

基于 Harris 角点的彩色图像文字检测

刘亚洲, 刘国荣, 王田甲

(西安交通大学 电子与信息工程学院, 陕西 西安 710049)

摘 要: 提出了一种基于 Harris 角点的文字检测算法. 首先, 根据彩色图像和视频中文字区域和背景之间的颜色分量大小的对比, 利用 Harris 角点检测算法, 得到图像的角点分布图; 然后对图像进行滤波, 去除相对孤立的角点; 将角点图像进行二值化, 利用形态学运算将角点聚合形成区域; 对区域进行轮廓跟踪算法, 得到文字区域的初定位图像; 最后对文字区域进行验证, 得到最终结果. 实验证明该算法具有较高的准确性.

关键词: 颜色分量; Harris 角点; 文字检测; 数学形态学; 轮廓跟踪

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2010)10-0136-04

Text Detection in Color Images Based on Harris Corner

LIU Ya-zhou, LIU Guo-rong, WANG Tian-jia

(School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A text detection algorithm based on Harris corner detection is proposed. Firstly, it is very different between the color components' sizes of the text regions and the backgrounds in color images, so make use of the Harris corner detection algorithm to get corners image. After that image filtering is necessary for removing isolated corners. Thirdly, convert corners image to bianry image, use mathematical morphology to connect remaining corners to be regions. Fourthly, find approximate location of text regions for tracing regions' contours. Finally, inspect and verify text regions to get the final result. Experiments show that this algorithm has a more precision compared with other algorithms.

Key words: color components; Harris corner; text detection; mathematical morphology; contour tracing

1 引言

近年来, 图像、音频和视频为主的多媒体信息正在迅速成为信息交流与服务的主流. 传统的数据库检索中采用的基于关键词的检索方式已不能满足人们的需要, 基于内容的多媒体检索成为一个研究的热点. 正确的识别图像、音频和视频中的内容是基于内容的多媒体检索的一个前提. 另外, 在多种应用中图像和视频中的文字检测与识别就显得尤为重要, 而要实现彩色图像和视频中的文字进行精确识别, 必须先实现对文字区域的准确定位.

目前, 图像和视频中文字的检测方法主要有基于边缘的方法、基于纹理的方法、基于区域的方法、基于视频的多帧平均的方法、基于学习的方法等^[1]. 由于图像和视频中的文字含有较丰富的角点

信息, 文中采用 Harris 角点检测算法和数学形态学运算实现对彩色图像和视频中的文字定位.

2 相关理论和概念

2.1 Harris 角点检测

Harris 算子是 C. Harris 和 M. J. Stephens 在 Moravec 算法的基础上提出的一种基于信号的点特征提取算子, 这种算子受信号处理中自相关函数的启发, 给出与自相关系数相联系的矩阵 M , M 矩阵的特征值是自相关函数的一阶曲率, 如果两个曲率值都高, 那么认为该点是角点. 在图像中进行角点检测, 原理为: 如果图像中某一点向任一方向的偏移都会引起灰度的很大变化, 则该点是角点^[2-3].

Harris 角点检测算法的处理过程可以表示如下:

$$\text{Gauss} = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1)$$

$$M = \text{Gauss} \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (2)$$

$$R = \det(M)/\text{tr}(M) \quad (3)$$

其中, I_x 为 x 的梯度, I_y 为 y 方向的梯度, $G(s)$ 为高斯模板, $\det$ 为矩阵的行列式, tr 为矩阵的迹. 当某点的 R 值大于设定的阈值 T 时, 则该点为角点.

Harris 算子是一种有效的特征提取算子^[2-4], 与其他的角点检测算子相比, 具有计算简单; 可以定量提取; 提取点稳定等特点.

2.2 数学形态学

数学形态学是建立在严格数学理论上的一种新兴学科, 是一种新型的数字图像处理方法和理论. 其基本思想是利用一个结构元素去探测一幅图像, 将图像中凡是很好地放入结构元素的位置做标记, 以便得到关于图像的结构信息^[5].

令 $f(x, y)$ 代表输入图像的像素值函数, $b(i, j)$ 代表结构元素函数, D_f 和 D_b 分别表示函数 $f(x, y)$ 和 $b(i, j)$ 的定义域. 作用在图像 f 上的基本运算有: 腐蚀、膨胀、开启和闭合. 主要介绍灰度形态学.

