

文章编号:1001-9081(2012)08-2305-04

doi:10.3724/SP.J.1087.2012.02305

基于笔画相关加权的视频图像文字识别

苏 畅^{1,2}, 胡晓冬^{1*}, 王斌辅¹, 尚凤军¹

(1. 重庆邮电大学 计算机科学与技术学院, 重庆 40065; 2. 美国康奈尔大学 计算机系, 纽约州 伊萨卡市 14853, 美国)

(* 通信作者电子邮箱 46765069@qq.com)

摘要:为了提取影视视频图像中的字幕信息,提出一套鲁棒的方法:首先采用图像的边缘特征对字幕信息进行区域定位,并给出结合边缘信息的方法对图像文字进行二值化;其次,采用投影法和区域生成方法定位单个文字;最后,充分考虑到文字笔画的拓扑结构,进行相邻子网格笔画结构相关性的判定,并采用笔画模糊隶属度完成弹性网格特征的提取。该方法在复杂的背景图像中能够有效得到文字的二值图像,并保证了提取特征的稳定性、健壮性,对二值化后的影视字幕的识别率达到92.1%,实验结果表明了方法的有效性。

关键词:视频图像; 文字识别; 文本定位; 二值化; 子网格特征; 笔画相关性

中图分类号: TP391.41 文献标志码:A

Video image character recognition based on stroke-related weight

SU Chang^{1,2}, HU Xiao-dong^{1*}, WANG Bin-fu¹, SHANG Feng-jun¹

(1. College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. Department of Computer Science, Cornell University, Ithaca, NY, 14853, USA)

Abstract: In order to extract the subtitle in the video image, a robust method was proposed. First, the image edge feature was adopted in caption location step, and the binarization method of text images with the edge information was given. Then, the method combined with projection and regional generation was used to locate a character. Finally, taking fully account of the topology of the text strokes, the stroke correlation among the adjacent sub-grids was determined and the stroke fuzzy membership was used to complete the elastic grid feature extraction. This method can effectively get the binary image of characters from a complex background image, ensure the stability and robustness in feature extraction. The experimental results show the method is effective, and its recognition rate has been up to 92.1%.

Key words: video image; character recognition; text location; binarization; sub-grid feature; stroke correlation

0 引言

近年来,随着多媒体技术和互联网的飞速发展,数字视频的容量正以惊人的速度增长。从数字视频中抓取到的图像中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。对嵌入图像中的文本信息进行检测与识别,使用户能更加快捷和准确地获取蕴含于图像中的丰富的信息。所以,如何从视频图像中快速而准确地定位和提取文本信息,成为现今国际上热门的研究课题^[1]。

视频图像色彩具有复杂多变性,针对这类具有复杂背景图像文字的定位方法,主要有基于连通域、基于纹理和基于边缘的方法。基于连通域方法,首先利用颜色或灰度等属性从图像中提取出连通域,然后根据几何特征来对它们进行鉴别,最后将通过鉴别的连通域合并成文字定位的结果。张佑生等^[2]提出一种变异灰度直方图,通过垂直子图像的凸台和水平子图像的凹谷检测到文字区域。基于纹理的方法是通过考察像素的邻域,应用纹理分析的方法确定出文字区域。田破荒等^[3]对图像进行小波变换,检测出高频分布下的候选文字区域,通过文字和背景穿越线特征的两次聚类实现定位。基于边缘的方法是利用图像中文本与背景的对比度来完成定位。宋砚等^[4]在边缘提取方法基础上,通过多尺度变换图像

得到的图像金字塔实现文字定位。李闯等^[5]通过边缘密度进行粗定位,再利用连接强度实现文字的精确定位。

特征提取在模式识别领域中应用十分广泛,而提取方法也是影响识别率的关键因素。弹性网格划分经常应用在特征提取之前^[6],它根据字符水平和垂直方向上的投影构成非均匀网格。刘伟等^[7]引入邻域网格的尺寸和与网格内字符像素的距离做模糊隶属度计算方法增加特征稳定性。杜佳明等^[8]提出利用质心角度的位移、拉伸和旋转不变的特征进行文字特征提取。林民等^[9]使用笔段网格字形描述方法,将笔段作为描述基元,保证了特征提取的稳定,准确地提取出局部特征。

