

基于文字穿越线和笔画连通性的 视频文字提取方法

田破荒,彭天强,李粥程

(信息工程大学信息工程学院,河南郑州 450002)

摘 要: 视频文字自动提取对提高视频内容的整体理解程度具有重要意义. 在复杂背景下,现有提取方法难以稳健地定位文字和有效抑制背景干扰. 我们提出一种复杂背景下稳健的视频文字提取方法,首先利用小波变换和非监督聚类方法获得候选文字区域;其次,引入穿越线及其幅度谱概念,通过提取穿越线特征进行聚类定位单行文字;最后,充分考虑到文字笔画的连通性,引入重线概念,通过检测重线获取文字像素关键点,从关键点出发进行区域生长,结合启发式规则,完成二值化过程. 实验表明,该方法能够在复杂背景下得到较理想的文字二值化图像,不受文字颜色、字体、大小等因素的影响,性能优于文献[7]和文献[8]提出的方法.

关键词: 小波变换;最小最大聚类;穿越线;重线;笔画连通性

中图分类号: TP391.43 **文献标识码:** A **文章编号:** 0372-2112 (2009) 01-0072-07

An Approach for Video-text Extraction Based on Text Traversing Line and Stroke Connectivity

TIAN Po-huang, PENG Tian-qiang, LI Bi-cheng

(Zhengzhou Information Science and Technology Institute, Zhengzhou, Henan 450002, China)

Abstract: Automatic video-text extraction is an important field in video content comprehension. In complex background, conventional extraction methods can not robustly locate text nor eliminate interference from background. A robust video-text extraction method is proposed in this letter. First, an unsupervised paradigm based on wavelet is applied to obtain candidate text region. Second, traversing line and its aptitude spectrum is introduced, and text refinement is carried out by extracting traversing line features. At the end, according to the stroke connectivity, MFVL is introduced and applied to obtain key-points in strokes, from those key-points region growth and geometrical restriction is performed to obtain binary text image. Experiment demonstrates that our scheme can extract text with different color, font, size in complex image, and is better than method in Refs. [7,8].

Key words: wavelet transform; min-max clustering principle; traversing line; heavy line; stroke connectivity

1 引言

随着多媒体技术和 Internet 的飞速发展,海量的视频出现在数字图书馆、电视广播和互联网上,随之带来的问题是,用户需要有效的手段迅速准确地从海量视频中找到自己感兴趣的内容^[1]. 着眼于视频流分析和管理,视频检索技术,目前已经成为研究热点. 视频中的文字能够提供很好的高层语义信息,有助于提高对视频内容的整体理解程度. 因此,视频文字提取和识别成为视频检索技术的重要内容.

视频文字大致分为两类:场景文字和叠加文字. 前者指在视频拍摄过程中产生的文字,它是原始画面的一部分,出现随机,通常并不包含视频内容信息,比如商店

招牌、交通标志等;后者指在视频制作过程中人工加入的文字,是对视频内容的注释,为视频索引和检索提供了语义线索,比如标题条、人物名字等. 为获得视频文字信息,需要经过文字提取和识别两个步骤,其中,用于文字识别的商用 OCR 技术已经成熟. 作为 OCR 的前端,文字提取的性能对 OCR 识别率有着重大的影响,然而它却由于面临着图像低分辨率、未知的文字颜色、大小、位置、复杂背景,以及由于压缩带来的图像受损等诸多困难,一直没有达到实用水平.

大体上,视频文字提取可分为两个步骤:检测和二值化. 检测是判断图像中是否存在文字并定位文字区域;二值化是根据检测结果获得干净的二值文字图像. 目前,依据所采用的特征来划分,文字检测方法主要有

三类:基于边缘的方法、基于区域的方法和基于纹理的方法。

基于边缘的方法主要利用文字区域和背景具有明显的对比度差异。Lienhart 等根据在多尺度下检测的边缘信息,利用多层神经网络来定位文字^[2]。Agnihotri 等^[3]利用边缘信息初步确定文字区域候选,然后利用连通区域方法确定文字块。该方法虽计算简单,但在文字和背景对比度相差不大的情况下检测效果并不理想。