2.2.1 灰度腐蚀

结构元素 $b(i, j)$ 对输入图像 $f(x, y)$ 的腐蚀定义为:

$$(f \ominus b)(x, y) = \min\{f(x+i, y+j) - b(i, j) \mid (x+i, y+j) \in D_f; (i, j) \in D_b\} \quad (4)$$

腐蚀运算是一个点一个点地进行的, 某点的运算结果是它在一个局部范围内的点与结构元素中与之对应的点的像素值之差, 选取其中的最小值. 可见, 经腐蚀运算后, 边缘部位像素值相对大的点的像素值会降低, 从而边缘会向像素值高于相邻区域的区域内收缩.

2.2.2 灰度膨胀

灰度膨胀定义为

$$(f \oplus b)(x, y) = \max\{f(x-i, y-j) + b(i, j) \mid (x-i, y-j) \in D_f; (i, j) \in D_b\} \quad (5)$$

膨胀计算的运算也是一个点一个点地进行的, 计算时涉及它周围点的像素值以及结构元素值, 实际上是局部范围内点与结构元素中对应的像素值之和, 选取其中的最大值. 所以, 膨胀运算后, 边缘得到延伸.

2.2.3 灰度开运算

用结构元素 b 对灰度图像进行先行腐蚀操作, 而后进行膨胀操作即为灰度图像的开运算, 即

$$f \circ b = (f \ominus b) \oplus b \quad (6)$$

2.2.4 灰度闭运算

用结构元素 b 对灰度图像先进行膨胀操作, 而后进行腐蚀操作即为灰度图像的闭运算, 即

$$f \cdot b = (f \oplus b) \ominus b \quad (7)$$

2.3 图像的预处理方法

由于很多图像质量差和实际研究图像的需要, 经常需要对图像进行预处理, 图像预处理方法有图上质量的增强、图像的灰度化、图像的二值化、边缘检测等. 文中主要介绍彩色图像的灰度化.

数字图像大致可分为彩色图像和灰度图像, 文中的文字检测方法是基于彩色图像的, 但文献[6]的方法有的是基于灰度图像的, 需要将彩色图像转化成灰度图像. 灰度处理的方法主要有以下3种:

(1) 最大值法: 使彩色图像的 R 、 G 、 B 值等于其中最大的一个, 即

$$R = G = B = \max(R, G, B) \quad (8)$$

(2) 平均值法: 求出 R 、 G 、 B 值的平均值, 即

$$R = G = B = (R + G + B)/3 \quad (9)$$

(3) 加权平均法: 设原图像为 $f(x, y)$, 图像的 R 、 G 、 B 分量分别用 $f_r(x, y)$, $f_g(x, y)$ 和 $f_b(x, y)$ 来表示, 转化后的灰度图像为 $g(x, y)$, 则

$$g(x, y) = W_r * f_r(x, y) + W_g * f_g(x, y) + W_b * f_b(x, y) \quad (10)$$

其中 W_r 、 W_g 、 W_b 分别为 R 、 G 、 B 的权值.

3 彩色图像角点检测的辅助定理

由于文字周围角点信息相当的丰富, 在进行图像的角点检测时, 尽可能的不要漏检图像中存在的角点. 然而在将彩色图像变换为灰色图像时, 图像中的部分信息会丢失, 这样会影响检测效果.

定理 1 设有一幅彩色图像 $f(x, y)$, 用图像灰度化方法将其转换为灰度图像 $g(x, y)$, 则对变换后的图像 $g(x, y)$ 进行角点检测时, 存在角点的丢失.

证明: 不失一般性, 假设用式(9)将彩色图像转换为灰度图像, 则当图像一幅图像的文字颜色 RGB 值为 (255, 0, 0), 而背景颜色 RGB 值为 (0, 255, 0) 或者是 (0, 0, 255), 对图像进行灰度变换后, 灰度图像的颜色只有一种, 即灰度值 $g(x, y) = 255/3 = 85$, 这种情况下就无法检测出文字的角点.

文字角点的丢失很可能会导致文字区域的漏检. 为减少漏检情况的发生, 文中对彩色图像的 R 、 G 、 B 分量分别进行角点检测, 对检测出的角点求并集.

