

基于 LHP 多尺度向性滤波的文字检测算法

许鹏飞¹⁾ 姚鸿勋¹⁾ 纪荣嵘¹⁾ 王积成²⁾ 孙晓帅¹⁾

¹⁾(哈尔滨工业大学计算机科学与技术学院, 哈尔滨 150001) ²⁾(中国人民解放军电子工程学院, 合肥 230037)

摘要 复杂光照条件和文字—背景的交融是自然场景图像中文字检测的主要难点。为解决该问题,提出了基于 LHP(local Haar binary pattern)多尺度向性滤波的文字检测算法。该算法首先采用对光强变化不敏感并具文字特征显式描述特点的 LHP 模式的纹理描述算子;并在 LHP 模式上采用多尺度向性滤波器 MDF(multi-scale directional filtering)来确定候选文字区域;最后使用基于 LHP 直方图的支持向量机法精确定位文字区域。实验结果表明,与其他主流算法相比,该算法能够去除复杂光照条件和文字—背景交融的影响,具有更好的性能。

关键词 文字检测 Haar 小波 LHP 支持向量机

中图法分类号: TP391.4 文献标志码: A 文章编号: 1006-8961(2010)08-1207-07

A Text Detection Algorithm Based on Local Haar Binary Pattern with Multi-scale Directional Filtering

XU Pengfei¹⁾, YAO Hongxun¹⁾, JI Rongrong¹⁾, WANG Jicheng²⁾, SUN Xiaoshuai¹⁾

¹⁾(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

²⁾(Electronic Engineering Institute, Hefei 230037)

Abstract The main difficulties for text detection are due to complicated illumination condition and text-background weak contrast in natural scene images. In order to resolve them, this paper presents a local Haar binary pattern(LHP) based algorithm for text detection. Firstly, LHP texture description operator is presented. It is insensitive to illumination variance and can effectively describe the text feature. Secondly, the Multi-scale Directional Filtering based on LHP is proposed for fast filtering to obtain candidate text regions. Finally, the LHP-histogram-based SVM is presented to refine the text location. Comparing with state-of-the-art algorithms, the experiment results demonstrate the robustness of the proposed method with better accuracy.

Keywords text detection, Haar wavelet, local Haar binary pattern(LHP), support vector machine(SVM)

0 引言

有效地理解图像和视频内容一直是计算机视觉、模式识别领域追求和探索的目标之一。为了更好地理解图像视频内容的深层语义,文字信息是一个重要线索。因此,研究针对于视觉场景的文字检测算法对多媒体的语义理解有着重要的意义。

文字检测的目的在于检测图像或视频中的文字,

包括场景中固有的文字和图像视频编辑过程中嵌入的叠加文字。根据所采用的视觉特征的关联程度不同,现有文献的图像中的文字检测算法可分为基于边缘、基于连接成分分析和基于纹理等3类方法。

基于边缘特征的方法^[1-2]是利用文字区域存在较多边缘的特点,使用形态学膨胀等方法确定文字块。当背景复杂、高频量多、文字边缘特征不明显时,该类方法误检率较高。基于连接成分分析的方法^[3-4]是利用颜色特征从图像中提取出连通区域,

基金项目:国家高技术研究发展计划(863)项目(2006AA01Z197);新世纪优秀人才支持计划(NCET-05-03 34);国家自然科学基金项目(60775024)

收稿日期:2008-11-18;改回日期:2009-04-27

第一作者简介:许鹏飞(1982—),男,哈尔滨工业大学计算机科学与技术专业博士研究生。主要研究方向为文字检测与多媒体检索。
E-mail: pfxu@vilab.hit.edu.cn

再根据区域几何约束使用阈值规则确定文字区域。该类方法遇到图像背景复杂时很难准确提取出连通区域,而算法中所使用的几何规则和阈值更缺乏鲁棒性,很难推广。基于纹理的方法^[5-7]则认为文字区域具有特定模式的纹理分布。由于纹理特征不受图像质量下降和背景噪声的影响,因此这类方法与前两类方法相比,对文字检测具有更好的鲁棒性,这也在国际文字检测公开测试中得到充分的验证^[8]。