基于区域的方法根据文字具有相近的颜色,采用颜色聚类、连通体分析、启发式规则等方法区别文字与背景。Jain 等^[4]人通过量化整个颜色空间进行颜色聚类,把图像分解成目标图像和背景图像,然后在每个子图利用连通区域的方法确定文字区域。张佑生等^[5]提出了一种基于子图像变异灰度直方图(VCH)的视频文本检测与定位方法,该方法通过对行分割子图像 VCH 中的凸台和列分割子图像 VCH 中的凹谷的识别与定位,实现在图像中快速检测与定位文本。

基于纹理的方法把文字看成一种特殊的纹理,通常将整幅图像分块,然后对图像块进行纹理特征提取,接着利用神经网络、支持向量基等分类器判决该图像块属于文字还是背景。Mao^[6]提出利用小波变换提取像素的局部能量特征进行文字检测。Glavata 等^[7]对图像进行小波分解,提取高频小波系数分布特征来区分文字和非文字区域。

二值化方法主要有基于颜色的方法和基于笔画特征的方法。前者认为文字像素和背景像素在颜色上存在明显差异,因此可以采用全局或者局部阈值进行分割^[8]。而后者利用特定的滤波器得到类似于笔画的像素,例如不对称滤波器^[9]、四向文字提取滤波器^[10],并对非笔画像素进行抑制,然而此类方法对于相交的笔画会产生误判。

通过以上的分析,我们认为,现有文字提取方法存在以下两个方面的不足。第一,现有的方法都假设文字检测模块能非常准确地定位单行或单个文字,在上面提到的各种因素制约下难以做到。事实上,目前普遍使用的基于投影的定位方法过分依赖阈值,鲁棒性差,在复杂背景下,容易丢失文字信息。第二,现有的二值化方法对文字定位精度要求过高,且难以去除具有类似文字颜色或者笔画的背景区域,另外没有考虑不同语种文字的结构差异对二值化性能的影响,难以适应多语种环境。针对以上两个问题,本文提出了一种基于文字穿越线和笔画连通性的视频文字提取方法。在文字检测阶段,利用小波变换和非监督聚类方法来获得候选文字区域,进而提取文字穿越线特征来定位单行文字;在二值化阶段,充分考虑到文字笔画的连通性,引

入重线概念,通过检测重线获取文字像素关键点,进而进行区域生长完成二值化过程。由于水平排列的文字占总视频文字量的 99% 以上^[2],本文只针对水平排列的文字。实验表明,该方法性能稳定,能够在复杂背景下得到较理想的文字二值化图像,优于文献[7]和文献[8]的方法。

2 基于文字穿越线和笔画连通性的视频文字提取方法

视频文字提取属于模式识别的范畴,实际上就是要克服各种不利因素,将文字像素和背景像素在某种最优准则下进行分类,本文提出了一种稳健的基于文字穿越线和笔画连通性的视频文字提取方法,其算法

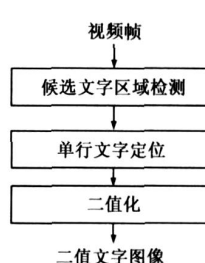


图1 文字提取算法流程图

框图如图 1 所示。将视频帧从 RGB 空间转换到 YUV 空间,提取其中的 Y 分量,获得灰度图。首先,做一层分解的 Haar 小波变换,对 HL 频带的小波系数进行分块处理,从每一个子块上提取特征进行聚类,得到候选文字区域;其次,提取文字区域穿越线特征做聚类,区分文字穿越线和背景穿越线,用连续的文字穿越线组成单行文字;最后,寻找重线,获得文字像素关键点,根据笔画连通性,做区域生长,完成文字区域的二值化。

