算法将所有被级联分类器接受的连通分量都作为 SVM 的输入,而且它们的特征将被归一化,然后只有那些被 SVM 接受的连通分量才会被视作文本,并最终输出到结果图像中。图 4 显示了 SVM 的处理效果。

在本文中,SVM 是在训练连通分量库的子集上进行训练,而不是在整个集合上进行训练。训练中,

(a) 输入:级联分类器接受的连通分量　　(b) 输出:SVM 接受的连通分量

图 4　SVM 处理结果

Fig. 4　Result of SVM

本文算法将区别地对待文本连通分量和非文本连通分量，即保留所有的文本连通分量，而将所有的非文本连通分量都输入级联分类器，只有那些被误分类为文本的非文本分量才和保留的文本分量一起组成原始训练库的子集，用来训练 SVM。其目的是使训练连通分量的分布更接近于 SVM 在实际应用中会遇到的情况，以避免 SVM 过度关注于级联分类器可以解决的问题。

4 实验结果

考虑到文本连通分量尺寸一般比较大，而非文本连通分量尺寸一般比较小，直接比较最终接受的连通分量数量的评价方法显得不够公平。为此本文提出了一种严格的基于像素点的评价规则，即将结果图像与标注真值图像进行比较（图 5），按照分类输出结果中正确与错误像素点的数量来评价算法。

（a）原始图像 （b）结果图像 （c）标注真值图像

图 5 评价规则

Fig. 5 Evaluation rules

在图 5 中，结果图像和标注真值图像都是二值图像，文本像素点用黑色表示，值为 1；背景像素点用白色表示，值为 0。为了保证评价的正确性，所有的真值图像均由手工标注。本文使用准确率 $R_{\text{precision}}$ 和响应 R_{recall} 两个性能指标（见式(8)）进行评价。式(8)中 $\boldsymbol{O}$ 代表二值输出图像；$\bar{\boldsymbol{O}}$ 代表其补图像；$\boldsymbol{G}$ 代表标注的真值图像；$a(\cdot)$ 代表图像中值为 1 的像素点数量；N_1 代表被检测到的文本像素点数目；N_2 代表被误检测的非文本像素点数目；N_3 代表被漏检的文本像素点数目。

$$R_{\text{precision}} = \frac{N_1}{N_1 + N_2} \quad R_{\text{recall}} = \frac{N_1}{N_1 + N_3} \tag{8}$$

其中，$N_1 = a(\boldsymbol{O}\&\boldsymbol{G})$ $N_2 = a(\boldsymbol{O}\&\bar{\boldsymbol{G}})$ $N_3 = a(\bar{\boldsymbol{O}}\&\boldsymbol{G})$。

本文算法的测试环境是 Pentium4 3.0GHz CPU 的 Windows 平台。为测试算法性能，还构建了一个包含 500 幅自然场景图片的测试库，这些图片分辨率均为 640×480 Pixels，包括各种不同语言、字体、尺寸、颜色、倾斜角度、光照条件和表面的文本。平均处理时间小于 1s，准确率 $R_{\text{precision}}$ 和响应 R_{recall} 如表 1 所示。

表 1 算法性能

Tab. 1 Performance of our algorithm

	准确率(%)	响应(%)
训练集	92.34	94.57
测试集	90.88	93.12

5 结 论

本文提出了一种基于连通分量特征的自然场景中文本检测与分割算法。该方法对于各种尺寸、字体、倾斜角度、光照条件下字体的检测表现出很强的鲁棒性，但对于金属材质的文本，由于其表面的镜面反射严重污染了文本自身，致使 Niblack 方法无法将文本和背景分割，因此结果不能令人满意。本文算法在性能上尚有不足，也将做进一步改善。

参考文献(References)

1 Clark P, Mirmehdi M. Finding text regions using localized measures [A]. In: Proceedings of 11th British Machine Vision Conference [C]. Bristol, UK, 2000: 675 ~ 684.

2 Chun B T, Bae Y, Kim T Y. Automatic text extraction in digital videos using FFT and neural network [A]. In: Proceedings of IEEE International Fuzzy Systems Conference [C], Seoul, Korea, 1999, 2:1112 ~ 1115.

3 Chen D, Shearer K, Bourlard H. Text enhancement with asymmetric alter for video OCR[A]. In: Proceedings of International Conference on Image Analysis and Recognition [C], Venice, Italy, 2001: 192 ~ 197.

4 Mao W, Chung F, Lanm K, *et al*. Hybrid Chinese/English text detection in images and video frames [A]. In: Proceedings of International Conference on Pattern Recognition [C], Quebec, Canada, 2002, 3:1015 ~ 1018.

5 Wang K Q, Kangas J A. Character location in scene images from digital camera[J]. Pattern Recognition, 2003, 36(10): 2287 ~ 2299.

6 Kim K C, Byun H R, Song Y J, *et al*. Scene text extraction in natural scene images using hierarchical feature combining and verification [A]. In: Proceedings of International Conference on Pattern Recognition[C], Cambridge, UK, 2004, 2:679 ~ 682.

7 Zhu K, Qi F, Jiang R, *et al*. Using adaboost to detect and segment characters from natural scenes [A]. In: Proceedings of Camera Based Document Analysis and Recognition[C], Seoul, Korea, 2005: 52 ~ 59.

8 Winger L, Robinson J A, Jernigan M E. Low-complexity character extraction in low-contrast scene images [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2000, 14(2): 113 ~ 135.