然而基于纹理的方法中仍然有两个问题没有解决好,这严重制约了自然场景文字检测算法性能。第一,复杂光照条件,如强光或图像中光照变化大,导致图像信息细节弱化,使文字—背景纹理差异淹没于光照差异和噪声之中而无法检测(图 1(a));第二,背景的交融,包括文字—背景对比度微弱和对比度变化,会直接影响前景与背景的界定,相对纹理量的改变导致阈值判断错误或者漏检(图 1(b))。



(a) 强光 (b) 对比度变化

图 1 图像示例

Fig. 1 Samples of image

本文提出了对光强变化不敏感的 LHPB 纹理描述算子,能够对复杂光照条件和背景的交融情况进行统一的纹理描述,解决了上述问题。并提出多尺度向性滤波器 MDF 和 LHPB 直方图支持向量机法,组成了新的文字检测算法,算法框架如图 2 所示。

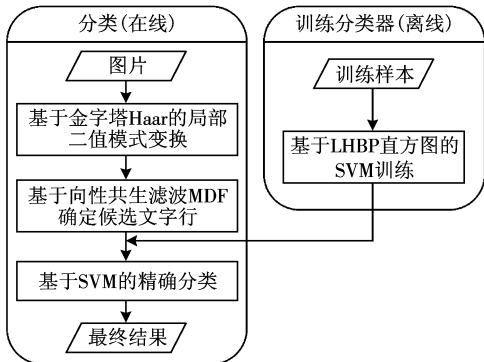


图 2 基于 LHPB 多尺度向性滤波的文字检测算法框图

Fig. 2 The flowchart of text detection algorithm based on LHPB with multi-scale directional filtering

1 LHPB 模式

局部二值模式(LBP)^[9]是一种局部纹理描述模式,它只记录模式的变化趋势,而不计较变化大小,用它来描述纹理,具有良好的抗光照变化的特性,同时具有旋转不变性和尺度不变性等特点^[9-11]。

提出了一种新的纹理描述方法——LHPB(local Haar binary pattern):在多尺度 Haar 小波的每个尺度的高频子带能量上使用带阈值的邻域算子来描述局部纹理模式,用 LHPB 值表示,定义为

$$LHPB_k(x_c, y_c) = \sum_{p=0}^7 S(f^k(x_p, y_p) - f^k(x_c, y_c)) \times 2^p \quad (1)$$

式中, $S(x) = \begin{cases} 1 & x \geq T \\ 0 & x < T \end{cases}$, $f^k(x, y) = \sum_{q \in \{HL, LH, LL\}} (h_q^k(x, y))^2$ 。其中, $f^k(x, y)$ 为第 k 个尺度在 (x, y) 处的小波能量, $h_q^k(x, y)$ 为点 (x, y) 的第 k 个尺度的 Haar 小波高频子带系数, (x_c, y_c) 为中心像素, p 为中心像素近邻编号, (x_p, y_p) 为近邻像素, 2^p 为编码权值。 T 为自适应纹理阈值,是根据小波能量分布情况确定的。

式(1)的意义在于,通过每个像素与其周围像素的 Haar 小波高频子带能量的差异来描述该点的纹理,并按这种差异进行编码,每种编码代表相应的纹理模式。判别函数 $S(x)$ 的作用为在比较像素能量时,只记录能量高于自身或低于自身两种情况(分别用 1 和 0 来表示),也就是在描述局部纹理时不刻画像素能量差异的多少,只描述差异的正向或逆向,并对不同位置的邻域赋予不同的权值进行编码,用不同的码值表示不同的纹理模式。这种纹理编码模式对图像中文字—背景交融的各个局部区域(图 1(b)),无论对比度差异多大,都能够采用统一的编码将相应的纹理描述出来。

式(1)中,阈值 T 的作用是通过一个阈值来去除图像中由强光或光照变化而产生的纹理。该纹理在相邻像素之间的差异较小,但在一定范围内都存在,对准确提取文字—背景的不同纹理模式有较大的影响。阈值 T 去除对光强变化的敏感,使算法有效地提取出真正由文字—背景差异而产生的纹理。在实验中发现,这种光照引起的纹理的大小与图像中纹理的整体分布相关,但根据具体的分布参数确定需要较高的运算复杂性,这里采用的是经验阈值,在实验中 T 如式(2)时效果较好。

