

文章编号: 2095-2163(2019)01-0203-05

中图分类号: TP391.41

文献标志码: A

超低分辨率视频中的字符识别技术研究

王 瑞, 李海峰, 马 琳

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘 要: 随着计算机技术的发展,具有录制视频功能的便携设备越来越多,而视频网站通知也便利了视频的发布,产生了海量的视频数据。研究中发现视频中的文本信息尤为重要,因其常常与视频内容主题直接相关,对于视频中的信息检索或者下一步的数据挖掘具有重要意义。而许多视频由于录制设备的影响,其分辨率往往很低。本文针对超低分辨率视频中出现的印刷体字符进行识别,使用基于模板匹配算法的核心字符识别引擎,同时为了最大程度减少用户的标注次数,提出了全自动生成模板库的算法。为了进一步提高识别率,在识别结果后处理环节加入了纠错反馈机制。最后,根据提出的算法构建了一套自动识别视频字符的实用系统,在实际应用中取得了很好的效果。

关键词: 字符识别; 视频; 超低分辨率; 图像处理

Research of ultra-low resolution video character recognition technology

WANG Rui, LI Haifeng, MA Lin

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

【Abstract】 There are more and more portable smart video-recording devices with the development of computer technology. Video website is so convenient that everyone can freely upload their videos. Now there has vast amount of video data. It is really important that the text is extracted within video, because these text have strong relation to video content. And the users can apply these text for video retrieval or other data mining purpose. Many videos have low resolution due to the limited record function of devices. This thesis focuses on the printed character recognition of ultra-low resolution video. The core recognition engine is based on the template matching algorithm. In order to avoid labeling too much data for training, the paper proposes a method which can generate templates automatically. After recognition result produced, the paper also uses the method of error correction feedback to improve the recognition accuracy. Finally, the paper practically develops a video OCR system according to the proposed algorithm, and achieves great performance in practical applications.

【Key words】 character recognition; video; ultra-low resolution; image processing

0 引 言

随着科技的发展,便携式的智能设备已经进入普及阶段,而设备中往往都具有着视频录制功能,通过智能手机、平板电脑等任何实体,人们录制视频也日渐趋于简易,而网络的日益发达,又促使人们将视频上传至网络,形成了海量的视频数据。作为一种信息传递的模式,相对于文字、声音,视频数据携带的信息量更为丰富。如何从当下人们产生的海量视频数据中提取有意义的信息一直以来即是学界研究热点,其中一个备受多方关注的课题方向就是视频中文本的识别技术,因为视频中出现的文本信息常常会与视频的内容主题直接相关,可以将其应用于视频检索领域,或者作为下一步数据挖掘的数据来源^[1]。

而这些视频当中,很多视频由于录制设备的限制(手机、平板电脑、监控设备等),其分辨率往往不高,甚至受到干扰噪声影响,质量上迹近低劣。本文主要对于超低分辨率视频的字符识别进行了研究。目前针对此类视频的字符识别,传统的视频 OCR 技术无法应用于其中,比如由于字符过小,二值化处理后信息会丢失许多,许多经典的特征提取方法也不再适用^[2]。时下,关于此类低分辨率的视频识别问题,Ohkura 等人^[3]提出使用超分辨率(Super Resolution)的识别方法,将低分辨率的图片复原为高分辨率图片,再使用传统的字符识别技术即可获得有效实现。Einsele-Azami^[4]的研究是将高清图片进行处理,模拟低分辨率视频的生成模型,生成低分辨率的字符后再用于其后的识别。而本次研究中,拟将使用基于模板匹配的核心识别算法,对超低分

作者简介: 王 瑞(1992-),男,硕士研究生,主要研究方向:图像处理、文本识别;李海峰(1969-),男,博士,教授,博士生导师,主要研究方向:智能信息处理技术、人工脑与认知科学、语言信息处理等;马 琳(1967-),女,博士,副教授,主要研究方向:模式识别、生物信息技术、认知科学等。

收稿日期: 2016-06-28

辨率的视频进行识别。同时为了减少用户标注的工作量,提出了自动生成模板库的算法,以及为了改进识别率而提出的纠错后处理算法。最终,又设计研发了一套完整的视频字符识别系统,在实际应用中取得了稳定良好的效果,具有高度重要的现实意义。