通过以上分析,本文提出一套有效的视频字幕提取方法。文字区域定位采用基于边缘的方法,并依据其边缘进行二值化,通过投影和区域生长完成单个字符的划分。识别部分利用弹性网格相关隶属度进行特征提取,并且针对隶属度的方法进行了改进,自动计算不同笔画方向上的相应隶属度。实验表明,即使视频图像有复杂的背景,也能够较理想的定位并提取出文字信息。

1 视频图像文字定位

在对视频图像中的文字进行识别之前,需要有效地确定

收稿日期:2011-11-14;修回日期:2012-01-09。 基金项目:重庆市教委科学技术研究项目(KJ110504);重庆市科委自然科学基金资助项目(2009BB2081);教育部留学回国人员科研启动基金资助项目(教外司留[2010]1174)。

作者简介:苏畅(1979-),女,辽宁鞍山人,副教授,博士,主要研究方向:无线传感器网络通信协议、物联网;胡晓冬(1986-),男,河南新密人,硕士研究生,主要研究方向:图像处理、文字识别;王斌辅(1990-),男,河南漯河人,主要研究方向:图像处理;尚凤军(1972-),男,内蒙古宁城人,副教授,博士,主要研究方向:移动 IPv6、无线传感网络。

出图像中文字区域，并将字符图像与背景图像分离开，为字符特征提取做准备。在该阶段要进行的操作主要有文字定位、二值化和字符分割。

1.1 文字区域定位

文字区域定位是为了检测到图像中出现字幕的位置，它包括两个步骤：依据边缘的粗定位与精确定位^[5,10]。因为生活中的视频图像大部分都是彩色图像，所以要对视频图像灰度化，并进行高斯平滑降噪处理。

粗定位主要有以下步骤：

1) 提取到 Sobel 边缘强度值 $pSrc[i]$ 之后，采用均值作为全局阈值 T_g 去除强度较弱的边缘特征。

$$T_g = \frac{\sum_{i=0}^{mPixelNum} pSrc[i]}{mPixelNum} \quad (1)$$

其中 $mPixelNum$ 为像素点数。

2) 使用 $n \times n$ 的窗口扫描边缘图，统计出边缘像素点周围局部区域内的边缘数量 $edgeNum$ ，区域内边缘强度的平均值 $avgEdge$ 和边缘强度最大值 $maxEdge$ 。由此计算出每个窗口内的一个局部阈值：

$$T_i = \left(1 - \frac{edgeNum}{n \times n}\right) \times (maxEdge - avgEdge) \quad (2)$$

检查水平和垂直方向上的连通边缘点数量，当连通量大于局部阈值时，则删除该边缘曲线。

3) 对文字边缘进行修复：统计像素点周围 $m \times m$ 的模板内的边缘的加权平均值，如果像素小于平均值，则将当前像素点置为平均值。

4) 对边缘曲线进行水平和垂直投影得到大小尺寸不同的矩形文字区域。其中通过水平投影确定出文字区域的上边界和下边界，通过垂直投影确定出文字区域的左边界和右边界。

精确定位就是对文字区域进行筛选，删除错误的文字区域。首先将文字区域与行高相差较大的去除掉，其次把极亮点去除，最后通过每个边缘的外接矩形的合并融合完成文字区域的精确定位。

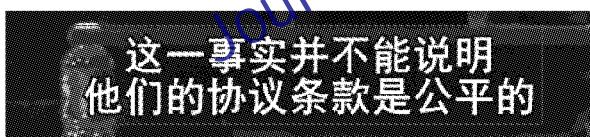


图 1 文字区域定位

1.2 文字二值化

文字二值化是为了将图像文字从背景中分割出来。文本以边缘点的像素灰度值作参考，通过计算与周围像素的加权平均值，完成二值化操作。主要有以下步骤：

1) 获取图像中的 Sobel 边缘，然后设定一个阈值将 Sobel 边缘点进行二值化。

2) 以它边缘的外接矩形作为基准线，向外延伸 w 行，取得其灰度图上该矩形内部的像素点局部均值 $avgGrayVal$ 。灰度图上与边缘曲线对应的所有像素的加权灰度平均值为：

$$avgEdgeVal = \sum grayVal(i, j) \times \left(\frac{edgeVal(i, j)}{255}\right)^2 \quad (3)$$

