

改进的 Harris 角点文本定位方法^①

汤二贝, 洪留荣, 郑 颖

(淮北师范大学 计算机科学与技术学院, 淮北 235000)

摘 要: 针对计算加权角点饱和度方法在文本定位中对非文本角点提取过密或文本角点提取不足的情况容易错误定位的问题, 通过运用候选区域块的高宽比(外接矩形高度与宽度之比)、面积比(白色区域面积与外接矩形面积之比)数据作为特征, 结合二元线性回归提出了文本定位方程, 解决非文本角点过密的问题. 运用文本区域块的高宽比比面积比的数值通常出现两极分布的特点, 在定位方程中加入常数 C_1 , 用来定位文本, 数值在两极之间时加入常数 C_2 , 定位非文本, 实验结果表明本文算法比计算加权角点饱和度方法取得了更好的效果.

关键词: Harris 算子; 角点检测; 二元线性回归; 文本定位; 定位方程

Improved Harris Corner Text Location

TANG Er-Bei, HONG Liu-Rong, ZHENG Ying

(School of Computer and Technology, Huaibei Normal University, Huaibei 235000, China)

Abstract: The calculation of weighted corner saturation gets the wrong location in text when the corners are over crowding, or the corners are insufficient. This paper solves the overcrowding of non-text corner by using the aspect ratio data (the ratio of external rectangular height and width), area ratio data (the ratio of white area and external rectangular area) of the candidate block as the eigenvalue. This text location equation is put forward with the combination of Binary linear regression. Using the numerical of regional block ratio area of high to width ratio of the text region is usually appear the poles distribution, it employs constant C_1 in the location equation to locate the text. Then, it adds constant C_2 when value between the poles to locate the non-text. The experimental results show that this algorithm has achieved better effect than the method of calculating weighted corner saturation.

Key words: Harris operator; corner detection; Binary linear regression; text location; location equation

1 概述

图像中的文本包含丰富的信息, 文本定位就是将文本从图像中自动定位出来, 对图像高层语义的自动理解、索引和检索有非常重要的意义^[1]. 当前文本定位的方法主要分为四类: 基于边缘、基于纹理、基于连通区域、基于角点. M.Basavanna 等人^[2]通过基于边缘的零交叉点的方法, 该方法对只适用于水平文本的定位; Zhou Yatong^[3]和 S.Nirmala^[4]是基于纹理的文本定位方法, 该类方法具有很好的抗噪性和适用性, 但是对图像的边缘提取有很强的依赖性, 且无法获得高层次的文本信息; Soo-chang Pei^[5]是基于连通区域的定位

方法, 虽然具有较强鲁棒性, 但容易受复杂背景干扰的文本信息; Soo-chang Pei^[5]是基于连通区域的定位方法, 虽然具有较强鲁棒性, 但容易受复杂背景干扰; 基于角点的定位方法分为基于轮廓和基于图像灰度两种^[6], Kitchen^[7]角点检测和 Schmid^[8]角点检测是基于轮廓的角点检测方法, Susan^[9]角点检测和 Harris^[10]角点检测都是基于图像灰度的角点检测方法. Harris 角点检测通过计算像素点所在位置的梯度进行检测, 从重复率和角点周围的局部区域两个方面进行对比^[11]. 文本区域通常都包含比较密集的点, 因此通过对角点

① 基金项目: 安徽省优秀青年人才资金(2011SQRL073)

收稿时间: 2014-06-04; 收到修改稿时间: 2014-06-30

的提取、膨胀可获得面积比较大的区域块,以此定位文本区域,Yong Zhang^[12]通过计算连接分量的加权角点饱和度来定位文本。

本文在使用面积比的同时,利用文本通常都是横向或者纵向排列的特性,高度与宽度的比例呈两极分布的规律,因此也引入了高宽比。以面积比、高宽比为特征值结合二元线性回归^[13],提出了文本定位方程,解决非文本角点过密定位困难的问题;对文本角点检测不足的问题,通过高宽比比面积比数值的计算,对比计算数值与设置的双阈值的大小,分情况加入不同常数对方程进行补充,提高定位精度。