1 超低分辨率字符识别算法研究

1.1 超低分辨率字符定义

目前,对于超低分辨率的界定并未形成公认的定义,本文根据已有研究者给出的定义^[5-6],认为字符大小在 3×5 像素至 10×10 像素下的字符为超低分辨率字符。如图1所示,左侧文本图片的字符大小为 5×8 像素,右侧文本的大小为 3×5 像素。在超低分辨率下,由于像素点很少,携带信息的能力有限,因此形态较为复杂的字符,比如汉字,在此分辨率下根本无法做到清晰识别。为此,定义本算法识别的字符范围为阿拉伯数字0-9、大小写英文字符,以及某些形状单一的特殊符号,如“.”、“-”、“+”等。



(a) 5×8 像素文本图片 (b) 3×5 像素文本图片
(a) 5×8 pixel text image (b) 3×5 pixel text image

图1 视频字符示意图

Fig. 1 Text image in ultra-low resolution video

1.2 系统模板库自动生成算法

基于模板匹配的字符识别算法,需要建立一个高质量的模板库。也就是模板库中需要包含不同字符的图片,这些图片与视频中待识别的图片大小相

同。为了取得较好的识别效果,这些模板的最佳选择就是直接来自于待识别视频,因为这样模板库的数据与识别数据即会属于同一个概率分布。

如果由人工进行模板挑选,需要较大的工作量,尤其是面对需要识别大量视频的情况下,工作量将更加可观。因此需要建立一种自动生成识别模板库的方法。自动生成模板库算法的步骤流程具体如下。

(1) 用户设置识别字符的字符集。研究中设计了一个人机交互界面,具有相应的选项,供用户选择字符集。比如数字字符集、大写英文字母字符集、小写英文字母字符集。同时,用户也可以自定义字符集,输入一些符号类型的字符,比如逗号、句号、问号等。

(2) 按照用户定义的字符集生成高清字模图片。将用户定义的字符集生成分辨率高的图片。由于未能确定用户的字符背景色是黑色、还是白色,有针对性地生成2套高质量字模图片。此步骤由编写的程序自动生成某一字体的单一字符图片。

(3) 将高清字模图片按照目标视频识别的大小进行缩放。对于(2)中生成的高分辨率字符图片使用图像处理算法进行处理,使得其图片大小与待识别字符整体保持一致。并且能够最大程度地保留字符原有信息,即依然可以分辨出处理后是哪个字符。

综上3个步骤生成的标准字模图片如图2所示。第一行是高清字符图片,每个字符高52像素,宽30像素。第二行是经过高斯平滑后的图片。第三行是最终加入缩放处理后生成的图片,字符高8像素,宽5像素。



(a) 高清字符图片
(a) High resolution character image

(b) 经过高斯平滑后的图片
(b) Gaussian smoothed image

(c) 加入缩放处理后的图片
(c) The scaled image

图2 高清字符缩放结果

Fig. 2 Down sampling result for high resolution image

(4) 对于待识别的字符背景是黑色、还是白色进行检测。此处需要开展是黑色背景、白色背景、至灰度图片中背景接近黑色、或者接近白色情况的应用辨识。该步骤是为下一步挑选模板提供基础准备,如果未能获知当前的字符背景信息,那么也将无

法生成和待识别字符相近的模板库。识别的基本假设如下:如果待识别字符是黑色背景,那么总体上待识别字符与黑色背景的距离要小于白色背景的标准字模,反之亦然。

(5) 从待识别的图片中挑选出与标准字模接近

的图片加入模板库。使用模板匹配算法,以(3)中生成的字模图片作为模板原型,对于视频进行字符识别,将与模板原型最接近的字符挑选出来。加入到模板库中。目前的选择策略是对于每一个标准字模,将其与每个待识别的字符计算距离形成距离分布,取前 K 个加入模板库。如图3所示,第一列是自动生成的标准模板,其后的图片是根据选择策略从待识别字符中选取的新模板。