3) 通过比较局部均值和边缘曲线强度值对灰度图像进行二值化。当 $avgGrayVal > avgEdgeVal$ ，且 $gray(i, j) \leq avgEdge$ 时，或当 $avgGrayVal \leq avgEdgeVal$ ，且 $gray(i, j) > avgEdge$ 时，该像素点 (i, j) 为黑点（背景为白点）， $gray(i, j)$ 为像素点 (i, j) 处的灰度值。

按照上述方法，将所有的边缘线条所在区域进行二值化，

并把没有线条的区域置为白色，于是得到了整个图像的 Sobel 二值图。

1.3 字符分割

图像分割是将图像细分成为构成其子区域的对象，其程度取决于要解决的问题^[11]。图像字符分割是指将文字区域分割成一个以单个字符为主的小区域。

1.3.1 文本行分割

对现代语言来说，文字的排列一般采用横排的形式出现，而行与行之间会留有少量的空白间隙，于是可以使用这条间隙作为文本行分割的依据。

在经过 Sobel 二值化处理之后，设文字区域内某个像素值为 $f(i, j)$ 。对视频图像中的文本区域进行水平方向投影，于是可得第 i 行的投影值：

$$H(i) = \sum_{j=1}^N f(i, j); i = 1, 2, \dots, M \quad (4)$$

其中： M 为文本区域的高度， N 为文本区域的宽度。

在确定文本行上下边界之前有几个变量需要声明：连续投影值不为 0 处的行高 p_m ，连续投影值 0 处的行高 g_n ，以及连续投影值不为 0 处的平均值 p_{Avg} 。其中： m 为连续文字区域数， n 为连续间隙数。

在文字区域图像上从上向下进行行搜索，分别得到可能有字符的行高序列 p_m 和确定没有字符出现的行高序列 g_n 。

$$\text{于是得到 } p_{Avg} = \frac{1}{m} \sum_{i=1}^m p_i$$

当相邻的两个文字区域满足以下 3 个条件的其中一条时，需要进行合并操作：

条件 1 上下相邻区域的 g 值和大于合并之后高度 $unionHeight \times \sigma_1$ ，或者该区域与左右相邻区域有边界的包含关系；

条件 2 左右距离小于其最大宽度，并且其高度差小于合并之后的 $unionHeight \times \sigma_2$ ；

条件 3 存在相交的情况。

1.3.2 行内字符分割

区域生长是一种根据事前定义的准则将像素或子区域聚合成更大区域的过程，基本的方法是以一组“种子”开始将与种子性质相似的相邻像素附加到生长区域的每个种子上。于是可以将字符像素作为种子，通过区域生长来完成行内字符分割。对于种子的定义：

$$f(i, j) = \begin{cases} 0, & \text{非种子} \\ 1, & \text{种子} \end{cases} \quad (5)$$

从文本区域的右上角开始进行种子搜索，遇到种子之后，则向当前种子的邻域继续进行搜索，如果发现有新的种子，那么将这些种子做统一标识。经过一次种子区域发现之后，搜索出多个种子汇聚成的种子区域。

接下来就对这些种子区域进行相应的组合，构成真正意义上的字符区域。将种子区域扩展成矩形，区域之间存在重叠，可以看成是一个字符的几部分，将它们进行合并，之后进行新的边界划定，得到完整的字符区域。由于前面已经执行过文本行分割，对于“二”这种上下结构的字符，在没有实质性的重叠或相交发生时，也要将种子区域进行合并处理。

