

中图法分类号: TP391.12 文献标志码: A 文章编号: 1006-8961(2011)06-0989-06

论文索引信息: 黄为, 王云宽, 李兵, 吴少泓. 结合笔画方向信息与SVM的英文文字检测[J]. 中国图象图形学报, 2011, 16(6): 989-994

结合笔画方向信息与SVM的英文文字检测

黄为, 王云宽, 李兵, 吴少泓

(中国科学院自动化研究所, 北京 100190)

摘要: 文字检测是文字信息提取系统中最重要的环节。针对在非均匀光照或背景图案复杂等恶劣环境下文本难以被检测的难题, 设计一种通用的基于笔画方向信息图的通用文字检测算法, 该算法采用由粗到细的实现框架。在粗的文字行定位中, 利用 Haar 小波和 LBP 描述符, 建立与原图像相对应的笔画方向信息图, 经滤波, 连通域分析以及 PPA 后得到候选的文字行; 在文字行精确分类中, 利用 SVM 分类器, 并结合多种文字的纹理特征, 确认最终的文本行区域。针对图片数据库 ICDAR03 的实验表明该算法能在不同条件下快速、准确地检测出文字区域, 文字检测的精确率为 0.64, 召回率为 0.67。

关键词: 文字检测; Haar 小波; LBP; 支持向量机

English text detection combining stroke direction information with SVM

Huang Wei, Wang Yunkuan, Li Bing, Wu Shaohong

(Institute of Automation, Chinese Academy of Sciences, Beijing 100190 China)

Abstract: Text detection plays an important role in the system of text information extraction. To solve the problem of detecting text in the bad settings such as complicated Text-background, uneven illumination and so on, a generation text detection method based on stroke direction map is proposed, which utilizes a coarse-to-fine framework. In the coarse step, it firstly builds the stroke direction Information map of the source images with Haar wavelet and LBP descriptor. Then the candidate text lines can be obtained through filtering, component analysis and projection profile analysis of the direction Information map. In the final step, it makes use of SVM classifier with kinds of text texture features to verify the real text lines. Experiment results on ICDAR03 show that the proposed algorithm can quickly and accurately detect texts region in different settings.

Keywords: text detection; Haar wavelet; LBP; SVM

0 引言

现今, 文字信息提取系统被广泛应用于各行各业, 例如工业检测中自动检测产品标号、生产日期; 智能交通系统中自动检测汽车的牌号; 视频信息的索引和检索; 自然场景中文字标志的辨识等。其中,

文字检测与定位是文字信息提取系统的起点, 能否精确检测和定位图像中的文字区域, 会直接影响文字分割和分子识别的准确度, 从而影响整个文字信息提取系统的性能。在过去的研究中, 研究人员已经提出了许多文字检测方法, 大致可以分为 4 类: 基于边缘和梯度的方法^[1-2], 基于连通域的方法^[3-4], 基于纹理的方法^[5], 以及基于学习的方法^[6-7]。4 种

收稿日期: 2010-05-07; 修回日期: 2010-07-21

基金项目: 中国科学院自动化研究所科技产业发展基金项目 (DC07J03)。

第一作者简介: 黄为 (1981—), 男, 中国科学院自动化研究所控制理论与控制工程专业博士研究生, 研究方向为机器视觉和智能检测。E-mail: huangfeihu1@gmail.com。

方法相比,基于边缘和梯度的方法计算简单,速度快,但不具有通用性,不能应用于对比度相差较小的情况;基于连通域的方法虽然定位效果比较好,但对图像的成像质量要求较高,并且不能应用于字符颜色或者灰度变化较大的图像,因此也不具备通用性;基于纹理的方法,能很好利用文字独立的纹理特征,具备一定通用性,但是定位精度不高;基于学习的方法,检测准确率较高,但由于需要对图像的每个区域进行判别,因此检测速度较慢,并且该方法十分依赖于训练样本和文字特征的提取。

鉴于以上所述,设计的文字检测定位算法摒弃了不具备通用性的基于边缘和基于连通域的方法,综合了基于纹理的方法和学习方法,并融入了多尺度的处理技术,采用由粗到细的检测框架,如图 1 所示,设计出了一个快速并且鲁棒的通用文字检测方法。