2.1 候选文字区域检测

小波变换系数可表达纹理的频率特性,这里采用二维 Haar 小波进行一层变换,将图像分解为一个低频区(LL)和三个高频区(HL、LH、HH)。为更好地研究和比较文字区域和背景区域在纹理上的区别,将变换图像的高频区域切分为若干 $M \times N$ (例如 16×8) 的子块,分析各个子块的系数直方图,可以发现,文字子块的系数分布范围较大,较为稀疏,背景区则相反。考虑到当文字水平排列时,文字区块的高频小波系数中 HL 区域系数的分布的离散程度最强,本文使用 HL 高频区的小波系数,从中提取两个参数:(1) 系数方差;(2) 系数范围,将两个参数归一化,组成二维特征向量,运用最小最大聚类方法^[11]区分文字和背景。为了适应复杂背景,将子块聚为三类,分别为:背景子块、复杂背景子块和文字子块。经过聚类,可以得到文字子块标记图。考虑到只检测水平排列的文字,认为文字区域长度应当不小于两个子块长度,因此去除在水平方向上没有相邻子块的子块。最后,采用四邻域法则(如果两个矩形块有一条边是重叠的,则认为二者连通)做连通分析,组合检测到的可能的文字子块。为文字连通区建立外接矩形并将外接矩形向四周各扩展向 5 个像素,以矩形划定的区域作为候选文字区域。

2.2 单行文字区域定位

由于假设文字是水平排列的,水平投影便成为一种有效的行分割方法,但是在复杂背景下,这类依赖阈值的分割方法的效果比较差.本文引入穿越线及其幅度谱的概念,利用文字区域中穿越线特征定位文字.

给定一幅灰度图像,令 $X = \{x_{mn}\}$ 表示其中一大小为 $M \times N$ 的矩形区域. X 可以被看作由 M 行像素依次排列组成,称这些像素行为穿越线,记为 $F_k, 0 \leq k \leq M-1$. 对于图像中含有文字的区域,其穿越线一部分穿越了文字行,另一部分只穿越了背景区,穿越线的灰度变化特性可以用一个二维图形来描述,它以穿越线上像素点在原图像中的列号为横坐标,由一系列垂直于横坐标的线段(这些线段的长度分别对应穿越线上各像素点灰度值)构成,称此图形为穿越线幅度谱.

图2给出了一幅图像中含有文字区域的不同类型穿越线的幅度谱.文字穿越线的幅度谱趋于周期性地剧烈上下波动,呈梳状,反映出文字穿越线上文字像素和背景像素依次交替出现的特点,如图2(a)所示;而穿越背景区的穿越线的幅度谱波动相对比较随机,轮廓相对平滑,表现了背景区域灰度变化相对缓慢的特点,如图2(b)所示.尽管这种近似并不在所有的情况下都成立,但是仍然可以通过它来区分文字穿越线和背景穿越线.事实上,这一特征与文字区较背景区,纹理更为丰富,边缘更为强烈的假设是一致的.为了区分文字穿越线和背景穿越线,用差分幅度谱的绝对值的标准差来度量二者的区别.

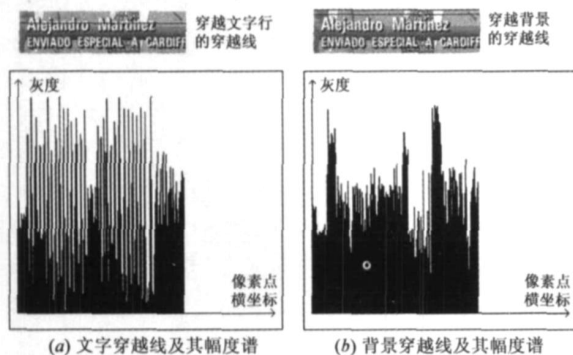


图2 穿越线与幅度谱

设穿越线 F_k 上像素点的灰度值为 $g_{k,i}, 0 \leq i \leq N-1$, 差分幅度谱的标准方差定义如下:

$$stvGS = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (\Delta g_{k,i} - \mu_k)^2} \quad (1)$$

其中 $\Delta g_{k,i} = |g_{k,i} - g_{k,i-1}|$, $\mu_k = \frac{1}{N-1} \sum_{i=1}^{N-1} \Delta g_{k,i}$. 通过大量统计实验发现,梳状特征越明显,则 $stvGS$ 越大,穿越线的穿越文字的可能性越大.