以上操作，完成了对常见有重叠结构字符的合并。而对于“圳”和“川”这种结构的种子区域有不重叠或相交的情况，采用以下步骤将其进行修正：

1) 对所有字符进行宽度估算。

$$w_0 = \min\left(curLine_{Height}, \frac{1}{M} \sum_{i=1}^M curRt_i_{Width}\right) \quad (6)$$

其中: $curLine_{Height}$ 为当前文本本行的行高, $curRt_i_{Width}$ 表示第 i 个字符的宽度, M 为字符总数量。

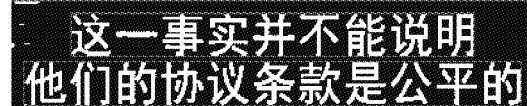
2) 为了减少被拆分开的字符和标点符号对字符合并时的影响, 当 $w_0 \times \sigma_1 < curRt_i_{Width} < w_0 \times \sigma_2$ 时, 对 w_0 进行修正。

$$w_0 = \frac{1}{N} \sum_{i=1}^N curRt_i_{Width} \quad (7)$$

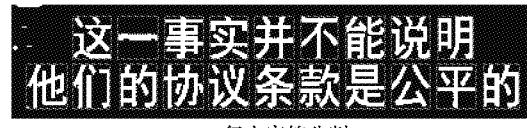
其中 N 为满足条件的种子区域个数。

3) 如果相邻两字符区域合并后的字符宽度 $\max(curRt_i_{Width}, curRt_{i+1}_{Width}) < unionRt_{i+1}_{Width} < line_{Height} \times \beta$, 则进行合并操作; 否则不再进行合并。

上述字符分割的结果如图 2 所示。



(a) 文本行分割



(b) 行内字符分割
图 2 字符分割

2 视频图像文字识别

文字特征提取在视频图像文字的识别过程中起着重要的作用, 稳定的特征提取能够保证好的识别结果。于是先对字符进行弹性网格划分, 再判定相邻网格的在对应笔画上的相关性, 最后采用相邻网格内笔画模糊相关隶属度计算特征值来完成文字特征提取。

2.1 构建文字弹性网格

在文字识别初期, 多是以均匀划分的固定网格作为特征提取的前提操作, 但是固定网格会因文字大小或笔画位移落入非固定网格中, 造成特征提取时有比较大的误差。而弹性网格的构建都是由当前所识别的图像文字自身的笔画密度所决定, 能有效避免以上问题。均匀网格划分和弹性网格划分结果如图 3 所示。

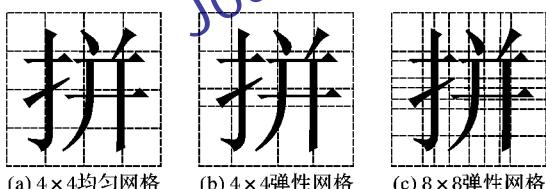


图 3 弹性网格划分

为了将大小为 $M \times N$ 的图像文字划分出 $P \times Q$ 个子网格^[6], 首先对图像文字在水平和垂直方向上进行笔划密度投影, 依据式(4)的方法分别得到 $H(i)$ 和 $V(j)$ 。最简单的笔画密度函数就是 $d(i, j) = f(i, j)$ 。

于是可以得到弹性网格在水平与垂直方向的坐标:

$$P_s = \min \left\{ k \left| \frac{s-1}{P} \sum_{i=1}^M H(i) \leq \sum_{i=1}^k H(i) \leq \frac{s}{P} \sum_{i=1}^M H(i) \right. \right\}; 1 \leq s \leq P \quad (8)$$

$$Q_t = \min \left\{ k \left| \frac{t-1}{Q} \sum_{j=1}^N V(j) \leq \sum_{j=1}^k V(j) \leq \frac{t}{Q} \sum_{j=1}^N V(j) \right. \right\}; 1 \leq t \leq Q \quad (9)$$

因此, 根据水平和垂直方向上的坐标, 划分出的网格有如下定义:

$$mesh(i, j) = \{(i, j) | P_s \leq i \leq P_{s+1}, Q_t \leq j \leq Q_{t+1}\} \quad (10)$$