4 文字定位

基于 Harris 角点检测和数学形态学的基本原理,文中图像和视频中的文字区域检测步骤如下:

- (1) 输入彩色图像或视频帧 I ;
- (2) 对图像 I 进行角点检测;
- (3) 对图像的角点进行滤波,去除噪声;
- (4) 将图像转换为二值的角点图像,并用数学形态学方法将图像链接为区域;
- (5) 提取链接区域的边界,并进行文字的初定位;
- (6) 对得到的文字区域进行验证,得到最终的文字区域.

4.1 图像的角点检测

设输入图像为 F ,要定位图像 F 中的文字,首先要检测到文字周围的角点,文中对图像进行 Harris 角点检测主要有以下几步:

- (1) 提取出图像的 R 、 G 、 B 分量信息,对各分量执行步骤(2)~(5).

- (2) 利用水平、垂直差分算子对图像的每个像素进行滤波求得 I_x 和 I_y ,进而求得矩阵 M 的值为

$$M = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- (3) 利用式(2)对矩阵 M 的四个元素进行高斯平滑滤波,离散二维均值高斯函数如式(1),文中高斯函数中 $\sigma = 0.5$.

- (4) 利用 M 计算出图像每个像素的角点量

$$R = \frac{I_x^2 * I_y^2 - (I_x * I_y)^2}{I_x^2 + I_y^2}$$

为避免分母为零的情况出现,在编程实现时分母加上了一个很小的数值 0.000 001. 设定阈值为 T ,当像素的角点 $R > T$ 时,此像素点为候选角点.

- (5) 进行局部非极大值抑制以获得最终角点.当候选角点的角点量在某邻域内取得局部极大值时才可定位为最终的角点.很明显,阈值 T 的大小和局部极大值邻域的大小都会影响到角点的数量.根据实验,文中选取局部极大值的邻域为 3,阈值 $T = 5000$.

- (6) 对在 R 、 G 、 B 分量上求得的角点集合求并集,即得到整个图像的角点.可以看到,但多数的角点集中在文字区域,但也有部分的角点分散在图像的其他地方.

4.2 角点滤波

文字区域之外也分布着很多的角点,这些角点

成为了文字定位的噪声.为了要去除这些噪声,文中进行角点滤波,步骤为:

- (1) 设图像高度为 H ,图像的宽度为 W , $\text{Corner}(i, j)$ 标志图像的位置 (i, j) 处是否有角点:当 $\text{Corner}(i, j) = 1$ 时,表示 (i, j) 处有角点;当 $\text{Corner}(i, j) = 0$ 时,则 (i, j) 处没有角点.令

$$\text{Count}[i] = \sum_{j=1}^W \text{Corner}(i, j)$$

其中 $0 < i \leq H, 0 < j < W$.

通过大量实验,文中取 $\text{Count}[i] < 4$ 时,将所有 $\text{Corner}(i, j)$ 设置为 0,即去除第 i 行的角点.

- (2) 由于文字角点相对比较密集,则当 (i, j) 处有角点时,在 (i, j) 周围一定范围内一定还存在着其他的角点.设 $\text{Corner}(i, j) = 1, m \times m (m > 1)$ 为是以 (i, j) 为中心的矩阵的大小,令

$$\text{roundcount} = \sum_{p=i-5}^{i+5} \sum_{q=j-5}^{j+5} \text{Corner}(i, j)$$

其中 $5 \leq i \leq H-5, 5 \leq j \leq W-5$.通过实验,文中取 $m = 11, \text{roundcount} < 3$ 时,将以 (i, j) 为中心的 $m \times m$ 矩阵范围内的角点全部去除.

- (3) 按照步骤(2)的做法去除孤立的角点.

4.3 角点连接为区域和定位文字

将去除噪声后的角点图像变为二值图像,具体过程是:先将彩色图像转换为灰度图像,若 $\text{Corner}(i, j) = 1$,则将 (i, j) 处、 (i, j) 处向上 4 个像素内和 (i, j) 向左 4 个像素内的像素值设为 0,即为黑色,其余像素的像素值设置为 255,即为白色.

从角点图像可以看出,角点的分布相对聚集在一起的,数学形态学的膨胀操作可将同一区域的角点聚合在一起,然后在利用形态学的腐蚀操作,对聚合在一起的区域进行细化.文中利用轮廓跟踪算法^[7]对图像中的文字进行定位.