2 基于Harris角点检测定位文本方法分析

2.1 Harris 角点检测简述

Harris 角点检测算法是基于信号的点特征提取的一种算子^[10],其原理是将所处理的图像窗口 w 向任意方向移动微小位移 (x,y) ^[6], x 方向和 y 方向的位移分别为 u 和 v ,其灰度变量定义为^[11]:

$$E_{x,y} = \sum_{x,y} w(x,y) [I(x+u,y+v) - I(x,y)]^2 = \sum_{x,y} w(x,y) [I_x u + I_y v + \alpha(u^2 + v^2)]^2$$

I_x 和 I_y 为一阶灰度梯度。定义 $A = I_x^2 \otimes w(x,y)$,

$B = I_x I_y \otimes w(x,y)$, $C = I_y^2 \otimes w(x,y)$ 和方阵

$$M = \begin{pmatrix} A & C \\ C & B \end{pmatrix}.$$

设 λ_1, λ_2 为方阵 M 的两个特征值, λ_1, λ_2 值的情况决定角点、边缘、和平坦区域^[11]:

1) λ_1, λ_2 的值都比较大,说明局部自相关函数有一个尖峰, $E_{x,y}$ 沿任何方向变化都很大,此处为角点;

2) λ_1, λ_2 的值一大一小,说明局部自相关函数呈现山脊状, $E_{x,y}$ 沿着山脊方向的变化很小,垂直山脊方向的变化很大,此处为边缘;

3) λ_1, λ_2 的值都比较小,说明局部自相关函数比较平坦,此处为平坦区域;

2.2 计算加权角点饱和度文本方法^[12]步骤

计算加权角点饱和度方法主要步骤如下:

1) 对输入图像候选块进行连接分量检测和 Harris 角点检测,得到每个连接分量和角点集。

2) 对每个连接分量计算 X 和 Y 的协方差矩阵:

$$C = \begin{pmatrix} E[x - \mu_x]^2 & E[x - \mu_x][y - \mu_y] \\ E[x - \mu_x][y - \mu_y] & E[y - \mu_y]^2 \end{pmatrix}$$

$E[\bullet]$ 为数学期望, $\mu_x = E(x)$, $\mu_y = E(y)$, $x_i \in X$ 和 $y_i \in Y$ 表示每个像素的横、纵坐标设置,通过对每个连接分量的最小转动惯量轴两次旋转得到最小边界框。

计算骨架和角点的测地距离 $d_A(\Omega \cap A, S)$. $d_A(p, q) = \min \{L(P) \mid p_1 = p, p_l = q, P \subseteq A\}$, A 为像素集, $d_A(p, q)$ 表示 p, q 在 A 中的最小路径长度, $P = (p_1, p_2, \dots, p_l)$ 加上 p 和 q , Ω 为输入图像角点集, S 为骨架集, $d_A(\Omega \cap A, S) = \bigcup_{p \in \Omega \cap A} \{d_A(p, S)\}$.

3) 对每个角点 c_i 定义 $w_i = e^{-\frac{d_i^2}{2\sigma^2}}$, 并计算每个连接分量的加权角点饱和度为:

$$S_w = \frac{\sum_{i=1}^n w_i}{w \cdot h}$$

当 S_w 值较大时,定位为文本。其中 d_i 是角点 c_i 与骨架的测地距离, n 为角点数, w 和 h 分别为连接分量外接矩形的宽度和高度。

以上方法通过对候选区域块单一加权角点饱和度的计算用于定位是否为本,但当文本连接分量角点提取不足时,角点数 n 减小,角点和骨架的测地距离 d_i 可能变大, w_i 减小, $\sum_{i=1}^n w_i$ 减小,角点减少连接分量外接矩形的 w 和 h 值将也变小, S_w 值可能变小,导致文本容易错误定位为非文本。同理一些非文本块连接分量角点过密时,也容易错误定位为文本。



(a) 原始图像



(b) 先后进行 Harris 角点检测、膨胀、形态学处理以及去除小面积块后图像

图 1 计算加权角点饱和度方法的结果

3 改进的Harris角点文本定位方法

3.1 方法思想

为了解决上述的问题, 通过对比图 1(a)和图 1(b), 图中文本区域通常都具有较大的面积比(白色区域面积与外接矩形面积之比)和较小的高宽比(外接矩形高度与宽度之比), 非文本这两个特征则不同, 为此可以用来定位文本和非文本. 同时使用两个特征相互关系的二元线性回归改善了计算加权角点饱和度方法中单一使用外接矩形的宽度 w 和高度 h 的乘积为特征的不足, 二元线性回归通过两个变量的使用, 会对角点过密的非文本区域块面积比数值进行修正, 比一元线性回归^[14]更符合实际.