2.2 自适应加权弹性网格特征提取

弹性网格的划分, 使完整的字符被分成 $P \times Q$ 个子网格。由于字符笔画都是相连的, 这样在划分出的子网格之间就存在一定的联系, 如横笔画会分布在水平相连的子网格中(如图 4 所示)。由于这种笔画结构之间的相关性, 为了能够合理地获取 8 邻域子网格的贡献, 这里拟采用邻域子网格的梯度方向特征与质心相结合的方法, 依据其共同的偏移方向来判断中心子网格是否受到其邻域的影响。而邻域子网格相关笔画的贡献率采用模糊隶属度函数来确定。

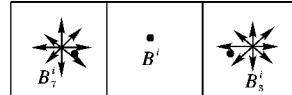


图 4 横笔画子网格相关性

对于子网格 B^i 的横笔画方向特征, 首先考虑其邻域的方向性, 而 B^i 的横笔画只受到水平相邻子网格 B_3^i 和 B_7^i 的影响, 邻接的其他子网格将不对其横笔画构成影响。当质心偏移的方向与其子网格的方向性相同时, 则说明该子网格对 B^i 的特征提取有贡献作用, 可以依据其贡献率获取特征。

相邻子网格内每个像素点的梯度方向和幅值^[12]分别为:

$$\theta(x, y) = \tan^{-1} \frac{(f(x, y+1) - f(x, y-1))}{(f(x+1, y) - f(x-1, y))} \quad (11)$$

$$m(x, y) = ((f(x+1, y) - f(x-1, y))^2 + (f(x, y+1) - f(x, y-1))^2)^{1/2} \quad (12)$$

窗口内每个像素点梯度的幅值利用高斯函数进行加权, 高斯函数表达式为:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp \left(-\frac{x^2 + y^2}{2\sigma^2} \right) \quad (13)$$

在窗口内每个 4×4 的小块上统计计算 8 个方向上的图像梯度方向直方图, 若某个像素点梯度方向落在某个方向的附近, 则其相应的梯度高斯加权幅值就累加在这个方向上, 箭头长度代表这个方向上的梯度加权幅值的累加和。

对于子网格 B^i , 它的质心位置计算公式如下:

$$x_c = \frac{\sum_{(x, y) \in B^i} f(x, y) \times x}{\sum_{(x, y) \in B^i} f(x, y)} \quad (14)$$

$$y_c = \frac{\sum_{(x, y) \in B^i} f(x, y) \times y}{\sum_{(x, y) \in B^i} f(x, y)} \quad (15)$$

若 B_3^i 水平左指向为主方向, 且质心在指向的附近位置, 则 B_3^i 为 B^i 的横笔画特征作贡献; 若 B_7^i 水平右指向为主方向, 且质心在指向的附近位置, 则 B_7^i 为 B^i 的横笔画特征作贡献。因此, 在计算的横笔画方向特征时只需要加入 B_3^i 和 B_7^i 的贡献率。

对于子网格 B^i 中含有的字符部分, 其他笔画方向特征(如图 5 所示)与相邻子网格的贡献有以下关系: 坚笔画受到垂直相邻子网格 B_1^i 和 B_5^i 的影响; 撇笔画受到相邻子网格 B_2^i 和 B_6^i 的影响; 捺笔画受到相邻子网格 B_4^i 和 B_8^i 的影响。它们的具体计算方法与横笔画相似。

若子网格质心落在指定笔画方向的附近, 则采用模糊隶属度获取其弹性网格特征, 隶属度函数^[7]设定为:

$$\mu(a, b) = \exp \left(-\frac{(3a/(a+b))^2}{2} \right) \quad (16)$$

其中: a 表示子网格指向 B^i 方向上的梯度幅值, b 表示 B^i 指向子网格在对应方向上的梯度幅值。若子网格质心没有落在

指定笔画方向的附近,则 $\mu(a, b) = 0$ 。

在对 B^i 的弹性网格特征计算时,横竖撇捺四个方向向上子图分别用 $f_H(x, y)$ 、 $f_V(x, y)$ 、 $f_L(x, y)$ 、 $f_R(x, y)$ 来表示。因为来自其 8 邻域子网格的影响程度不同,所以为相邻子网格提供不同的隶属度 $\mu(a, b)$ 值。从而结合考虑 B^i 与其 8 邻域子网格的笔画贡献情况(如图 6 所示),可获知子网格 B^i 的弹性网格特征。

