文章编号：2095－0411（2014）01－0047－05

基于自适应阈值 Canny 算子的视频文本定位方法*

贾冬勤，王洪元，程起才
（常州大学 信息科学与工程学院，江苏 常州 213164）

摘要：为实现视频图像中人工文本的快速定位，提出了一种有效的字幕定位方法。该方法首先对视频帧进行灰度变换，去除冗余颜色信息，利用自适应阈值 Canny 算子对文本区域进行边缘检测，再将得到的子图像在水平方向进行投影行定位，最后对行定位结果进行文字区域精确化。结果表明：该方法能避免人为设置阈值不当使得在不同背景情况下产生伪边缘或边缘漏检，能够有效定位文本，不受文字字体、大小等因素的影响。

关键词：文本定位；自适应阈值；Canny 算子

中图分类号：TP 319　　**文献标识码**：A　　**doi**：10.3969/j.issn.2095－0411.2014.01.011

Video Text Localization Method Based on Adaptive Threshold Canny Operator

JIA Dong-qin，WANG Hong-yuan，CHENG Qi-cai
（School of Information Science and Engineering，Changzhou University，Changzhou 213164，China）

Abstract：The paper proposes an effective method for location of the artificial text in video images. This algorithm transmits the grey level of video frame first，and wipes off the information of redundancy colors. Then makes use of adaptive threshold Canny algorithm to detect the boundary of text area. Row location will be in progress in horizontal direction of the sub－image. And accurate the text area of the horizontal direction results. The consequence shows that this method can avoid faults，which will lead to unsatisfactory results of Canny boundary detection，in setting threshold by human. This algorithm can locate the text effectively，ignoring the impact of the typeface and size of the text.

Key words：text localization；adaptive threshold；Canny algorithm

近年来，随着多媒体技术、互联网的飞速发展，视频的容量正以惊人的速度增长。视频中往往包含有重要的文字信息，它们在一定程度上对图像主要内容进行简练描述和说明。如何从视频中快速而准确地定位文本信息的研究具有重大的实际意义[1]。

视频文本一般可以分为两种[2－3]：一种是视频本身存在的并由设备记录下来的文字信息，例如摄像机拍到的路牌名、汽车照片中的车牌照字符、广告牌等，这类文字通常被称为场景文本或背景文本；另一种是在视频制作后期人为加入的，例如新闻标题、对白文本等，这类文字通常成为人工文本。场景文本相对于人工文本难以检测，而且场景文本对视频所表达信息的理解意义不大，所以本文

* **收稿日期**：2013－09－11
基金项目：国家自然科学基金项目资助（61070121）
作者简介：贾冬勤（1988－），女，江苏南通人，硕士生；通讯联系人：王洪元。

只对人工文本进行研究。

文字定位常用方法有基于连通分量的方法[1,4]、基于纹理的方法[5-6]和基于边缘的方法[7]。基于连通分量的方法可以快速地定位到文字的连通区域，缺点是当背景比较复杂的时候容易失败；基于纹理的方法具有一定的通用性，不受图像分辨率高低、文本尺寸大小、字符字体不同的影响，但其计算量偏大、定位精度不够高，还需要结合其他特征进行文本验证；基于边缘检测的方法计算量小，检测效果明显，但是当背景过于复杂，背景线条很多时，对于字符的边缘会造成一些噪声影响。针对以上问题，本文采用自适应阈值 Canny 算子的视频文本定位方法具有计算量小，不受文本尺寸大小影响，避免了传统边缘检测算子对字符边缘的漏检或伪边缘的产生对定位精度的影响。实验证明了该方法的有效性。

1 算法过程

采用自适应阈值 Canny 算子并与投影相结合的视频文本定位方法，具体步骤如图 1 所示，主要思路是：通过自适应阈值的 Canny 算子边缘检测得到边缘子图像，在其基础上利用投影对子图像进行投影行定位，最后对行定位结果进行文字区域精确化。