(6)对于每一个字符,保存其垃圾模板的阈值。利用上一步中每个标准字模的距离分布,可以确定垃圾模板的距离阈值。当进行模板匹配时,会检查是否超过此阈值,若超过则认为识别结果无效,该字符质量过差,无法识别。

(7)保存模板库至文件,模板库建立完成。将模板库中的多幅图片转换为一个灰度值矩阵,并将此矩阵保存为特殊格式的文件存储在磁盘中。其它视频进行字符识别时可以共用此模板库。



图3 模板自动生成算法结果

Fig. 3 Result of auto-generating templates algorithm

1.3 核心识别算法

OCR系统的识别引擎基于模板匹配算法。其输入为已分割好的待识别字符以及模板库,输出为图片中表示的字符内容。

本字符识别模板匹配算法的步骤主要包括:图像预处理,提取特征,进行特征匹配,记录候选名单,输出识别结果。在此,对于每一步骤的实现解析可给出如下论述。

(1)预处理阶段。将分割的图片进行灰度化与标准化处理,同时检测分割图片中没有文字内容的空白图片。对于极低分辨率的字符,由于二值化后图片的信息丢失严重,研究中并未采用传统的二值化方法进行预处理。

(2)特征提取阶段。本识别引擎使用的是图像像素的灰度值数据,在极低分辨率的情况下,无法提取经典的特征,比如点、线等特征,因此本次研究采用了最直观的像素特征。

(3)特征匹配。计算待识别图片与模板库中图片的相似度,经过实验,使用余弦距离的识别效果要更显优势。

(4)记录候选名单。对于每一个待识别的字符,各自记录模板库中与待识别字符的相似度在前3位的字符,形成第一候选(即识别结果)、第二候选、第三候选。注意,这里要求3个字符分别属于不同类别。为此,则将第一候选的内容作为结果输出。同时将第二候选、第三候选的记录也保存起来,供后处理模块调取使用。

1.4 系统纠错后处理算法

在自动建立模板库后,此时的模板库的质量并不高,因为使用的高清字符的字体可能与视频中的字符不同,存在一定的差异,因此生成的模板库也不会是完全正确,导致第一次识别的准确率不会太高。综上分析可知,识别后处理过程就尤为重要,在后处理过程中对于模板库进行逐步的优化,使用人工纠错与自动纠错互相配合的方法使得识别结果不断提高。

1.4.1 AR模型自动纠错算法

自动纠错算法针对的是数值类型的变量的纠错。其基本假设是,视频中的数值一般都代表自然界中的物理量,而这些物理量随着时间的变化都是连续的,很少会发生数值的突变。如果出现突变,自动纠错算法会认为是错误,并且自行纠正,同时提醒用户。

本系统使用自回归模型(Autoregressive model, AR模型)进行纠错。AR模型是统计中一种处理时间序列的方法。对于变量 x ,则有:

$$x(t) = \sum_{i=1}^p \phi_i X_{t-i} \quad (1)$$

由式(1)可知,当前 t 时刻的值可由之前的历史数据的线性组合形成。本系统中使用的AR模型为5阶模型,即:

$$x(t) = \sum_{i=1}^p \phi_i X_{t-i} \quad p=5 \quad (2)$$

AR模型的自动纠错步骤,可分述如下:

(1)对于包含错误的识别结果 $X(t)$,可以使用AR模型参数估计得到预测结果 $X'(t)$ 。

(2)若 $|X'(t) - X(t)| \geq \text{thresh}$,则认为这里出现了异常的突变点,需要对识别错误的数据进行纠

正。根据实验,阈值的取值为预测与识别结果均值的 N 倍,此处 N 取为 3。阈值计算的数学表述如式(3)所示:

$$thresh = \frac{3}{T} \sum_{t=1}^T |X'(t) - X(t)| \quad (3)$$

这样的阈值将可以达到较好的纠错结果。

(3)纠正的时候,首先确定是哪一位导致错误最大,接着从识别产生的第一候选与第二候选列表中尝试纠错,如果使用候选名单中替换后小于阈值,则纠错结束。如果候选列表纠错效果未臻理想,则直接使用 AR 模型拟合所得值。同时,配合识别可信度进行检查,对于可信度高的字符不予纠错。