基于以上分析,这里利用最小最大聚类方法来区

分备选文字区域的文字穿越线和背景穿越线.认为连续8条以上的文字穿越线组成有效的单行文字区域.

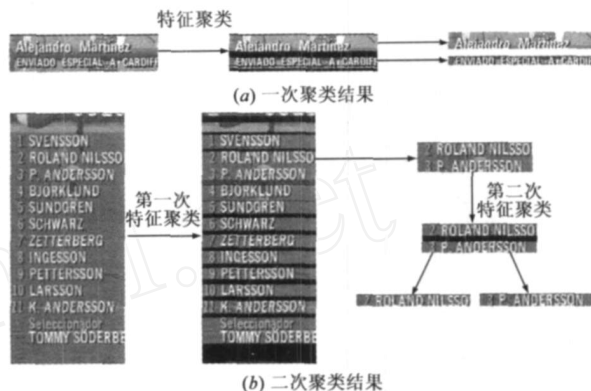


图3 穿越线聚类效果图

在大多数情况下,候选文字区域包含的文字行较少,其高度都小于某个阈值,把穿越线聚为两类便可以很好地区分文字穿越线和背景穿越线, (图3(a),背景穿越线被置为黑色).但是对于高度较大的文字区域,由于纵向背景变化比较大,若聚为两类,部分文字会穿越线被划为背景穿越线.本文采用二次聚类的方法,获得了比较满意的划分结果(如图3(b),背景穿越线被置为黑色).步骤如下:

(1) 根据文字区域高度确定划分的类别数 k : 高度小于50个像素的区域聚为两类,以后高度每增加若干个像素(例如90),聚类数就增加1.以特征值最小的一类为背景穿越线.这样的聚类使得无法把所有的文字行分开.

(2) 设一次聚类中划分出的有效文字行的最小高度为 $H_{\min}$, 对高度大于两倍的 $H_{\min}$ 区域做 $k=2$ 的最小最大聚类.

经过两次聚类,可以分开不同的文字行.

从几幅效果图中可以看出:每个划分出的区域只含有一个文字行,但是一部分结果包含了太多的背景区域(图8(d)),另一部分并没有把所有的文字像素都包含进去(图9(d)).事实上,这样的结果完全可以满足后续的二值化算法的要求.

2.3 二值化

传统的二值化方法都建立在对文字精确定位的基础上,但是在复杂背景下,难以准确定位文字,除此之外,它们还难以去除类似文字颜色或者纹理的背景.

深入观察,大部分文字存在以下特点:(1)文字和背景在交界处存在明显的颜色变化;(2)文字内部颜色相同或相近;(3)同行文字颜色变化不大.基于这些特点和上面介绍过的文字穿越线,对于一个给定的文字区域,称其中 $stvGS$ 值最大的穿越线为重线(如图4中白线标注的扫描行).因为文字穿越线的 $stvGS$ 值大于

背景穿越线,从而重线一定为文字穿越线,大量实验证明重线穿越文字行的概率极高。

由于复杂背景下视频文字多为亮色,如图 4 所示,重线由若干交替出现的高灰度的文字像素游程和低灰度的背景像素游程组成。重线的幅度谱从灰度变化的角度表现了这种交替现象。可以注意到,重线幅度谱呈现明显的梳状特征:随着横坐标的变化,幅度由高(低)到低(高)迅速交替,且高低幅度的落差十分明显。其中,谱中的峰对应于文字像素游程,谷对应于背景像素游程。通过设置灰度阈值,可以方便地将获得文字、背景像素游程。根据同行文字灰度变化不大的假设,这样的划分不会遗漏文字游程。称各个文字像素游程中灰度值最高的像素点为关键点(图 4 中红点标注的像素点),显然关键点属于文字像素点。