4.4 文字区域的验证

为去除虚假文字区域,可对得到的文字区域进行验证,方法如下:

假设矩形区域为

$$\text{rectangle}(x_0, y_0, x_l, y_l)$$

其中, (x_0, y_0) 为矩形的左上角坐标, (x_l, y_l) 为矩形的右下角的坐标.遍历矩形区域,查找角点数,若矩形区域内的角点数少于某个正整数 N ,则认为该矩形区域为虚假文字区域,应予以去除.通过大量实验,文中取 $N = 9$.

5 实验结果与分析

采用 VC++ 6.0,在 PC 机(PIV 3.0 GHz,内存

512MB,显卡 NVIDIA GeForce FX 5200,显存 128 MB)上实现了文中提出的彩色图像和视频的文字检测算法,文献[5]中提出的基于灰度图像的文字区域的检测算法.文中将基于彩色图像和视频的文字区域检测算法与基于灰度图像的文字检测算法进行了对比,抽取了 135 张文字彩色图像和视频帧进行了实验,表 1 是两种算法的对比结果.

表 1 彩色图像算法和灰度图像算法的对比

算法	文字检测率 (%)	文字区域 检测率(%)	文字区域 误检率(%)
彩色图像算法	97.90	98.53	16.87
灰度图像算法	95.26	96.14	20.32

从表 1 可以看出,彩色图像算法在文字检测率和文字区域检测率方面都有提高,在文字区域误检率方面有明显的降低.

不管是自然场景中的图片或是视频帧的图像,其背景都有复杂背景和简单背景之分.因此,文中对彩色图像算法在两种背景下的文字图像进行了实验,表 2 为实验结果.

表 2 不同背景下的文字检测效果对比

图像类型	文字检测率 (%)	文字区域 检测率(%)	文字区域 误检率(%)
简单背景图像	97.97	99.12	8.87
复杂背景图像	97.79	96.70	25.21

简单背景图像在文字区域检测率和文字区域检测率方面比复杂背景要高些,而在文字区域误检率方面简单背景的误检率要比复杂背景的误检率低的多.

6 结束语

文中根据 Harris 角点检测和数学形态学原理,

提出了一种彩色图像和视频中的文字区域检测算法.通过实验证明,在彩色图像和视频中直接进行文字区域的检测要比将彩色图像转换为灰度图像后再进行文字检测更加的精确.另外,文字区域的误检率还有待降低,在以后的研究中可根据文字的自身特点来进一步去除非文字区域.

参考文献:

[1] 王勇,郑辉,胡德文. 图像和视频中的文字获取技术[J]. 中国图像图形学报, 2004,9(5):532-538.

[2] 冯宇平,戴明. 一种基于角点特征的图像拼接融合算法[J]. 微电子与计算机, 2009,26(7):21-28.

[3] Rolf P Wurtz, Tino Lourens. Corner detection in color images by multiscalecombination of end-stopped cortical cells [M]. Springer Verlag: Lecture Notes in Computer Science, 1997:901-906.

[4] Wang Yuzhu, Yang Dan, Zhang Xiaohong. Robust corner detection based on mult iscale curvature product in B-spline scale space[J]. Acta Automatica Sinica, 2007, 23 (4):414-417.

[5] 范立南,韩晓微,张广渊. 图像处理与模式识别[M]. 北京:科学出版社,2007.

[6] 赵永志. 彩色图像和视频中的文本提取算法研究[D]. 西安:西北工业大学,2007.

[7] 王占全,徐慧. 精通 Visual C++ 数字图像处理技术与工程案例[M]. 北京:人民邮电出版社,2009.

作者简介:

刘亚洲 男,(1983-),硕士研究生.研究方向为图形图像处理.

刘国荣 男,(1957-),副教授.研究方向为人工智能与数理逻辑.

王田甲 女,(1984-),硕士研究生.研究方向为数字图像处理.

(上接第 135 页)

作者简介:

仲元昌 男,(1968-),博士,副教授,研究生导师.研究方向为通信与测控系统、嵌入式智能仪器系统、无线传感器网络、

虚拟仪器.

李 飞 男,(1985-),硕士.研究方向为通信与测控系统.

李彩玲 女,(1983-),硕士,工程师.研究方向为通信与测控系统.