实践中发现,AR 模型的局限性在于:对于识别错误率较大的情况下(大于 5%),AR 模型无法使

用,此时需要进行手动纠错。另外,对于比较小的突变,比如数值类型末尾,由于变化量小,也无法纠错,因此自动纠错主要就是在手动纠错修正了几次后,错误率较低的情况下获得使用。

1.4.2 人工纠错

人工纠错算法通过一个用户界面,将识别结果与实际图像同时显示,由用户进行手动纠错。用户可以在人机界面中输入正确的标签,而后将该字符图片以及标签加入模板库。人工纠错时,对于质量处理特别差的字符,可以标注为垃圾模板。

2 超低分辨率字符识别系统设计

2.1 系统整体架构

视频 OCR 系统的整体运行流程如图 4 所示。

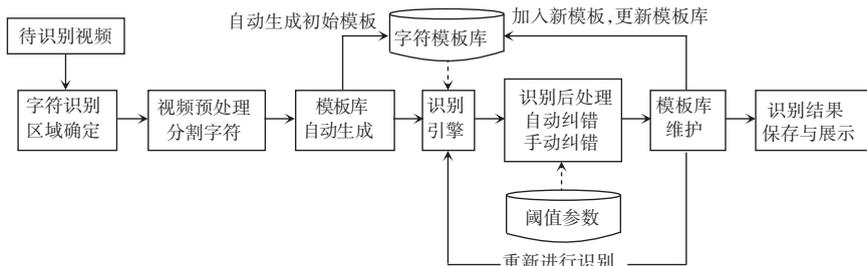


图 4 系统运行流程图

Fig. 4 Operation flow diagram of OCR system

由图 4 可知,用户对于一个新类的视频,首先进行识别区域的选择,对待识别的每个字符分辨进行框选。接着将视频每一帧的图像预处理按照确定的识别区域予以分割,之后开启预处理,比如灰度化、标准化等。

在此基础上,运行系统初始模板库建立算法,建立初始模板库,接着进行第一次识别。识别后将启用后处理模块对于识别结果进行分析,在这一模块会用到一些先验知识来进行自动纠错,例如数值型物理量连续变化的特点,以及识别引擎对于识别结果的可信度信息。如果可信度高,那么不进行纠错。

模板库维护算法可以判定甄别是否将纠错后的字符加入模板库,而由更新后的模板库对视频给出二次优化识别。重复如上纠错步骤,使得模板库质量不断提升,识别效果持续改进。直到用户认为识别准确率满足要求,保存最终结果。

系统整体功能结构如图 5 所示。系统开发环境为 VS2013,开发语言为 C++,使用 Qt GUI Framework 设计界面,研发定制功能。使用 OpenCV 开源库辅