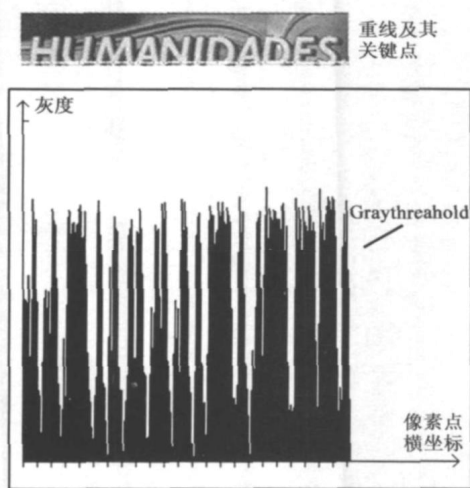


图4 重线和关键点及对应的幅度谱

图 5 给出了该方法的流程:首先确定单行文字的重线位置,获取在笔画与重线的交叉处的文字像素关键点,然后从关键点出发在全图范围内做区域生长,获得文字笔画,最后,应用启发式规则滤除掉干扰背景,得到最终的文字二值图像。考虑到不同语种在笔画连通性上的差异,分别针对英文和汉语做了不同处理。该方法从文字像素自身出发,很大程度上削弱了复杂背景对文字二值化的影响,并能够在粗略的单行文字定位下获得理想的结果。

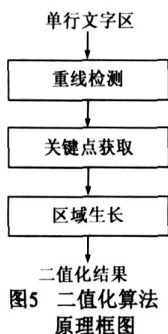


图5 二值化算法原理框图

2.3.1 英文二值化

英文字母由单个或多个连通的笔画构成。根据文字内部像素点颜色相同或相近的假设,各个英文字母的所有像素点根据相同或相近的颜色形成一个连通体。如图 4,经过定位重线并分析其幅度谱,每个字母都获得一个(字母“U”)或几个相邻的关键点(字母“H”)。以字母“U”为例,设字

母“U”的所有像素点构成集合 X ,处于“U”边界处的背景像素点构成集合 Y ,显然,在图像中,集合 Y 的元素包围了所有集合 X 中的元素。进一步设属于“U”的关键点为 $p, p \in X$,集合 X 中所有元素与 p 的颜色距离小于 ε ,集合 Y 中所有元素与 p 的颜色距离大于 δ ;根据文字内部颜色相同或变化很小,而文字边界存在明显颜色变化的假设,有 $\varepsilon < \delta$ 。从这些关键点 p 出发,设定颜色阈值 t ,使得 $\varepsilon < t < \delta$,在全图范围内做区域生长^[8]。由于 X 中的元素(包括 p)相互连通且与 p 的颜色距离小于 t , X 中的所有元素,即所有“U”的像素点,被并入关键点所生成的区域。集合 Y 的元素与 p 的颜色距离大于 t , Y 中的元素,即“U”的边界点不会被包括。又由于在图像中, Y 包围了 X ,所有其他的背景像素点不论如何变化,都不会被包括,从而获得干净的“U”的二值化图像。这个过程可以分别对每个字母做二值化。在实验中,我们在 RGB 空间实现二值化过程, t 取固定值。需要说明的是,在关键点确定后,由于区域生长在全图范围内执行,所以本方法不受文字定位结果的束缚,在定位结果不理想的情况下,仍然可以做有效的二值化。

由于背景复杂,重线穿越可能会穿越亮色背景区域,这导致灰度阈值设置后,出现对应于这些亮色背景的高灰度游程,及其关键点。经过区域生长,这些亮色背景区域也被提取出来。但是这些背景关键点所生成的区域往往很大。本文设定以下启发式规则去除亮色背景干扰,获得干净的文字:

- (1) 字母所占像素数不得超过 600;
- (2) 设当前文字行内的各个字母区域的高度平均值为 $\bar{w}_H$, 标准差为 $D(W_H)$, 认为高度大于 2 倍 $D(W_H)$ 的区域为非文字区;
- (3) 设当前文字行内的各个字母区域的宽度平均值为 $\bar{w}_W$, 标准差为 $D(W_W)$, 认为宽度大于 2 倍 $D(W_W)$ 的区域为非文字区;