$$T = 0.05 \times \max(f(x, y)) \quad (2)$$

与以往纹理描述方法相比, LHPB 模式纹理描述算子在多尺度的小波特征上提取局部纹理模式,

阈值 T 能够在保留文字—背景对比的情况下, 滤掉光照复杂和文字—背景交融情况的影响, 有效地提取文字区域的纹理特征(图 3, 图 4)。



图 3 小波能量特征和 LHPB 的对比

Fig. 3 Haar wavelet energy feature and LHPB feature

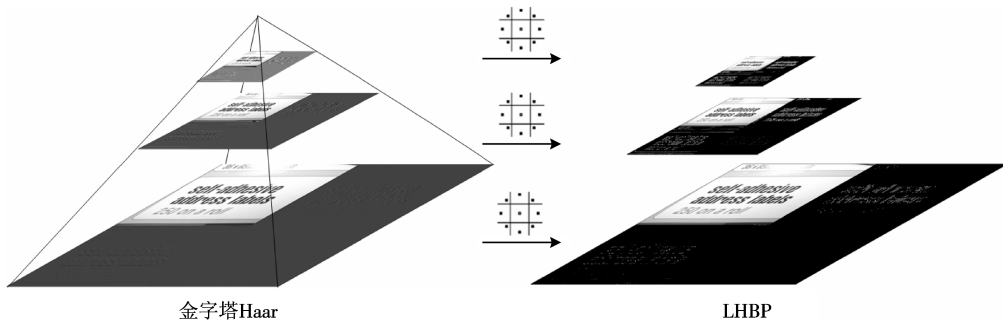


图 4 基于 LHPB 的多尺度纹理特征提取

Fig. 4 LHPB based multi-scale feature extraction

2 多尺度向性滤波器 MDF

2.1 文字的向性分析

与非文字区域相比, 文字区域在水平、竖直、对角线和反对角线 4 个方向上的纹理特征很明显。通常的汉字可以看成是横(水平)、竖(竖直)、撇(对角线方向)、捺(反对角线方向) 4 个笔画的组成(折等笔画可以由以上笔画组合而成), 可以认为文字区域的向性纹理由以上 4 个方向中的至少 2 个方向组成。对于英文和阿拉伯数字, 表 1 中的统计结果显示, 除数字 1, 英文字符 l 以外, 其他英文字符和阿拉伯数字的笔画也符合上述向性纹理特征。

据上述分析, 可根据纹理的方向性来确定候选文字区域: 统计区域中水平、竖直、对角线、反对角线 4 个方向中每个方向的像素个数, 如果有 2 个以上超过规定的阈值, 则认为这个区域为候选文字区域。图 5 中文字区域 LHPB 直方图中水平、竖直、对角线、反对角线 4 个方向(圆圈位置)的 bin 值比非文字区域高很多, 这证明了以上推断的正确性。

表 1 常见字符笔画方向相关性

Tab.1 Relativity of characters' strokes

笔画方向相关性	常见字符
横—竖相关	B C D E F G H J L O P Q R S T U a b c d e f g h j m n o p q r s t u 0 2 3 4 5 6 8 9
横—撇相关	A Z z 7
横—捺相关	A
竖—撇相关	K M N Y k 4
竖—捺相关	K M N R Y
撇—捺相关	A K M V W X Y k v w x y

2.2 LHPB 向性纹理编码表 (DTCT)

与 LBP 类似, 不同的 LHPB 编码, 反映了中心像素周围纹理的分布及方向, 因此可以通过 LHPB 编码来判定纹理走向。图 6(a) 为图像经过 Haar 变换后的一个 3×3 区域的小波能量特征分布的例子, 对图 6(a) 的实例采用如图 6(b) 给定的权值表进行编码, 对应的 LHPB 编码为 1100001_2 (图 6(c) 所示), 显然该编码反映的是竖直方向的纹理特征。