助实现图像处理算法。

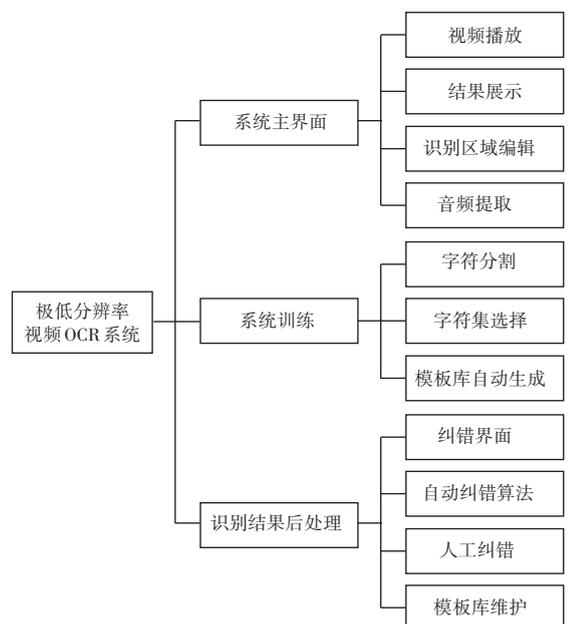


图 5 系统整体功能结构图

Fig. 5 Structure diagram of system function module

2.2 系统实验结果

实验时,使用项目合作方提供的 5 个视频数据。6 个视频的基本信息见表 1。

表 1 测试数据基本信息

Tab. 1 Basic information of test video data

视频名称	字符大小/ 像素	每一帧 字符个数	视频 长度/s	视频编码 格式	分辨率 (横纵比)
视频一	5×8	14	79	MPEG2	720×576
视频二	5×8	14	48	MPEG2	720×576
视频三	5×9	12	48	MPEG2	720×576
视频四	6×8	14	49	MPEG2	720×576
视频五	3×5	19	49	MPEG2	720×576

为了验证后处理纠错反馈算法的有效性,记录了每一轮纠错迭代过后的识别结果,这里定义一轮迭代为用户纠错 5 次,且每一次的纠错类型不同(处于不同位置,错误类型不同),5 次纠错后更新模板库,再次识别。具体数据可见表 2。

表 2 迭代识别错误率统计

Tab. 2 Error rate of iterative character recognition %

视频名称	自动生成模板 识别错误率	第一轮 纠错迭代	第二轮 纠错迭代	第三轮 纠错迭代
视频一	4.7	2.1	1.0	0.4
视频二	3.8	1.5	0.9	0.4
视频三	2.9	2.0	0.8	0.7
视频四	3.2	2.4	1.1	0.5
视频五	9.3	4.3	2.0	0.9

由表 1、表 2 可以看出,使用自动生成的模板库进行首次识别的时候,识别率都不高。这是由于在这一环节并未让用户进行任何的标注,因此缺少关于问题域的知识,这就使得自动生成的模板库比较粗糙,掺杂了一些错误模板,导致识别率降低。而由之后每一轮纠错迭代后的识别率可以看出,使用本文的纠错算法,模板库的质量快速提高,从而使得识别的错误率显著下降。这也充分证明了后处理纠错算法的有效性。

3 结束语

从超低分辨率视频中提取有效的字符信息,目前研究成果较少,现阶段大多都是针对高清分辨率下的视频字符开展研究。在本文中,设计的核心识别引擎采用了模板匹配算法,为了尽量减少用户标注的工作量,再次提出了全自动生成模板库的算法,大大提高了用户的工作效率。同时为了弥补自动生成模板库导致的模板库质量的退化,又专门提出了一种纠错反馈机制,包含人工纠错与自动纠错,使得模板库的质量快速提高,从而保证了优异的字符识别准确率。最后,则综合研发了一套完整的视频 OCR 识别系统,提供了人性化的用户界面,满足了项目方实践使用的需求。在接下来的工作中,主要将重点研究低分辨率视频下的文本区域自动定位,以及文本跟踪技术。目前识别区域的确定是需要用户编辑,今后则将通过自动文本定位算法,进一步优化系统识别流程,让用户的操作更加简单、便捷。

参考文献

- [1] 朱成军, 李超, 熊璋. 视频文本检测和识别技术研究[J]. 计算机工程, 2007, 33(10): 218-219, 245.
- [2] SHARMA N, PAL U, BLUMENSTEIN M. Recent advances in video based document processing: A review [C]// 2012 10th IAPR International Workshop on Document Analysis Systems. Gold Coa: IEEE, 2012: 63-68.
- [3] OHKURA A, DEGUCHI D, TAKAHASHI T, et al. Low-resolution character recognition by video-based super-resolution [C]//10th International Conference on Document Analysis and Recognition. Barcelona, Spain: IEEE, 2009: 191-195.
- [4] EINSELE-AAZAMI F. Recognition of ultra low resolution, anti-aliased text with small font sizes [D]. Switzerland: University of Fribourg, 2008.
- [5] TOBA S, KUDO H, MIYAZAKI T, et al. Ultra-low resolution character recognition system with pruning mutual subspace method [C]// 2015 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW). Taiwan: IEEE, 2015: 284-285.
- [6] SLIMANE F, KANOUN S, HENNEBERT J, et al. A study on font-family and font-size recognition applied to Arabic word images at ultra-low resolution [J]. Pattern Recognition Letters, 2013, 34(2): 209-218.