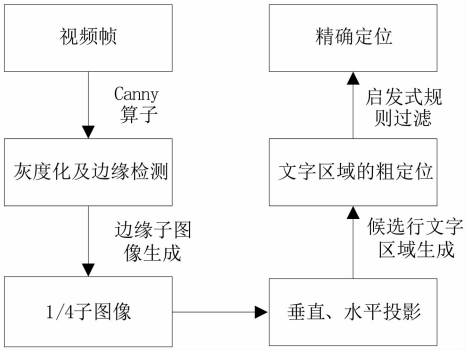


图 1 算法流程图

Fig. 1 Algorithm flow chart

1.1 视频帧灰度变换

提取单个视频帧对图像进行图像灰度变换，目的是降低图像的干扰信息从而突出有用信息，根据三色原理，采用公式（1）计算得到彩色图像中各个像素的灰度值。

$$\mathrm{Gray}(x,y)=0.3*R+0.59*G+0.11*B \tag{1}$$

式中：Gray（x，y）为像素点（x，y）的灰度值，R、G 和 B 分别代表像素（x，y）的 RGB 颜色的红、绿、蓝分量。

1.2 边缘检测

边缘检测是多数图像处理必不可少的一步，是字幕定位提取的基础和前提。与其他边缘检测算子相比，Canny 算子[8]由于其有效、严谨的边缘检测效果而受到广泛的应用。其优点在于：①高定位精度，即精确地把边缘点定位在灰度变化最大的像素上②边缘响应是单值的③低误码率，即非边缘点极少标为真实边缘点。

传统的 Canny 算子对边缘进行检测，由于阈值是人为确定的，往往会检测出假边缘。针对此问题，在 Otsu 算法[9]的基础上，采用一种使用暴力穷举算法的搜索使目标函数取得最佳阈值的方法。下面给出自适应算法的推导和证明[10-11]：将 Canny 算子的最大值点分为 3 类：C_2，C_1，C_0。C_2 类为边缘点集合，C_1 类包含可能是边缘点的集合，C_0 类为非边缘点集合。则有：

$$P_i=\frac{n_i}{N} \tag{2}$$

式中：n_i 为梯度幅值为 i 的像素数量；N 为图像中像素总数；P_i 为梯度幅值为 i 的像素所占的比率。

设 C_0 类中包括梯度幅值为［0，1，2…k］的像素，C_1 类中包括梯度幅值为［$k+1$，$k+2$…m］的像素，C_2 类中包括梯度幅值为［$m+1$，$m+2$…$l-1$］的像素。其中的 $l-1$ 为图像的最大梯度幅值，总像素梯度幅值期望为：

$$\mu_T=\sum_{i=0}^{l-1} ip_i \tag{3}$$

3 类点的对应比率和类内期望为：

$$\begin{cases}\omega_0=\sum\limits_{i=0}^{k} p_i\mu_0=(\sum\limits_{i=0}^{k} ip_i/\omega_0)\\ \omega_1=\sum\limits_{i=k+1}^{m} p_i\mu_1=(\sum\limits_{i=k+1}^{m} ip_i/\omega_1)\\ \omega_2=\sum\limits_{i=m+1}^{l-1} p_i\mu_2=(\sum\limits_{i=m+1}^{l-1} ip_i/\omega_2)\end{cases} \tag{4}$$

式中：C_0 类对应比率和类内期望分别为 ω_0、μ_0；C_1 类对应比率和类内期望分别为 ω_1、μ_1；C_2 类对应比率和类内期望分别为 ω_2、μ_2。

各类的类内方差为：

$$\begin{cases}\sigma_0^2=\dfrac{\sum\limits_{i=0}^{k}(i-\mu_0)^2p_i}{\omega_0}\\ \sigma_1^2=\dfrac{\sum\limits_{i=k+1}^{m}(i-\mu_1)^2p_i}{\omega_1}\\ \sigma_2^2=\dfrac{\sum\limits_{i=m+1}^{l-1}(i-\mu_2)^2p_i}{\omega_2}\end{cases} \tag{5}$$

基于梯度幅值直方图和类内方差最小确定双阈值的评价函数为：

$$\sigma^2(k, m) = \omega_0\sigma_0^2 + \omega_1\sigma_1^2 + \omega_2\sigma_2^2 \quad (6)$$

基于梯度幅值直方图和类间方差最大确定双阈值的评价函数为：

$$\sigma^2(k, m) = \sum_{j=0}^{2}(\mu_j - \mu_T)^2\omega_j \quad (7)$$

若使公式（6）值最小和使公式（7）值最大，则（k，m）为统计意义上的最佳阈值。

算法的步骤如下：①初始化试探阈值：令 $k=1$，$m=k+1$，设 Highthr＝0，Lowthr＝0；②计算 3 类像素点的对应比率、梯度的期望幅值，从而计算评估函数 $\sigma^2(k, m)$；③比较 $\sigma^2(k, m)$ 与 maxVar：如果 $\sigma^2(k, m) >$ maxVar，则令 Highthr＝m，Lowthr＝k，maxVar＝$\sigma^2(k, m)$；④若 m 小于图像的最大梯度阈值减 1，则 m 加 1，返回第②步；⑤若 k 小于图像的最大梯度阈值，则 k 加 1，返回第②步；⑥输出此时 Highthr，Lowthr，即为最佳阈值。