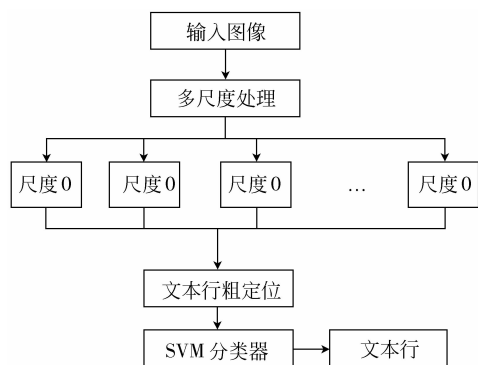


图 1 文字检测算法流程

Fig. 1 Flow chart of the text detection algorithm

1 文本行粗定位

粗定位文本行的流程如图 2 所示。



图 2 粗定位文本行流程

Fig. 2 Flow chart of coarse text line detection

1.1 高频系数的 LBP 特征图

Haar 小波变化具有良好的频域和时域的局部信息,能很好地反映图像文本区域的高频信息^[8];而 LBP^[9-10]是一种非常鲁棒的纹理描述符,能真实地反映文本行的纹理信息,将这两种特征结合在一起,就能得到图像高频系数的 LBP 特征图,该特征

图可直接反映文本区域高频信息下的纹理信息,该特征图的具体求取方法如下:

1) 对输入图像进行 Haar 小波分解^[8],得到图像 3 个方向的高频系数即水平 I_H , 竖直 I_V , 对角线 I_D , 以及 Haar 小波能量特征图 I_E :

$$I_E = \text{abs}(I_H) + \text{abs}(I_V) + \text{abs}(I_D) \quad (1)$$

2) 二值化 3 个方向的高频系数 I_H 、 I_V 、 I_D , 得到 3 个方向上的二值化图像即水平 I_{HB} , 竖直 I_{VB} , 对角线 I_{DB} , 二值化的方法如下:

选取一个 16×16 的滑动窗口 M , 按照步长 $(16, 16)$ 在高频系数图像上滑动, 计算局部阈值 T_L , 并利用该阈值二值化窗口滑动窗口, 直到遍历完整个图像:

$$T_L = \max(T_C, T_{L0}) \quad (2)$$

$$M_B(i, j) = \begin{cases} 0 & \text{abs}(M(i, j)) < T_L \\ 255 & \text{其他} \end{cases} \quad (3)$$

T_{L0} 是通过 Otsu 法计算出滑动窗口的局部阈值, 为保证速度, 此处 Otsu 法使用的直方图 Bin 的个数为 64。 T_C 是全局阈值:

$$T_C = \max(T_{C0}, 30) \quad (4)$$

30 是基本阈值^[11], 表示肉眼所能分辨的最低数值, 也就是说如果小波系数比这个数值还小, 那么人的肉眼几乎都看不到文字的存在, 不能分辨背景和文字像素。 T_{C0} 是使用 Otsu 法计算出高频系数图像的一个全局阈值, 为了保证阈值的精度, 此处 Otsu 法所使用的直方图 Bin 的个数为 256。

3) 利用 LBP 描述符描述每一个方向上的二值化图像 (I_{HB}, I_{VB}, I_{DB}), 最终得到 3 个方向高频系数的 LBP 特征图即水平 I_{HBL} 、竖直 I_{VBL} 、对角线 I_{DBL} 。 LBP 的描述符的定义如下:

对于输入图像 $I(x, y)$ 中一个给定的像素点 (x_c, y_c) , 其二值化 LBP 的映射值为^[9]

$$\text{LBP}(x_c, y_c) = \sum_{n=0}^7 s(G_n - C) 2^n \quad (5)$$

其中, C 表示像素点 (x_c, y_c) 的灰度值, 如图 3 所示, G_n 表示像素点 (x_c, y_c) 周围相邻点的像素值, 函数 $s(x)$ 的定义如下:

$$s(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases} \quad (6)$$

1.2 笔画方向信息图

英文字母和阿拉伯数字的笔画基本上是由横线、竖线和斜线组成。因此定位文字笔画的最佳方法就是找到上述 3 个方向上的信息点。

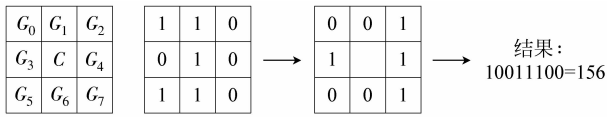


图 3 LBP 描述符

Fig. 3 LBP descriptor

经分析可知,组成文字中的横线、竖线的像素点的邻域结构和 LBP 值如图 4 所示。

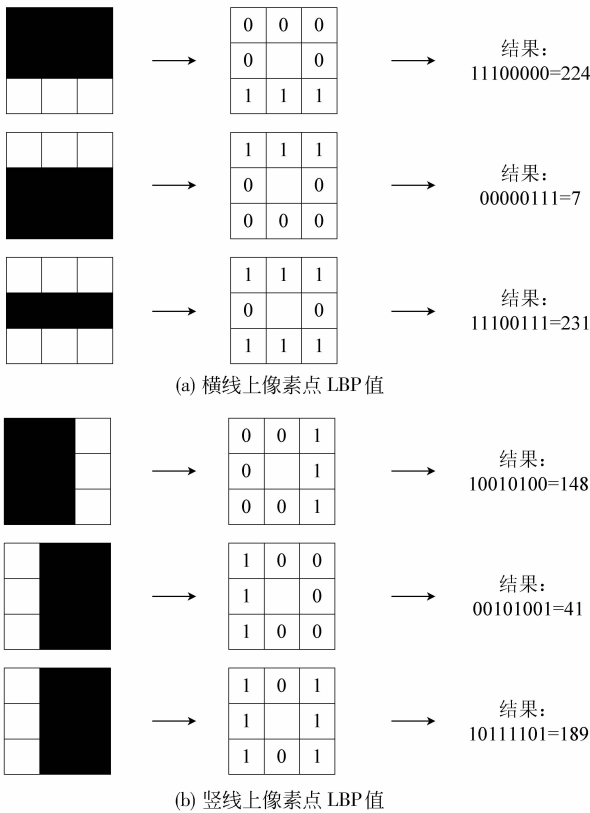


图 4 像素点的 LBP 值

Fig. 4 LBP of pixels

水平方向上高频系数的 LBP 特征图 I_{HBL} 包含大量的水平方向信息。因此原始图像中的所有横线的相关信息都会在 I_{HBL} 中出现。通过式 (7), 即可得到横线信息图 $I_{HM}(x, y)$:

$$I_{HM}(x, y) = \begin{cases} I_{HBL}(x, y) & I_{HBL}(x, y) \in \{224, 7, 231\} \\ 0 & \text{其他} \end{cases} \quad (7)$$

同理, 通过式 (8) 得到竖线信息图:

$$I_{VM}(x, y) = \begin{cases} I_{VBL}(x, y) & I_{VBL}(x, y) \in \{148, 41, 189\} \\ 0 & \text{其他} \end{cases} \quad (8)$$

较复杂的背景中, 即使没有横线或竖线的局部地方也会出现 148, 41, 189, 224, 7, 231 等值, 但数量很少, 可以被认为是孤立点, 通过形态学腐蚀可去除这些孤立点。

文字中的斜线不是标准的 45° 或者 135° , 存在如图 5 虚线所示的多种情况, 因此, 无法直接用特殊的 LBP 值去滤波高频系数图像来得到斜线信息图。经分析可知, 在出现斜线信息的局部范围内, 所呈现的 LBP 值应该是固定的, 它代表了该斜线方向的一种趋势。采取如下办法确定局部范围内的斜线信息的 LBP 值, 进而确定全局范围内的斜线信息 LBP 值:

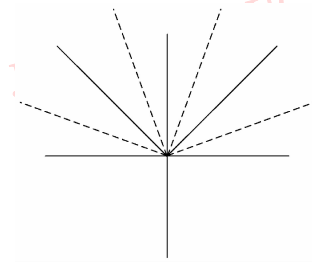


图 5 斜线的方向示意图

Fig. 5 LBP of pixels in Horizontal line

1) 计算 LBP 特征图的局部直方图 $Hist_N(i)$, 其中 $i = 0, 1, 2, \dots, 255$;

2) 不考虑横线和竖线, 因此将 $Hist_N(j) = 0$, 其中 $j = 0, 148, 41, 189, 7, 224, 231$;

3) 选取直方图中 4 个最大值所对应的索引号 ($k_{max1}, k_{max2}, k_{max3}, k_{max4}$);

4) 假设局部图像为 N , 根据式 (9), 得到过滤后的局部图像 N' :

$$N'(x, y) = \begin{cases} N(x, y) & N(x, y) \in \{k_{max1}, k_{max2}, k_{max3}, k_{max4}\} \\ 0 & \text{其他} \end{cases} \quad (9)$$

5) 遍历整个 LBP 特征图, 在得到整个 LBP 特征图所对应的斜线信息图后, 用形态学腐蚀去掉斜线信息图中的一些孤立点。

将上述方法用公式描述为

$$I_{out} = Dialog(I_{inLBP}) \quad (10)$$

式中, I_{inLBP} 表示为输入 LBP 特征图; I_{out} 表示为输出的斜线信息图。

由图 5 可知, 一条斜线不仅可以投影在对角线方向上, 还可以投影在水平和垂直方向上, 因此, 在考虑斜线信息时, 需综合水平、垂直、对角线 3 个方

向的 LBP 特征图的斜线信息图:

$$I_{DM} = Dialog(I_{HBL}) + Dialog(I_{VBL}) + Dialog(I_{DBL}) \quad (11)$$

综合斜线信息图 I_{DM} 、横线信息图 I_{HM} 、竖线信息图 I_{VM} ,即可得到笔画方向信息图 I_M :

$$I_M = I_{HM} + I_{VM} + I_{DM} \quad (12)$$

1.3 确定文字候选区域

英文字母和阿拉伯数字中的任何一个字符都会包含横线、竖线、斜线信息,因此在出现文本的区域,笔画方向信息会非常集中。鉴于此,文字候选区域由如下方法确定,选取一个 8×8 的滑动窗口 A ,按照步长(8,8)在笔画方向信息图上滑动,当滑动窗口 A 中非零元素的个数大于 σ 时,说明该窗口包含了丰富的方向信息,因此可被确定文字候选区域。在遍历完整幅笔画方向信息图后,就可以得到文字的候选区域。

1.4 文字候选区域分割文字行

通过连通域分析找到每个独立的候选区域的最大外接矩形,然后利用“projection profile analysis”^[12]的方法将多行的文字分开。对于切分得到的文字行,使用简单的结构特征去除那些不可能是文字的候选矩形区域。通过观察发现文字行的高度一般都超过 16 个像素,如果小于 16 个像素,肉眼很难分辨。因此,高度小于 16 个像素的候选将被删除。同时长宽比小于 0.5 的候选文字行也将删除。

2 基于 SVM 的文字行精确分类

通过文本行的粗定位,所有包含强烈规则方向信息的图像区域都可能被检测为候选文字,特别是对于图像中的条形状区域,例如砖头、门框、栅栏等。这些对象都有可能被检测为候选文字区域。为了去除这些与文字相似的非文字区域,本文选择了 SVM,并结合文字独特的纹理特征来判断候选的文本区域,进而决定最终的文本区域。

SVM 由 Vapnik^[13]最先提出,主要是解决二类分类问题,并在二类分类问题上,表现出极其优秀的性能。SVM 有多种类型的核函数,本文选取了具有通用性的核函数 RBF。

文本行可以看成一种特殊的纹理模式,有着独特的性质^[14],为了能准确反映这些特性,本文选取了以下统计量来描述文本行的纹理特征:Haar 小波矩、Haar 小波能量直方图、笔画方向信息图的直方