2.3.2 汉字二值化

事实上,上述二值化过程获得的是被重线穿越的笔画连通体,适于提取文字笔画相互连通的文字,例如英文。而汉字则不同,其构成可以被描述为:相互连通的笔画构成若干笔画连通体,再由互不相连的笔画连通体组成单个汉字。这种特殊的结构导致重线虽然可以穿越每个汉字,却无法穿越所有的笔画连通体,从而上述二值化法只能获得单行汉字中被重线穿越的那部分笔画连通体,组成不完整的二值化结果(例如,图 6(a)中的“精”由“米”、“主”、“月”三个笔画连通体构成,图 6(c)英文二值化方法只提取了被重线穿越的“米”和“主”,丢失了“月”)。

考虑到同行汉字的高度宽度都相当,且汉字笔画丰富,文字行中总是存在某些被重线穿越的笔画连通

体抵达文字行的上沿、下沿和左右两端. 显然, 不完整的二值化结果提供了文字行的精确定位. 从精确定位区域的顶部到底部, 每隔若干像素(例如 2 个像素)设置一条穿越线, 保证所有的笔画连通体都被穿越. 分析所有穿越线的幅度谱, 获得关键点. 显然, 这些关键点覆盖了所有笔画连通体. 它们可以分为三类: (1) 属于已获得笔画连通体的关键点; (2) 属于已确认非文字亮色区域的关键点; (3) 新关键点. 从新关键点出发, 在全图范围内做二次区域生长, 可以获得剩余的笔画连通体. 因为可能存在某些新关键点属于亮色背景, 我们将两次区域生长的结果合并, 再次使用上述启发式算法, 滤除背景区, 获得最终结果(如图 6(d)).

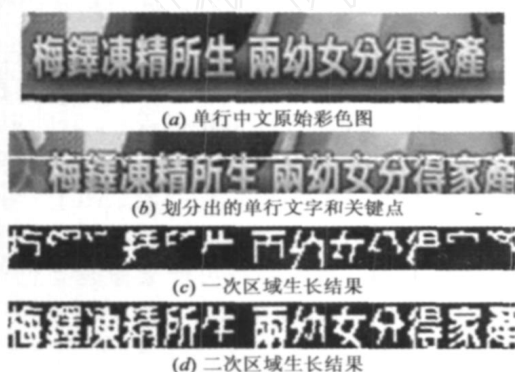


图6 中文二值化过程示例图

3 实验结果与性能比较

算法测试图像来自于美国 CNN 新闻、台湾东升新闻、新闻连播以及一些流行的测试图像, 共 100 幅图像, 包含 596 个汉字, 1903 个西文字母; 图像分辨率从 320 × 240 到 352 × 480 不等.

由于文本检测和二值化的评价标准不同, 实验分检测和二值化两个部分.

3.1 文字检测算法结果

根据本文二值化方法的特殊性, 采用新标准判定文字区域、单个文字是否被检测到: 如果定位区域中只含有一行文字, 且存在三条以上穿越线穿越了当前该文字区域(人工划定的理想单行文字区域) 80% 以上的文字(单词), 则认为该文字区域检测定位正确; 如果算法定位的单行文字区域中存在一条以上穿越线穿越了某个文字(汉字或单词), 则认为该文字正确定位. 算法性能评价标准包括文字区域定位准确率、文字区域定位查全率和文字定位查全率.

本文在上面提到的测试集上做测试, 并与文献[7]中基于纹理的检测方法、文献[8]中基于边缘特征的方法进行了性能比较. 表 1 给出了性能比较结果.

可见, 本文方法的文字区域定位准确率达到 91.23%, 查全率达到了 88.14%, 文字定位查全率达到

了 86.33%, 从查全率和查准率两方面都优于文献[7]和文献[8]中的方法. 图 7 给出了部分检测定位效果比较图.