传统 Canny 算子与自适应阈值算法的边缘检测结果如图 2 所示。

(a) 原始灰度图像

(b) 低阈值35，高阈值110

(c) 低阈值180，高阈值255

(d) 自适应阈值算法

图 2　边缘检测结果

Fig. 2　Results of edge detection

图 2 中（b）、（c）为人工选取阈值。候选边缘点的梯度值大于高阈值的点作为边缘保留，梯度值小于低阈值的点则删除，梯度值介于两阈值间且与边缘点邻接的点作为边缘保留。由结果可见，当人工选取阈值过低时，背景信息的边缘也被检测出来；阈值选择过高时，字幕边缘信息将会丢失；自适应算法能够较好的检测出我们所需定位的文本边缘。

1.3　文本区域定位

1.3.1　文本行定位

自适应阈值算法得到的边缘子图像为二值化图像，对二值化图像进行自上向下的方式进行扫描，计算每行的白点数，创建一幅与二值化图像相同大小的背景为黑色的图像，将计算得到的每行的白点数的总值赋给新的图像，结果如图 3（b）所示。

设置 y_b 和 y_e 为水平投影的起始行和结束行，文字区域的高度值为 N（$N=y_e-y_b+1$），由于字幕的大小有限制，当字符的高度小于 8 个像素时，人的眼睛已经难以识别视频中的文本，所以 $N\geqslant 8$，当检测到 y 行的白点数的总值为 0 时，计算文字区域的高度值 N_1，然后继续寻找新的 y_b、y_e，计算新文字区域的高度值 N_2，比较 N_1，N_2 的大小，通常对新闻内容高度概括的文字与其他文本相比具有更多的像素和，故选择大的 N 值作为最终文本行高度。当 $N_1=N_2$（值相等），将其合并作为最终的文字区域。根据图 3（b）信息得到的行定位结果 $N_1=34$、$N_2=20$，候选文字区域的初始行和结束行为：283～317。粗定位到的文本行由图 3（c）所示。

1.3.2　文本列定位

利用相同的投影方法对子图像进行列定位，垂直投影如图 3（d）所示。这些区域，有的是文字区域，要将其合并；有的非文字区域，要将其过滤掉。采用启发式规则对候选列文本区域进行过滤与合并以生成文本区域：若白点像素总值不为 0 且 x 列之间相近则将其合并为文本区域，否则视为非文本区域将其滤掉；若在两段文字列之间如有间隔，但在一定像素之内则需要将其合并（需要考虑到文本区标点符号的存在）。则文字起始列和结束列为：。结合行定位和列定位信息则可将文字区域定位，行号：283～317，列号：97～451。文本区域精确定位如图 3（e）所示。

2　实验结果

下面对视频帧做对比试验，将图 2（b）和图 2（c）粗定位的文本行进行列投影，比较人工选取阈值的方法能否准确定位到文本区域，如图 4；并且运用自适应阈值 Canny 算子边缘检测的方法又进行了一组实验来验证本文方法的有效性，如图 5。

(a)原始灰度图

(b) 二值化图像的行投影

(c) 行定位得到的子图像

(d) 垂直投影

(e) 定位结果

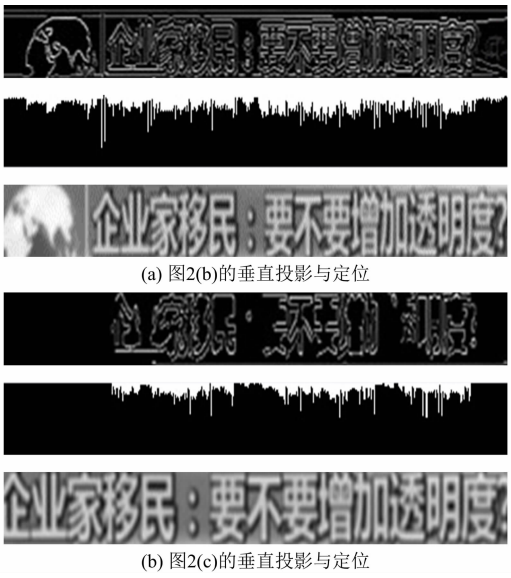
图 3　文本定位

Fig. 3　Text location

图 4（a）的文字列号：0～482，即便可以用规则将左右边界小于 10 像素的候选区域去除，但是投影区域非文本区域和文本区域粘连严重，无法准确选择文本区域；图 4（b）的文字列号：100～449，而文本区域的精确定位列号：97～451。可见，人工选取阈值的方法多少都存在误差，自适应阈值法能更加有效地消除了背景对文本区域字幕检测的干扰，也避免了边缘漏检。