图以及文本行高度与宽度。

Haar 小波矩反应了小波系数的大小和分布均匀等情况,它包括均值、二阶中心距和三阶中心距。由于 Haar 小波分解,有三个子带的高频系数,即水平方向、竖直方向和对角线方向,因此在一个固定的尺度提取 3 个小波子带上的 3 维特征,总共得到 9 维 Haar 小波矩特征。

Haar 小波能量直方图统计 Haar 小波能量特征图(式(1))的一阶信号变换,反映文字区域的能量分布,该直方图 Bin 的个数为 16。

笔画方向信息图的直方图能很好地描述文字区域中的方向信息,体现出文本区域中方向的规则性和统一性。由上述可知,笔画方向信息图从总体上来说只记录横线即水平方向信息点、竖线即垂直方向信息点、斜线即特殊方向信息点和以及没有方向的信息点 4 类,因此在计算笔画方向信息图的直方图时,可以将笔画方向信息图中等于 0 的点即没有方向的信息点作为一类,横线信息的点作为一类,竖线信息的点作为一类,斜线信息的点作为一类。将统计好的每类点的数目除以总像素点数目,即可得到每类点的直方图信息。

对于文字而言,不同大小的文字,由于尺度不一致,因此纹理特征也会有稍许区别,为避免这一问题,将文本行的高度和宽度也作为特征,由此得到了总共所需要的 31 维的文本行特征。这些特征能很好地反映文本区域的纹理特征和方向信息特征。

3 实验结果与分析

实验选用的数据库是一个开放的大型数据库 ICDAR03 中的文字检测定位数据库,它包含一个测试和一个训练的图片集合。本文所用到的数据库为文字检测定位数据库,首先利用它的训练集合训练 SVM 分类机,该训练集合总共有 258 幅图片,然后在测试集合上来测试本文所述的方法,测试集合总共有 249 幅图片,在建立多尺度图像金字塔模型时,将图像金字塔中的尺度步长固定为 2,图像金字塔的层数为 3 层。整个算法是在 PC 上完成的,其中 PC 的 CPU 为 AMD 4200+,主频为 2.2 GHz。代码是以 C++ 编写。

图 6—9 给出了 ICDAR03 中 4 幅经典图像使用基于笔画方向信息图的文字检测算法得到的结果。这 4 幅图中,图 6 的原图含有阴影,图 7 的原图背景

复杂,图 8 原图背景还有大量噪声,图 9 原图光照不均。对比这 4 种类型的检测结果,可以看出本文所述的基于笔画方向信息的检测方法大致可以定位出不同背景条件下图像中的文本行。与此同时,仍然会有检测不到的文字行,例如图 8 中的单个字符“5”与图 9 的“7”和“a”,在文字行粗定位时可以找

到这些单个字符的位置,由于单个字符无法反映文本本行的特征,因此在基于 SVM 精确分类时会被认为是非文字行,但从图 6—9 整体的文字行的检测效果来看,基于笔画方向信息的检测方法还是比较鲁棒的。

在基于 SVM 文字行精确分类时,使用不同的文字行特征将会有不同的分类效果。表 1 对比了几种

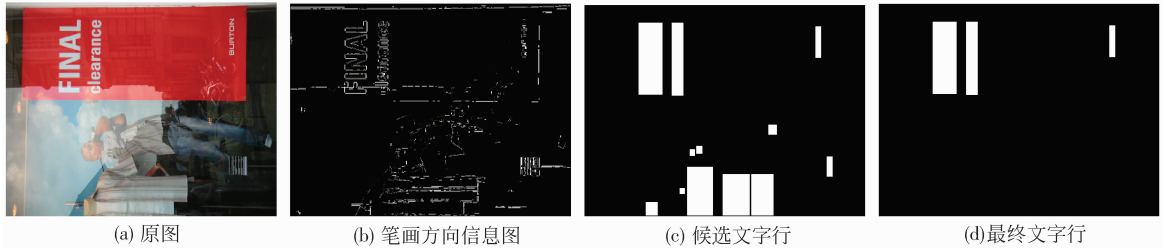


图 6 基于笔画方向信息图的文字检测示例 1

Fig. 6 Example No. 1 for text detection based on Stroke direction information map