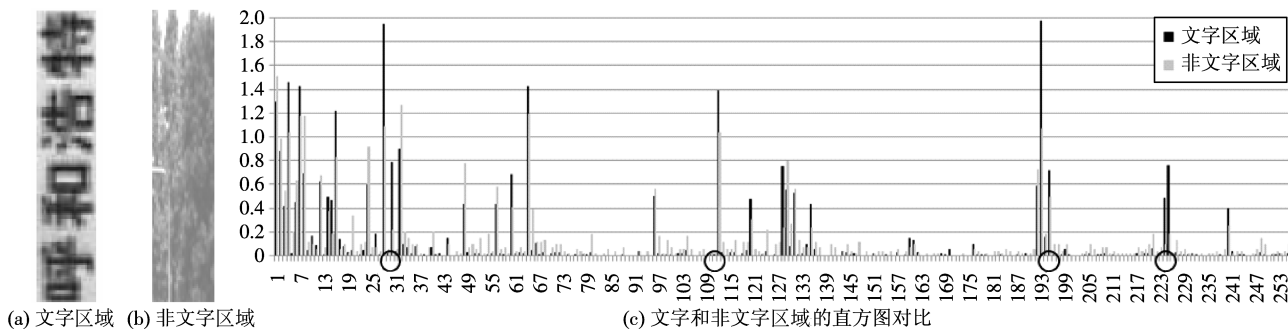


图 5 LHPB 直方图分析

Fig. 5 Analysis of LHPB histogram

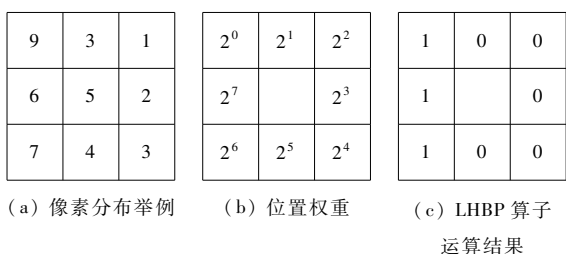


图 6 LHPB 算子计算示例 ($LHPB = 11000001_2 = 193$)

Fig. 6 The example of LHPB operator

对 LHPB 的各个编码进行分析,可获得各个不同方向上的纹理所对应的不同的 LHPB 编码,表 2 给出了水平、竖直、对角线、反对角线 4 个方向上的 LHPB 向性纹理编码表(DTCT),根据 DTCT 可以很清楚地判断某个 LHPB 编码所对应的纹理方向,进而利用其方向特征来判断是否为文字纹理。

表 2 LHPB 向性纹理编码表(DTCT)

Tab.2 Directional texture coding table of LHPB

纹理方向	加权的 LHPB 编码
水平	7,112,119
竖直	28,193,221
对角线	4,14,64,224
反对角线	1,16,56,131

2.3 基于 MDF 的候选文字区域定位

根据文字纹理的方向性分布特性和 LHPB 编码对纹理方向性的刻画,提出了基于 LHPB 的多尺度向性滤波器 MDF 文字定位方法,具体描述如下:

输入:待检测的 LHPB 图像。

输出:确定相应区域是否为文字区域的二值判定。

具体步骤:

1) 使用 $m \times n$ 的检测窗口模板在 LHPB 图像的

LH,HL,HH 域上滑动,使用式(1)计算当前检测窗口在 3 个域上 LHPB 模式分别为编号为 d 的方向上的方向纹理特性 ($d = 0, 1, 2, 3$, 分别代表水平、竖直、对角线和反对角线 4 个方向)。采用的邻域检测模板为

$$DirThreshold_{(i,j)}^d = \sum_{k=-m/2}^{m/2-1} \sum_{l=-n/2}^{n/2-1} DirTexture_{(i,j)}^d(k,l) \quad (3)$$

式中,

$$DirTexture_{(i,j)}^d(k,l) = \begin{cases} 1 & LHPB(i+k,j+l) \in DTCT^d \\ 0 & LHPB(i+k,j+l) \notin DTCT^d \end{cases} \quad (4)$$

式中, i, j 分别为检测模板起始点行列坐标。

2) 根据是否有超过 2 个方向的纹理共生情况来判断该区域是否为文字区域

$$DirFiler = \begin{cases} 1 & DirFlag \geq 2 \\ 0 & DirFlag < 2 \end{cases} \quad (5)$$

$DirFlag$ 由式(6)求得

$$DirFlag = \sum_{d=0}^3 Flag(d) \quad (6)$$

式中,

$$Flag(d) = \begin{cases} 1 & DirThreshold_{(i,j)}^d \geq T^d \\ 0 & DirThreshold_{(i,j)}^d < T^d \end{cases} \quad (7)$$

式中, T^d 为在方向编号为 d 的方向上的阈值。

3) 如果 $DirFalg = 1$,即当前检测窗口模板中的纹理分布满足向性文字阈值滤波的要求,则将当前滑动窗口检测模板对应区域置为文字区域;