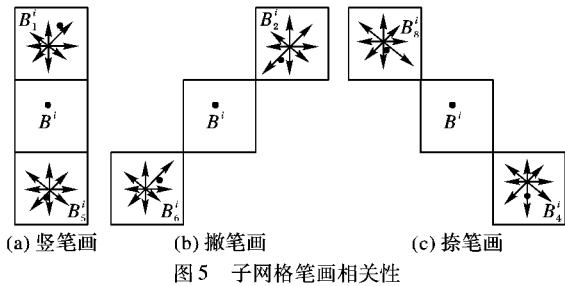


图 5 子网格笔画相关性

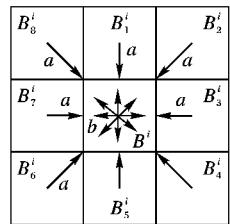


图 6 邻域子网格的笔画贡献

对于 B^i 的横笔画特征有:

$$D_H^i = \frac{\iint f_H(x, y) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_H(x, y) \mu_r(a, b) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_H(x, y) \mu_l(a, b) dx dy}{\iint f(x, y) dx dy} \quad (17)$$

其中: μ_r 为 B_3^i 对 B^i 的横笔画的贡献率, μ_l 为 B_7^i 对 B^i 的横笔画的贡献率。

对于 B^i 的竖笔画特征有:

$$D_V^i = \frac{\iint f_V(x, y) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_V(x, y) \mu_t(a, b) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_V(x, y) \mu_b(a, b) dx dy}{\iint f(x, y) dx dy} \quad (18)$$

其中: μ_t 为 B_1^i 对 B^i 的竖笔画的贡献率, μ_b 为 B_5^i 对 B^i 的竖笔画的贡献率。

对于 B^i 的撇笔画特征有:

$$D_L^i = \frac{\iint f_L(x, y) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_L(x, y) \mu_n(a, b) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_L(x, y) \mu_{lb}(a, b) dx dy}{\iint f(x, y) dx dy} \quad (19)$$

其中: μ_n 为 B_2^i 对 B^i 的撇笔画的贡献率, μ_{lb} 为 B_6^i 对 B^i 的撇笔画的贡献率。

对于 B^i 的捺笔画特征有:

$$D_R^i = \frac{\iint f_R(x, y) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_R(x, y) \mu_{nb}(a, b) dx dy}{\iint f(x, y) dx dy} + \frac{\iint f_R(x, y) \mu_h(a, b) dx dy}{\iint f(x, y) dx dy} \quad (20)$$

其中: μ_{nb} 为 B_4^i 对 B^i 的捺笔画的贡献率, μ_h 为 B_8^i 对 B^i 的捺笔画的贡献率。

3 实验结果及分析

本文实验数据集来源于数字影视和网络视频网站,共 100 幅图像,包含了 1072 个字符,涉及到的字符字体有黑体、宋体、雅黑、幼圆和楷体等常见形式。图像的分辨率有从 640×350 到 1280×800 等多种形式。

对于文字定位和识别的测评标准不同,于是实验分为两部分:定位和识别。

在对文字定位进行测量时,为了衡量方法的性能,对文字区域定位参照文献[3]的实验方法给出以下评价标准:如果定位区域中存在有一行以上文字,且真实文字区域占据的了该定位区域中所有方字行区域的 75% 以上,则可认为正确定位到该图像中的文字区域;如果定位到的单个字符区域覆盖了真实字符区域所有像素,且不超过字符外接矩形的 20%,则可认为正确定位到该字符。于是对实验数据的文字区域准确定位率、文字区域全定位率、单个字符准确定位率和单个字符全定位率四个方面进行测量评判。实验结果如表 1 所示。