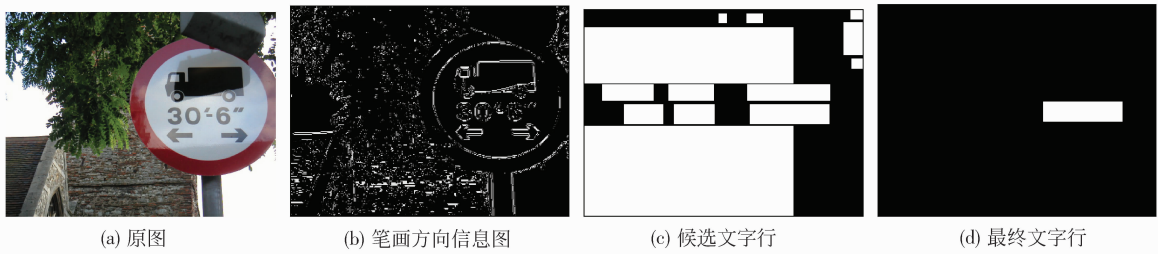


图 7 基于笔画方向信息图的文字检测示例 2

Fig. 7 Example No. 2 for text detection based on Stroke direction information map

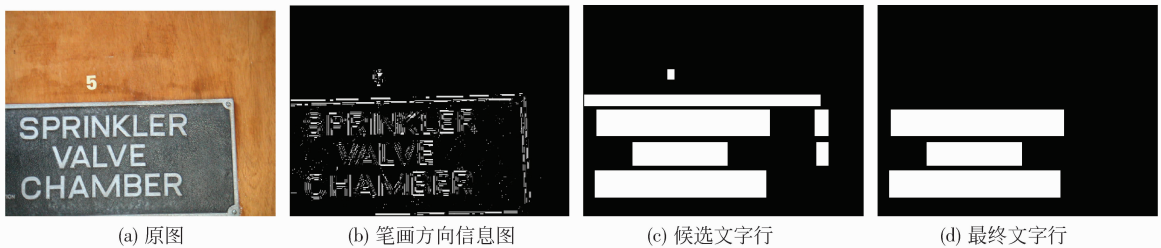


图 8 基于笔画方向信息图的文字检测示例 3

Fig. 8 Example No. 3 for text detection based on Stroke direction information map

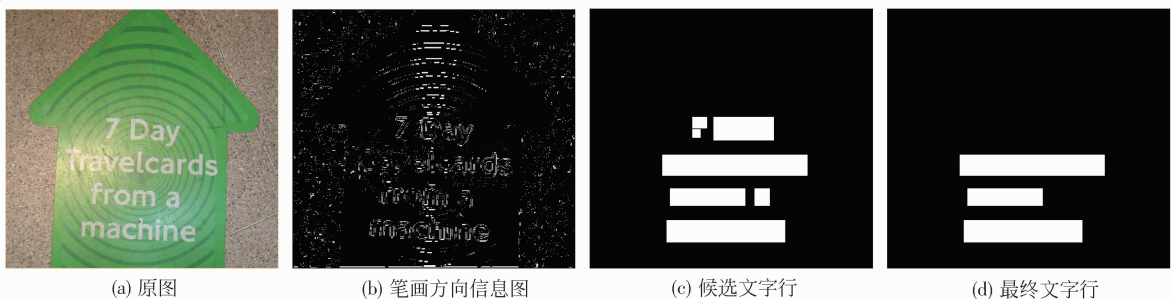


图 9 基于笔画方向信息图的文字检测示例 4

Fig. 9 Example No. 4 for text detection based on Stroke direction information map

不同文字特征在进行 SVM 文字行精确分类时的正确率。从对比结果可以看出,使用本文所述的文字行特征,在精确文字行分类时,正确率最高,这主要是因为在本文特征中加入了笔画方向信息图的直方图,该直方图能很好地反映出文字行的方向信息,这是其他两种文字行特征所不具备的。