表 1 文本检测性能比较

	文献[7]方法	文献[8]方法	本文方法
文字区域定位准确率	89.02 %	90.22 %	91.23 %
文字区域定位查全率	85.61 %	82.13 %	88.14 %
文字定位查全率	81.21 %	80.94 %	86.33 %



(a) 本文方法

(b) 文献[7]的方法

图7 部分文字检测定位实验结果

3.2 文字二值化算法结果

重线能否穿越文字是文字提取成功的关键, 称重线穿越文字行的事件为重线命中. 实验共测试了被正确定位的 139 行具有复杂背景的单文字行, 重线命中 138 次, 仅失败 1 次. 命中率为 99.26%.

在重线命中的前提下, 使用错误概率 (PE) 评价文本二值化算法^[12]. 错误概率定义如下:

$$PE = P(O) \times P(B|O) + P(B) \times P(O|B) \quad (2)$$

其中 $P(B|O)$ 表示把文本像素划分为背景像素的概率, $P(B|O)$ 表示把文本像素划分为文本像素的概率. $P(O)$ 和 $P(B)$ 分别表示人工确定的文本像素和背景像素的先验概率. 参与性能比较测试的二值化方法包括相关熵二值化法^[13] (下面称方法 1) 和文献[7]的方法 (下面称方法 2). 其中, 方法 1 是经典的全局二值化方法; 方法 2 是近期提出的二值化算法, 该法综合纹理、颜色信息. 测试本文二值化方法的图像来自于前端的定位结果, 而方法 1 和方法 2 的测试图像为相应的人工划定的理想文字图像. 表 2 列出了性能比较结果. 该结果表明本文方法具有明显的优势.

表 2 文字提取错误概率

	本文的方法	方法 1	方法 2
错误概率	0.121	0.421	0.354

图 8~11 列出了四组二值化性能比较图,每组图 (a)~(f) 子图依次为人工定位的原始彩色文字图,方法 1 效果图,方法 2 效果图,本文文字检测结果及重线、关键点,本文二值化结果,理想二值化结果。四幅测试图涵盖了纹理丰富的背景(图 8(a),图 9(a))、具有文字颜色的背景(图 10(a),图 11(a))以及不同大小不同

字体的文字。可以看到,法 1 效果都很差,显然无法应对复杂背景。法 2 把大量具有纹理丰富的或亮色的背景区域的像素点划为文字像素点。而新方法则显著抑制了复杂背景的干扰,其二值化结果都大大优于法 1 和法 2。