4) 各尺度检测完毕,进入 5); 否则返回 1);

5) 融合多尺度检测结果,在最终的二值检测模板上标示出文字区域。

MDF 中首先对检测窗口中的纹理分布进行统计,如果在水平、竖直以及 2 条对角线方向中同时有 2 个或者有 2 个以上方向上的纹理分布强度大于相

应的阈值,将检测模板所检测到的区域设置为文字区域,这有效地利用了 LHPB 的编码特性,以及文字区域的向性纹理分布特性。

在 MDF 中,采用根据图像的纹理分布情况来自适应确定参数的方法,避免了固定阈值对数据集的依赖问题。具体参数设定方法如下:

模板大小:考虑到 OCR 文字识别的最小的高度为 12 个像素以及每行文字分布的连续性,检测窗口模板大小定义为 16×12 ,这样保证了通过多尺度的变换,在不需要调整模板大小的情况下检测到不同大小、不同粗细的文字。

向性滤波阈值 T^d :该阈值反映的是检测区域中相应向性纹理像素所占的比例,在实验中发现,一个区域内文字像素的个数常常在一定的范围内,不同尺度采用一个统一的与检测窗口总像素个数 s 相关的阈值即可(式 8),不同疏密的文字可以在不同的尺度上满足阈值而被检测出来。

$$T^d = s \times 0.3 \quad (8)$$

在多个测试集上的测试结果表明,本文的自适应的阈值确定方法是有效的。

3 基于 LHPB 直方图 SVM 的精确定位

LHPB 模式值给出了图像中对应局部纹理的特征,LHPB 直方图可以合并反映某一块图像的整体纹理分布模式(图 5)。因此,使用 LHPB 直方图精确定位文字区域,可以去除与文字区域纹理相似、向性分布相近的背景区域,进一步提高检测精度。

本文选用 SVM 分类器^[12]对候选区域进行特征统计分类。对 MDF 粗定位的候选文字区域进行 LHPB 直方图统计,LHPB 直方图特征输入到 SVM 分类器中进行训练,以识别文字和非文字区域对应的纹理分布的整体特点并进行正确分类,有效地分辨出文字与非文字区域(图 7)。

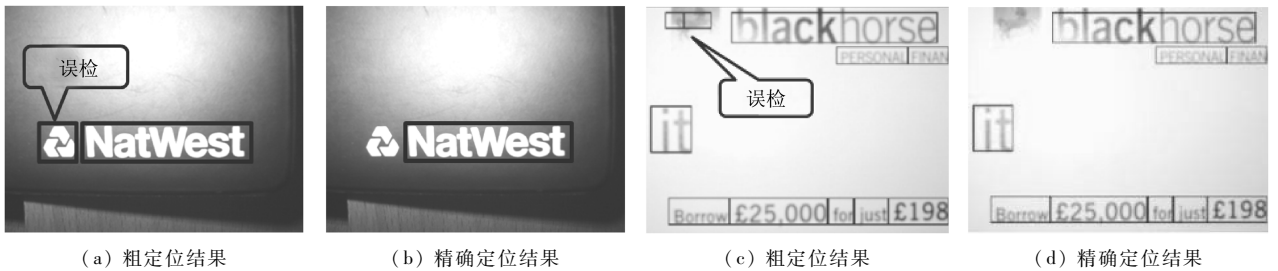


图 7 文字粗定位和精确定位的结果

Fig. 7 Results of the candidates and the precise of text area

4 实验结果及讨论

为了验证本文算法的性能,采用了 2003 年国际自然场景文本阅读比赛(ICDAR03)^[8]提供的文字检测数据集进行了测试。该数据集包含 258 幅训练图像和 249 幅测试图像。本文使用训练样本集中标定的 1 127 个文字区域以及随机从训练样本中选取的不含文字的图像块分别作为正例和反例样本,训练系统的精确分类部分的 SVM 分类器。实验测试机为 P4 3.0G CPU 和 1G 内存的 PC 机。