表 1 不同特征的分类正确率

Tab. 1 Classifying right rate of different features

本文所述特征	0.967
Haar 小波矩和 Haar 小波能量特征图	0.931
LBP 特征	0.879

表 2 对比了本文所述算法与文献[15]中的 3 种检测算法在 ICDAR03 上实验结果,即精确率、召回率和检测时间,由对比结果可以看出,本文所述算法在精确率和召回率性能指标上,都优于文献[15]的 3 种检测算法。而在检测时间上,由于本次实验用到的 PC 性能要优于文献[15]近 2 倍,经过简单的换算,如果在与文献[15]相同的 PC 上运行本文的检测算法,大约需耗时 6.4 s,仅仅比 Alex 和 Qiang 的慢,但却比 Hinnerk 要快。因此,从整体的检测性能来说,本文的算法仍然会优于文献[15]中所提到算法。

表 2 与文献[15]中方法的对比

Tab. 2 Comparison with competitions in reference[15]

	精确率 P	召回率 R	检测时间/s
本文算法	0.64	0.67	3.2
Hinnerk	0.62	0.67	14.4
Alex	0.60	0.60	0.35
Qiang	0.33	0.40	1.6

4 结 论

介绍一种新的基于笔画方向信息图的文字检测方法,该方法采用由粗到细的检测结构,首先利用笔画方向信息图中的方向信息进行文字行的粗定位,然后利用支持向量机 SVM 结合文字 Haar 小波矩特征, Haar 小波能量特征图的直方图以及文字的笔画方向信息图的直方图等特征,确定最终的文字区域。实验结果表明,笔画方向信息图能够非常准确地反映文字区域的笔画方向信息;基于笔画方向信息图的文字检测方法能快速准确地检测背景环境复杂的文字区域。

参考文献 (References)

- [1] Lyu M R, Song Jiqiang, Cai Min. A comprehensive method for multilingual video text detection, localization, and extraction [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(2): 243-255.
- [2] Ye Qixiang, Gao Wen, Wang Weiqiang. A robust text detection algorithm in images and video frames [C]//Proceedings of the Fourth International Conference on Information, Communications and Signal Processing. Singapore: IEEE Computer Society, 2003: 802-806.
- [3] Hase H, Shinokawa T, Yoneda M. Character string extraction from a color document [C]// Proceedings of the Fifth International Conference on Document Analysis and Recognition. Bangalore, India: IEEE Computer Society, 1999: 75-78.
- [4] Wang Kongqiao, Kangas Jari A. Character location in scene images from digital camera [J]. Pattern Recognition, 2003, 36(10): 2287-2299.
- [5] Ye Qixiang, Huang Qingming, Gao Wen, et al. Fast and robust text detection in images and video frames [J]. Image and Vision Computing, 2005, 23(6): 565-576.
- [6] Kwang In Kim, Keechul Jung, Se Hyun Park, et al. Support vector machine-based text detection in digital video [J]. Pattern Recognition, 2001, 34(2): 527-529.
- [7] Kwang In Kim, Keechul Jung, Jin Hyung Kim. 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.
- [8] Mao Wenge, Chung Fulai. Hybrid Chinese/English text detection in images and video frames [C]// Proceedings of 16th International Conference on Pattern Recognition. Quebec, Canada: IEEE Computer Society, 2002: 1015-1018.
- [9] Ojala Timo, Harwood D. A comparative study of texture measures with classification based on featured distributions [J]. Pattern Recognition, 1996, 29(1): 51-59.
- [10] Timo Ojala. 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] Ye Qixiang. Fast and robust text detection in images and video frames [J]. Image and Vision Computing, 2005, 23(6): 565-576.
- [12] 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.
- [13] Vapnik V N. Statistical Learning Theory [M]. New York: John Wiley & Sons, Inc., 1998.
- [14] Zhong Yu, Zhang Hongjiang. Automatic caption localization in compressed video [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(4): 385-392.
- [15] Lucas S M. ICDAR 2005 text locating competition results [C]// Proceedings of the Eighth International Conference on Document Analysis and Recognition. Seoul, Korea: IEEE Computer Society, 2005: 80-84.