表 1 视频文字区域定位实验结果

测量指标	值	测量指标	值
文字区域 准确定位率/%	56.8	文字区域 全定位率/%	89.7
单个字符 准确定位率/%	93.8	单个字符 全定位率/%	91.8

从实验结果可以看出,本文的区域定位方法能够保证对文字区域 89.7% 的全定位率。但是由于在定位时主要是依据图像的边缘信息来检测文字区域,以及利用字符的间距进行字符分割,因此对于边缘较为丰富的图像,会出现较多误检的情况。对于单个字符的划分,在边缘二值化方法的基础上进行,通过投影和区域生长相结合的方法,在实验数据集上单个字符准确评估率达到 93.8%,结果较为满意。

识别实验测评部分中,采用距离测试里最常用的欧氏距离作为其距离度量准则,特征距离相差最小的作为识别结果。实验结果如表 2 所示。

表 2 视频文字区域识别实验结果

方法	可识别率/%	拒识率/%	正确识别率/%	误识率/%
均匀网格	98.9	1.1	87.2	11.7
弹性网格	97.7	2.3	89.2	8.6
本文方法	97.5	2.5	92.1	5.5

(下转第 2312 页)

横向移动的滑动窗口,对不同样本截取不同数量的子图像;对子图像用均匀网格划分,统计网格中像素密度,作为观测向量;观测序列通过局部保持投影映射到低维空间,通过抽样的方法降低了由于邻接图过大而产生的过高的空间和时间复杂度;最后用新形成的观察序列训练隐马尔可夫模型。实验表明,本文方法保持了隐马尔可夫模型对维文字符的结构动态建模能力,对手写维文字符鲁棒性较好,在利用 LPP 的局部保持能力降维的同时,获得了优于原观测空间的聚类效果。本文算法将 LPP 与 HMM 构成一个统一的模型,实验表明,LPP-HMM 算法适用于维文单字母建模。考虑到手写维文的书写习惯,今后的研究重点,应为手写维文连接词识别。

参考文献:

- [1] 赵继印, 郑蕊蕊, 吴宝春, 等. 脱机手写体汉字识别综述[J]. 电子学报, 2010, 38(2): 405–414.
- [2] 袁保社, 吾守尔·斯拉木. 一种手写维吾尔文字符识别算法[J]. 计算机工程, 2010, 36(2): 186–188.
- [3] UBUL K, HAMDULL A, AYSA A, et al. Research on Uyghur off-line handwriting-based writer identification [C]// ICSP 2008: 9th International Conference on Signal Processing. Piscataway: IEEE, 2008: 1656–1659.
- [4] 哈力木拉提, 阿孜古丽. 多字体印刷维吾尔文字符识别系统的研究与开发[J]. 计算机学报, 2004, 27(11): 1480–1484.
- [5] RABINER L R. A tutorial on hidden Markov models and selected applications in speech recognition [J]. Proceedings of the IEEE, 1989, 77(2): 257–286.
- [6] LORIGO L M, GOVINDARAJ V. Off-line Arabic handwriting recognition: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(5): 712–724.
- [7] VINCIARELLI A. A survey on off-line cursive word recognition[J]. Pattern Recognition, 2002, 35(7): 1433–1446.
- [8] XIANG DONG, YAN HUAHUA, CHEN XIANQIAO, et al. Offline Arabic handwriting recognition system based on HMM [C]// 2010 3rd IEEE International Conference on Computer Science and Information Technology. Piscataway: IEEE, 2010, 1: 526–529.
- [9] MOHAMAD R A-H, LIKFORMAN-SULEM L, MOKBEL C. Combining slanted-frame classifiers for improved HMM-based Arabic handwriting recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(7): 1165–1177.
- [10] EL-HAJJ R, LIKFORMAN-SULEM L, MOKBEL C. Arabic handwriting recognition using baseline dependant features and hidden Markov modeling [C]// Proceedings of the Eighth International Conference on Document Analysis and Recognition. Piscataway: IEEE, 2005, 2: 893–897.
- [11] VINCIARELLI A, BENGIO S. Offline cursive word recognition using continuous density hidden Markov models trained with PCA or ICA features [C]// ICPR'02: Proceedings of 16th International Conference on Pattern Recognition. Washington, DC: IEEE Computer Society, 2002, 3: 81–84.
- [12] HE XIAOFEI, NIYOGI P. Locality preserving projection [C]// Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2004: 153–160.
- [13] 李真. 基于图像优化局部保留投影的人脸表情识别[J]. 计算机工程与应用, 2011, 47(27): 210–211.
- [14] 刘敏, 李晓东, 王振海. 一种新的有监督局部投影人脸识别算法[J]. 计算机应用, 2009, 29(5): 1416–1418.