通过对 ICDAR2003 数据集依次经过 Harris 角点检

测、膨胀、形态学处理以及去除小面积块后, 得到它们的面积比、高宽比的数据, 如图 2, 并将它们代入二元线性回归方程:

$$y = w_1 x_1 + w_2 x_2 + w_3 \quad (1)$$

其中 x_2 为面积比、 x_1 为高宽比, 文本 y 取值 1, 非文本 y 取值 0, 分别求解, 结果保留四位小数, 得到 $w_1 = -0.1347$, $w_2 = 2.3465$, $w_3 = -1.1129$, 代入(1)中, 确定二元线性回归方程.

在将大量、数量接近的文本和非文本的 x_1 、 x_2 代入后, 统计 y 值大致符合[0,1]上的均匀分布, 均值为 0.5, 结合计算(1)时, 候选区域块在定位时 y 取值的两种情况, 因此定义 y 值分类的临界点为 0.5. 归纳在当 $y \geq 0.5$ 时, 定位为文本; $y < 0.5$ 时, 为非文本.

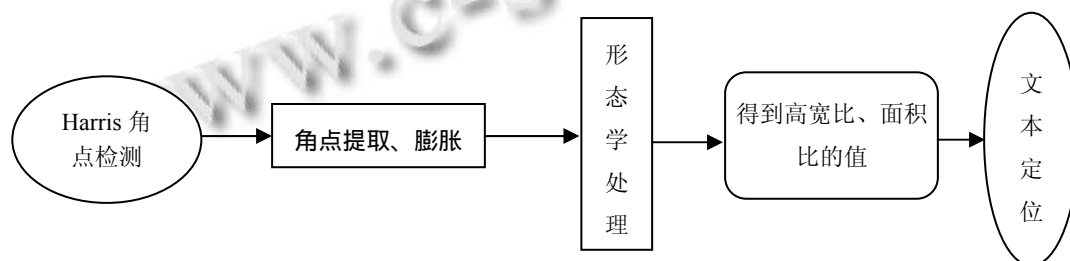


图 2 改进的 Harris 角点检测步骤

$0.5 \leq -0.1347x_1 + 2.3465x_2 - 1.1129$, 定位为文本, 其中 x_1 取非负值. 即:

$-0.1347x_1 + 2.3465x_2 - 1.1129 \leq 2.3465x_2 - 1.1129$, 解得 $x_2 \geq 0.6874$, 因此定位为文本 x_2 的最小值为 0.6874. 当 $x_2 < 0.6874$ 时, 不能对文本区域块正确定位.

为了解决这一问题, 通过对 ICDAR2003 数据集依次经过图 2 步骤处理后的候选区域块高宽比比面积比数据的分析, 文本区域块的数值通常呈两极分布, 利用这一特征可以在方程中加入常数 C_1 , C_1 的大小为定位临界点 0.5, 数值分布在两极之间加入 C_2 , C_2 的大小为 0, 即仍然定位为非文本. 区分进行定位, 完善了二元线性回归方程.

3.2 方程描述

使用 ICDAR2003 数据集经图 2 步骤处理后组成的候选区域块的面积比、高宽比数据, 分别将它们代入

方程(1)中, 结合原始图片中区域块的文本与否, 获得处理后图片中对应区域块的 y 取值的情况, 提出二元线性回归. 当有一些区域块因面积比偏小而可能造成误判时, 利用文本通常具有横向或者纵向的特征, 在对其高宽比比面积比数值分析时, 文本的数值通常呈两极分布, 非文本的分布则在两极之间, 因此可以通过设置两个阈值 ∂_1 , ∂_2 ($\partial_1 < \partial_2$) 的方法来区分. 大量实验证明, 两个阈值分别设为: 0.8610 和 5 时, 定位精度最高. 当 $x_1 / x_2 < 0.8610$ 或者 $x_1 / x_2 \geq 5$ 时, 在二元线性回归式中加入常数 $C_1 = 0.5$, 定位为文本; 当 x_1 / x_2 的值在两个阈值之间时, 为非文本, 式中加常数 $C_2 = 0$, 定位为非文本.