本文在上述独立于训练集的测试集上进行了测试,采用类似于该比赛中的评测方法,通过准确率(p)和召回率(r)评价系统的性能,如式(9)所示。

$$p = c / |T|, r = c / |E| \quad (9)$$

式中, c 为正确检测到的文字个数, T 为检测结果的

集合, E 为测试集中标定的文字集合。综合评测结果 f 如式(10)所示,本文中取 $a = 0.5$ 。

$$f = \frac{1}{a/p + (a-1)/r} \quad (10)$$

首先,为了评估本文算法在复杂光照和文字—背景交融情况下检测的有效性和鲁棒性,针对性地与单纯小波能量纹理特征检测方法^[7]进行了比较实验,结果如图 8 所示。本文算法综合性能高出 20% 左右。这是因为本文算法中 LHPB 编码对复杂光照等具有鲁棒性,去除了强光或文字—背景对比度变化等对纹理检测的影响,如图 9(a),图 9(b)所示。

其次,为验证本文算法多尺度检测的有效性,将本文算法与直接使用 LBP 运算的方法进行了比较,该对比方法不使用多尺度 Haar 变换而直接在灰度图像上进行 LBP 操作,后续步骤与本文算法相同。

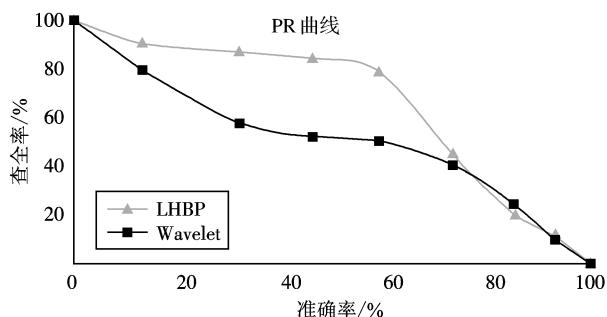


图 8 实验结果 PR 曲线

Fig. 8 P-R curve of experiment results

所提方法比单纯的 LBP 方法,综合性能高出 47% (如表 3 所示)。由此可见,多尺度 Haar 小波模式可以在不同尺度上很好地检测到不同大小、不同笔画

粗细的文字纹理。实验表明,本算法能够检测不同大小的文字(图 9(c),图 9(d))。

再次,将本文算法与基于边缘和连接成分分析的文字检测代表性算法^[4]以及国际自然场景文本阅读比赛的算法^[8]进行了对比。本文算法与边缘特征的方法准确率相近,但查全率高出 15%,综合性能高出 6% (如表 3 所示);在与 2003 年比赛结果的比较中,本文算法比其性能最佳算法(Ashida)的查全率和准确率分别高出 4% 和 33%,综合性能高出 18%。采用基于 MDF 的方法根据文字纹理向性特征来对图像中的文字进行粗定位,有效地减少了误检,基于 LHPB 直方图的 SVM 精确分类能够进一步提高系统性能,实验表明本文算法是有效的。

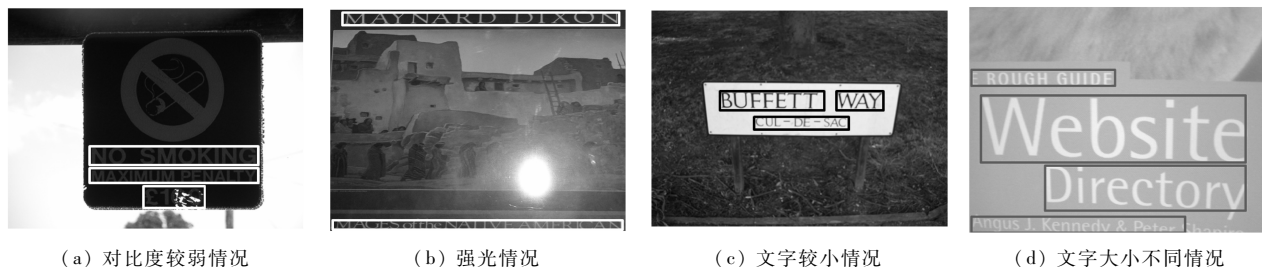


图 9 所提算法检测结果

Fig. 9 Results of proposed method

表 3 各个方法准确率和查全率比较

Tab. 3 Precision/Recall comparison

方法	准确率	查全率	综合性能	备注
基于 LHPB	0.59	0.79	0.68	
基于 LBP	0.29	0.17	0.21	
边缘特征 ^[4]	0.60	0.64	0.62	*
Ashida	0.55	0.46	0.50	比赛第 1 名 ^[8]
HWDavid	0.44	0.46	0.45	比赛第 2 名 ^[8]
Wolf	0.30	0.44	0.35	比赛第 3 名 ^[8]
Todoran	0.19	0.18	0.18	比赛第 4 名 ^[8]