图8 提取效果比较图组A



图9 提取效果比较图组B

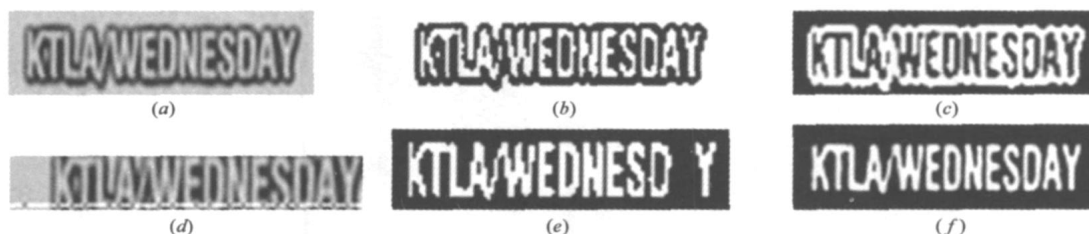


图10 提取效果比较图组C

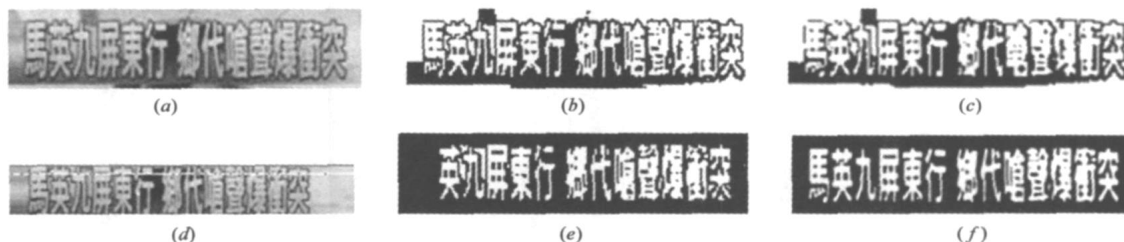


图11 提取效果比较图组D

4 结论

本文提出了一种复杂背景下的稳健的视频文字提取方法,该方法主要分为三个步骤:候选文字区域检测、重线检测和关键点获取、基于区域生长的二值化。其中,检测、定位方法可同时适应西文和中文;考虑到不同文种的笔画连通性差异,二值化方法略有不同。本文方法具有以下三个特点:第一,检测方法可以稳健地定位单行文字。第二,二值化方法很大程度上抑制复杂背景的干扰,第三,二值化方法可以在单行文字定位不佳的情况下有效提取文字。实验测试与性能比较表明整个方法对于不同文字颜色、字体、大小都保持了较好的鲁棒性,性能优于经典方法。

参考文献:

- [1] 章毓晋,基于内容的视觉信息检索[M]. 北京:科学出版社,2003.
- [2] R Lienhart, A Wernicke. Localizing and segmenting text in images, videos [J]. IEEE Transactions on Circuits Syst Video Technol, 2002, 12(4): 256 - 268.
- [3] Agnihotri L, Dimitrova N. Text detection for video analysis [A]. IEEE Workshop on Content-Based Access of Image and Video Libraries [C]. Fort Collins, CO, USA: IEEE Press, 1999. 109 - 113.
- [4] K Jain, B Yu. Automatic text location in images and video frames [J]. Pattern recognition, 1998, 31(12): 2055 - 2076.
- [5] 张佑生, 彭青松, 汪荣贵等. 基于子图像 VCH 的文本检测

- 与定位方法研究[J]. 武汉大学学报 信息科学版, 2003, 28 (3): 354 - 358.
- [6] Wenge Mao, Furlai Chung, Lam, K K M, Warr-chi Sun. Hybrid Chinese/ English text detection in images and video frames [A]. Proceedings of 16th International Conference on Pattern Recognition, 2002 [C]. Washington, DC, USA :IEEE Computer Society, Volume (3), Aug 2002. 1015 - 1018.
- [7] J Gilavata, R Ewerth, B Freisleben. A text detection, localization and segmentation system for OCR in images [A]. Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering [C]. Washington, DC, USA :IEEE Computer Society, 2004. 310 - 317.
- [8] Michael R Lyu, Jiqiang Song, Min Cai. A comprehensive method for multilingual video text detection, localization, and extraction [J]. IEEE Transaction on circuits and systems for video technology, 2005, 15(2): 243 - 255.
- [9] D Chen, K Shearer, H Bourlard. Text enhancement with asymmetric filter for video OCR [A]. In Proceedings of 11th International Conference Image Analysis Processing [C]. Palermo, Italy: IEEE Press, 2001, 192 - 197.
- [10] T Sato, T Kanade, E K Hughes, M A Smith. Video OCR for digital news archive [A]. In Proceedings of IEEE Workshop Content-Based Access Image Video Database [C]. Bombay, India: IEEE Press, 1998, 52 - 60.
- [11] C Ding, X He, H Zha, M Gu, H Simon. A min-max cut algorithm for graph partitioning and data clustering [A]. In Proceedings of IEEE International Conference Data Mining [C]. San Jose, CA, USA: IEEE Press, 2001, 107 - 114.
- [12] S U Lee, S Y Chung, R H Park. A comparative performance study of several global thresholding techniques for segmentation [J]. Computer Vision, Graphics and Image Processing, 1990, 52(2): 171 - 190.
- [13] N R Pal, S K Pal. Entropy thresholding [J]. Signal Processing, 1989, 16(2): 97 - 108.

作者简介:



田破荒 男. 1982 年出生于江西省南昌市. 硕士研究生, 研究方向为图像处理、模式识别.
E-mail: tph2003 @yeah. net



彭天强 男. 1978 年出生于湖北省随州市. 博士研究生, 讲师, 研究方向为视频处理、模式识别.
E-mail: ptq. drumboy @163. com

李弼程 男. 1970 年出生于湖南省衡南县. 现为信息工程学院教授、博士生导师, 研究方向为智能信号处理、海量信息检索.