结合 3.1 中确定的二元线性回归方程以及 3.2 中对方程的补充改进, 总结文本定位方程如公式(2).

$$y = \begin{cases} -0.1347x_1 + 2.3465x_2 - 1.1129 & \text{if } x_2 \geq 0.6874 \\ -0.1347x_1 + 2.3465x_2 - 0.6129 & \text{if } x_2 < 0.6874 \text{ or } x_1 / x_2 < 0.8610 \text{ or } x_1 / x_2 \geq 5 \\ -0.1347x_1 + 2.3465x_2 - 1.1129 & \text{if } x_2 < 0.6874, 0.8610 \leq x_1 / x_2 < 5 \end{cases} \quad (2)$$

4 实验结果与分析

为了验证结果的正确性, 本文选取ICDAR2003数据集作为样本进行实验, 并与同样使用此数据集的 Yong Zhang 的实验结果进行了对比, 性能参数对比如表 1.

表 1 性能参数对比

	正确率(%)	召回率(%)
YongZhang	80	76
本文算法	86.7	75.3

通过表 1 性能参数对比: 本文方程在定位的正确率方面有 6%提升, 证明了本文方法的有效性. 就具体实例分析如下:

图 3(a)图为原始图片, (b)图为依次进行图 2 步骤处理后的候选区域块图像, 通过大量实验, 发现去除小面积块设置为 1190 像素时, 效果最佳. 实验中, 分别获取(b)图中六个区域块(分别记为 m_1 到 m_6)的高宽比、面积比数参数值(保留四位小数), 通过(2)式计算相应 y 值, 并与定位的临界值比较, 得到定位结果. m_1 参数 $x_1=0.1738, x_2=0.5527$, 计算 y 值为 $0.6606>0.5$, 定位为文本; m_2 参数 $x_1=3.0588, x_2=0.6352$, $y=-0.0344<0.5$, 定位为非文本; m_3 参数 $x_1=3.0345, x_2=0.7335$, 值 $y=0.1996$, 定位为非文本; m_4 参数 $x_1=0.2477, x_2=0.7292$, 计算的值 $y=0.5648$, 定位为文本; m_5 参数 $x_1=0.1520, x_2=0.5804$, 得 $y=0.7285$, 定位为文本; m_6 参数 $x_1=0.4952, x_2=0.7217$, 计算 $y=0.5139$, 定位为文本. 定位结果与(a)图一致. m_3 结果表明, 本文方法对角点检测过密的非文本候选区域块有良好的定位效果.



(a) 原始图片



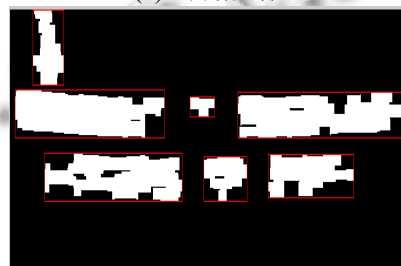
(b) 候选区域块图像

图 3 原始与其候选区域块图像对比(一)

方程对文本角点检测不足的区域块也有很好的定位效果. 如图 4 中第 3 块、第 5 块和第 7 块文本区域块的定位结果. 它们的数值如表 2.