*注:该数据为文献[4]作者从 ICDAR'2003 数据集选择的部分图像中的测试结果。

最后,为了验证本文算法的时间开销,与使用小波特征及神经网络分类的方法^[5]、使用灰度纹理及 SVM 分类的方法^[6]进行了对比实验。本文算法每秒钟能够处理 5 张以上平均大小为 640×480 的彩色图像,若应用在 25 帧/s 的视频中,采用 1/5 亚采样,即能够达到实时处理。从表 4 可以看出,本文算法处理时间明显低于其他方法,同时,算法的综合性

能也大大高于其他方法。这是因为运用 LHPB 模式能直接有效地描述文字区域纹理特征,MDF 多尺度向性滤波的快速过滤机制,使得算法能够在保证查全率的同时有效地过滤绝大部分非文字区域,有效地缩减了需精确检测的范围,大大压缩了时间开销。

表 4 各算法执行时间比较

Tab. 4 Speed of executive comparison

方法	准确率	查全率	处理速度
LHPB-MDF	0.59	0.79	5.3 images/second
Wavelet-NN ^[5]	0.48	0.68	1.2 images/second
Gray-SVM ^[6]	0.50	0.71	2.1 images/second

5 结 论

提出了一种 LHPB 模式的局部纹理描述算子,及该模式下的多尺度向性滤波器 MDF 和基于 LHPB 直方图支持向量机法,构成基于 LHPB 的多尺度向性滤波文字检测算法。LHPB 模式纹理描述

算子通过统一描述图像中纹理信息,解决了复杂光照和文字—背景交融的图像的文字检测的难点问题;多尺度向性滤波器 MDF 和基于 LHP 直方图的支持向量机法,利用了文字区域的向性纹理分布特性,能够快速、准确地检测到图像中的文字。在国际标准测试集上的实验证明,本文算法在时间开销上是性能相近的最快速算法的 $1/4 \sim 1/3$,比其他主流算法综合性能提高了 $6\% \sim 18\%$ 。

在未来的工作中,考虑到图像中的文字走向常常具有一定的旋转及存在文字形变的情况,打算利用 LHP 多点采样集合旋转不变性来解决文字走向变化和文字形变的问题,进一步提高文字检测的准确率。

参考文献 (References)

- [1] Lienhart R, Wernicke A. Localizing and segmenting text in images and videos [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(4): 256-268.
- [2] Liu X Q, Samarabandu J. Multiscale edge-based text extraction from complex images [C]//Proceedings of IEEE International Conference on Multimedia and Expo. Los Alamitos, USA: IEEE Computer Society, 2006: 1721-1724.
- [3] Jain A K, Yu B. Automatic text location in images and video frames [J]. Pattern Recognition, 1998, 31(12): 2055-2076.
- [4] Ezaki N, Bulacu M, Schomaker L. Text detection from natural scene images: towards a system for visually impaired persons [C]//Proceedings of International Conference on Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2004: 683-686.
- [5] Li H, Doermann D, Kia O. Automatic text detection and tracking in digital video [J]. IEEE Transactions on Image Processing, 2000, 9(1): 147-156.
- [6] Kim K I, Jung K, Kim H. 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.
- [7] Ye Q X, Huang Q M. A new text detection algorithm in image/video frames [C]//Proceedings of Pacific-Rim Conference on Multimedia. Heidelberg: Springer Berlin, 2004: 858-865.
- [8] Lucas S M, Panaretos A, Sosa L, et al. ICDAR 2003 robust reading competitions [C]//Proceedings of International Conference on Document Analysis and Recognition. Washington, USA: IEEE Computer Society, 2003: 682-687.
- [9] Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on feature distributions [J]. Pattern Recognition, 1996, 29(1): 51-59.
- [10] Ojala T, Pietikainen M, Maenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.
- [11] Zhao G Y, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 915-928.
- [12] Vapnik V N. The Nature of Statistical Learning Theory [M]. New York, USA: Springer-Verlag, 1995: 105-150.