图 5 中选择了一幅大小为 590×420 像素的灰度图，选择图像行号：314～419 处进行文本区域投影定位，根据图 5（c）、图 5（d）检测到两个文本区域，由字符的宽高比规则，得出文字区域行号：374～394，列号：188～403。结果如图 5（e）所示。

本文对一系列视频帧进行实验，所选取的测试视频帧在不同的背景下（不同背景、不同字体、不同大小、不同字体颜色等）进行测试。实验中采用查全率 R 和查准率 P 两个经典的评判指标来检测系统对文本区域定位的性能。



(a) 图2(b)的垂直投影与定位

(b) 图2(c)的垂直投影与定位

图 4　投影对比

Fig. 4　Projection contrast

(a)灰度图

(b)自适应边缘检测

(c)水平投影

(d)垂直投影

(e)定位结果

图 5　文本定位

Fig. 5　Subtitle position

$$R=\frac{n_c}{n_c+n_m}100\%,\quad P=\frac{n_c}{n_c+n_f}100\%$$

式中：n_c 准确检测到的文本区域的个数；n_m 为文本存在却漏检的文本区域的个数；n_f 表示非文本

区域却被误认为是文本区域的个数。

在相同测试集上执行传统 Canny 算子边缘检测的方法，利用 2 个标准参量查全率和查准率进行比较，比较结果见表 1。

表 1 算法性能比较 %

Table 1 Performance evaluation of algorithms %

方法	查全率	查准率
本文方法	89.1	85.3
传统方法	78.0	69.5

3 结 论

本文采用自适应阈值 Canny 算子对视频帧进行边缘检测，能够有效的消除噪声和背景干扰，同时还可以避免人为选取阈值造成的伪边缘及漏检。本文文字区域定位方法的查全率和查准率都优于传统方法。视频字幕定位之后，将继续研究字幕的分割及识别，视频文字检索系统的产生与应用是下一步的目标。

参考文献：

[1] 叶静. 基于内容的新闻视频检索 [D]. 上海：上海大学，2001.

[2] Zhang Dong - qing，Chang Shih - fu. Learning to detect scene text using a higher - order MRF with belief propagation [C] // Computer Vision and Pattern Recognition Workshops. Washington D C：IEEE Computer Society，2004.

[3] LI Chuang，Ding Xiao - qing，Wu You - shou. Automatic text location in natural scene images [C] // Proc of ICDAR. Seattle：IEEE Computer Society，2001.

[4] 蔡波，周洞汝，胡宏斌. 数字视频中字幕检测及提取的研究和实现 [J]. 计算机辅助设计与图形学学报，2003，15 (7)：898 - 903.

[5] Wu V，Manmatha R，Riseman E M. Text finder：An automatic system to detect and recognize text in images [J]. IEEE Trans on Pattern Analysis and Machine Intelligence，1999，20 (11)：1224 - 1229.

[6] Jung K. Neural network - based text location in color images [J]. Pattern Recognition Letters，2001，22 (14)：1503 - 1515.

[7] Chen Da - tong，Odobez Jean - Marc，Bourlard H. Text detect and recognition in images and video frames [J]. Pattern Recognition，2004 (37)：595 - 608.

[8] 冈萨雷斯. 数字图像处理 [M]. 2 版. 北京：电子工业出版社，2007：463 - 491.

[9] Otsu N. A threshold selection method from gray - level histogram [J]. IEEE Transactions on system Man Cybernetics (S1083 - 4419)，1979，9 (1)：62 - 66.

[10] 李牧，闫继红，李戈，等. 自适应 Canny 算子边缘检测技术 [J]. 哈尔滨工程技术学报，2007，28 (9)：1002 - 1007.

[11] 唐露露，张启灿，胡松. 一种自适应阈值的 Canny 边缘检测算法 [J]. 光电工程，2011. 38 (05)：127 - 132.