(a) 原始图像

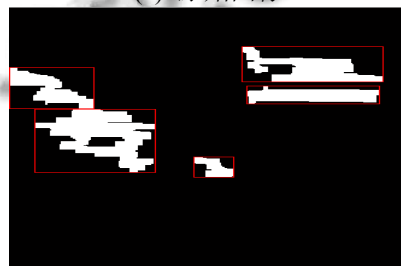


(b) 候选区域块图像

图 4 原始与其候选区域块图像对比(二)



(a) 原始图像



(b) 候选区域块图像

图 5 原始与其候选区域块图像对比(三)

表 2 图 4(b)候选区域块数据及定位结果

x_1	x_2	x_1/x_2	原图是否是文本	y 值	判别是否为文本
0.1339	0.7445	未计算	是	0.6161	是
1.0202	0.6273	1.6263	非	0.2217	非
0.1455	0.6796	0.2141	是	0.9622	是
0.3418	0.7085	未计算	是	0.5036	是
0.4275	0.6535	0.6542	是	0.8629	是
0.1134	0.7596	未计算	是	0.6542	是
0.2132	0.6583	0.3239	是	0.9031	是

但是, 在一些图中会出现误判的情况, 如图 5, 数

值如表 3.

表 3 图 5(b)候选区域块数据以及与图 5(a)的对比

x_1	x_2	x_1/x_2	原图是否 是文本	y 值	判别是否为 文本	是否 误判
0.9706	0.4100	2.3673	非	-0.2815	非	否
1.0621	0.4845	2.1922	非	-0.1191	非	否
1.0426	0.5962	1.7487	非	0.1457	非	否
0.5030	0.4590	1.0959	是	-0.1037	非	误判
0.2767	0.6728	0.4113	是	0.9285	是	否

第四组数据被误判为非文本, 主要是原图角点提取不足, 膨胀又使这些角点未能连接在一起, 形态学开闭运算只使这些角点形成小的像素区域, 并且因为形态学处理造成高宽比变大, 造成误判. 另外, 原图中有一些文本没有被定位出, 主要是这些文本区域角点提取不足或者区域间断造成像素过小, 判为小面积区域而被去除, 造成文本区域的误判.

5 结语

本文结合 Harris 角点检测和二元线性回归来定位场景中的文本. 该方法在保留 Harris 角点检测的准确性与广泛性的同时, 利用高宽比和面积比两个特征的二元线性回归改善单个特征的一元线性回归的不足, 经过大量实验得到了分析, 给出了利用文本的高宽比、面积比以及两者的比值来约束判别方程, 提高了文本定位正确率.

参考文献

- 晋瑾, 平西建, 张涛, 陈明贵. 图像中的文本定位技术研究综述. 计算机应用研, 2007, 24(6): 8–11.
- Basavanna M, Shivakumara P. Multi-oriented text detection in scene images. International Journal of Pattern Recognition and Article Intelligence, 2012, 26(7): 11–19.
- Zhou YT, Jiang H. Video text detection via robust minimax probability machine. International Journal of Digital Content

Technology and its Applications, 2013, 7(1): 679–686.

- Nirmala S, Nagabhushan P. Foreground text segmentation in complex color document images using Gabor filters. SIVip, 2012, 6(1): 669–678.
- Pei S, Chuang YT. Automatic text detection using multi-layer color quantization in complex color images. Proc. of the Intel IEEE International Conference on Multimedia, 2004, 4(2): 190–195.
- 赵万金, 龚声蓉, 刘纯平, 沈项军. 一种自适应的 Harris 角点检测算法. 计算机工程, 2008, 34(10): 212–214.
- Kitchen L, Rosenfeld A. Gray-level corner detection. Pattern Recognition Letters, 1982, 1(2): 95–102.
- Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors. International Journal of Computer Vision, 2000, 37(2): 151–172.
- Stephen M, Michael J. A new approach to low-level image processing. International Journal of Computer Vision, 1997, 23(1): 45–78.
- Harris CG, Stephens MJ. A combined corner and edge detector. Proc. of the 4th Alvey Vision Conference. Manchester, England., 1988.
- 屈喜文. 一种改进的 Harris 角点检测方法. 机电技术, 2012, (4): 40–42.
- Zhang Y. Arbitrarily oriented text detection using geodesic distances between corners and skeletons. 21st International Conference on Pattern Recognition (ICPR 2012). Tsukuba, Japan, November 11–15, 2012. 1896–1899.
- Joshi D. Databased comparison of sparse Bayesian learning and multiple line regression for statistical downscaling of low flow indices. Journal of Hydrology, 2013, 488: 136–149.
- Tabri RV. Testing for normality in linear regression models using regression and scale equivariant estimators. <http://ssrn.com/abstract=2345221>.