(上接第 2308 页)

从表 2 可以看出,本文方法的可识别率比均匀网格特征提取降低了 1.4%,而与普通弹性网格没有明显的差距。对于识别结果比均匀网格特征提取的正确识别率提高了 4.9%,而误识率下降了 6.2%;与普通弹性网格特征提取相比较,正确识别率提高了 2.9%、误识率下降了 3.1%。识别率较之前方法有所提高,同时误识率也得到降低,说明了本文在提取弹性网格的特征加入相邻子网格笔画相关性的判定后,通过笔画相关性计算出子网格的隶属度,而不采用弹性网格方法划分出的相邻子网格尺寸大小计算,使提取出的特征更稳定,证实了本方法的有效性。

4 结语

本文对视频图像字幕提取的定位与识别进行了研究,给出了一套复杂背景情况下稳健的文字提取方法。在视频图像字幕定位方面,结合边缘特征的文本区域查找方法,充分利用了文字的原始形状特征,较好地完成了复杂背景下的文字二值化和定位。在视频图像字幕识别方面,采用弹性网格对字符进行划分,引入相邻网格笔画相关性的判断,并使用字符子网格与相邻子网格对应笔画计算出模糊隶属度,之后再结合所有 8 邻域进行特征提取,从而有效地抑制了子网格非相关笔画的影响度,扩展了弹性子网格特征的附加信息量,增强了特征提取的鲁棒性和稳定性,取得了良好的实验结果。

参考文献:

- [1] MERLER M, KENDER J R. Semantic keyword extraction via adaptive text binarization of unstructured unsourced video [C]//

- ICIP'09: Proceedings of the 16th IEEE International Conference on Image Processing. Piscataway: IEEE, 2009: 261–264.
- [2] 张佑生, 彭青松, 汪荣贵. 一种基于变异灰度直方图的视频字幕检测定位方法[J]. 电子学报, 2004, 32(2): 314–317.
- [3] 田破荒, 彭天强, 李弼程. 基于文字穿越线和笔画连通性的视频文字提取方法[J]. 电子学报, 2009, 37(1): 72–78.
- [4] 宋砚, 刘延安, 张勇长, 等. 基于聚类的视频字幕提取方法[J]. 通信学报, 2009, 30(2): 136–140.
- [5] 李闯, 丁晓青, 吴佑寿. 基于边缘的字符串定位算法[J]. 清华大学学报: 自然科学版, 2005, 45(10): 1335–1338.
- [6] 高学, 金连文, 尹俊勋. 一种基于笔画密度的弹性网格特征提取方法[J]. 模式识别与人工智能, 2002, 15(3): 351–354.
- [7] 刘伟, 朱宁波, 何浩智, 等. 基于弹性网格模糊特征的手写汉字识别方法[J]. 中文信息学报, 2007, 21(3): 117–121.
- [8] 杜佳明, 王朝立, 朱枫. 基于质心角度的字符图像识别算法[J]. 信息与控制, 2010, 39(3): 320–325.
- [9] 林民, 宋柔. 一种笔段网格汉字字形描述方法[J]. 计算机研究与发展, 2010, 47(2): 318–327.
- [10] LYU M R, SONG J, CAI M. 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.
- [11] 杨武夷, 张树武. 一种视频中字符的集成型切分与识别算法[J]. 自动化学报, 2010, 36(10): 1468–1476.
- [12] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91–110.