

## 两阶段的视频字幕检测和提取算法

王智慧<sup>1</sup> 李佳桐<sup>2</sup> 谢斯言<sup>2</sup> 周 佳<sup>2</sup> 李豪杰<sup>1</sup> 樊 鑫<sup>1</sup>

(大连理工大学国际信息与软件学院 辽宁 大连 116621)<sup>1</sup> (大连理工大学软件学院 辽宁 大连 116621)<sup>2</sup>

**摘 要** 视频字幕检测和提取是视频理解的关键技术之一。文中提出一种两阶段的字幕检测和提取算法,将字幕帧和字幕区域分开检测,从而提高检测效率和准确率。第一阶段进行字幕帧检测:首先,根据帧间差算法进行运动检测,对字幕进行初步判断,得到二值化图像序列;然后,根据普通字幕和滚动字幕的动态特征对该序列进行二次筛选,得到字幕帧。第二阶段对字幕帧进行字幕区域检测和提取:首先,利用 Sobel 边缘检测算法初检文字区域;然后,利用高度约束等剔除背景,并根据宽高比区分出纵向字幕和横向字幕,从而得到字幕帧中的所有字幕,即静止字幕、普通字幕、滚动字幕。该方法减少了需要检测的帧数,将字幕检测效率提高了约 11%。实验对比结果证明,相比单一使用帧间差和边缘检测的方法,该方法在 F 值上提升约 9%。

**关键词** 视频字幕,检测和提取,帧间差,动态特征,Sobel 边缘检测

**中图法分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.009

### Two-stage Method for Video Caption Detection and Extraction

WANG Zhi-hui<sup>1</sup> LI Jia-tong<sup>2</sup> XIE Si-yan<sup>2</sup> ZHOU Jia<sup>2</sup> LI Hao-jie<sup>1</sup> FAN Xin<sup>1</sup>

(Department of International Information and Software Technology, Dalian University of Technology, Dalian, Liaoning 116621, China)<sup>1</sup>

(Department of Software Technology, Dalian University of Technology, Dalian, Liaoning 116621, China)<sup>2</sup>

**Abstract** Video caption detection and extraction is one of the key technologies for video understanding. This paper proposed a two-stage approach which divides the process into caption frame and caption area, improving the caption detection efficiency and accuracy. In the first stage, caption frame detection and extraction is conducted. Firstly, the motion detection is performed according to the gray correlation frame difference, the captions are judged initially, and a new binary image sequence is obtained. Then, according to dynamic characteristics of ordinary captions and scrolling captions, the new sequence is screened two times to get caption frame. In the second stage, caption area detection and extraction is conducted. Firstly, the Sobel edge detection algorithm is used to detect the caption region, and the background is eliminated according to the constraint height. Then according to the aspect ratio, the vertical and horizontal captions are distinguished, and all captions in the caption frame can be obtained, including static captions, ordinary captions and scrolling captions. This method reduces the frames which need to be detected and improves caption detection efficiency by 11%. The experimental results show that the proposed method can approximately improve the F score by 9% compared with the methods of separately using the gray correlation frame difference and edge detection.

**Keywords** Video caption, Detection and extraction, Gray correlation frame difference, Dynamic characteristics, Sobel edge detection

## 1 引言

视频字幕检测是计算机视觉领域的一个经典研究课题。视频字幕可分为静止字幕、滚动字幕和普通字幕 3 类,由于字幕出现的位置和方式难以预测,因此如何检测并提取出视频中的 3 种字幕是一大挑战。根据检测字幕时是否利用了时序信息,可将现有文献中的视频字幕提取工作分为两大类:处理

帧序列得到字幕检测结果;独立地处理每一帧得到字幕检测结果。

对于第一类,基于 ICA 的视频字幕提取方法<sup>[1]</sup>去除了信号间的相关性,简化了独立分量的提取。该方法通过在视频帧图像中计算独立成分得到字幕,减小了背景抖动造成的影响,能较好地处理相同背景下出现不同字幕内容或相同字幕内容时的运动情况,但是无法检测静止字幕。基于多帧结合

到稿日期:2017-10-24 返修日期:2017-12-11 本文受国家自然科学基金(61472059,61772108)资助。

王智慧(1982—),女,博士,副教授,CCF 会员,主要研究领域为计算机视觉、图像安全,E-mail:zhwang@dlut.edu.cn;李佳桐(1996—),女,主要研究领域为图像处理、深度学习;谢斯言(1996—),女,主要研究领域为数字媒体;周 佳(1996—),女,主要研究领域为软件工程、金融与信息  
服务;李豪杰(1973—),男,博士,教授,CCF 会员,主要研究领域为计算机视觉、多媒体信息检索,E-mail:hjli@dlut.edu.cn(通信作者);樊 鑫  
(1977—),男,博士,教授,CCF 会员,主要研究领域为计算机视觉与图像处理、医学影像分析,E-mail:xin.fan@ieee.org。

的视频标题检测算法<sup>[2]</sup>根据时间序列对多帧图像进行最小(或最大)像素值搜索,然后以块为单位进行文字与非文字的分类,最后通过迭代文字区域分解,更精确地定位文字区域的边界。但是,该方法只能检测视频中后期添加的静止标题文字。

对于第二类,Zhao 等<sup>[3]</sup>提出了一种基于 Harris 角点的方法。他们首先检测出每一帧的角点图,经过形态学膨胀后,再使用一些特殊的特征把非字幕区域过滤掉,从而达到字幕检测的目的。但是,对于复杂的背景,这种方法很难保证精确度。Sato 等<sup>[4]</sup>提出了一种基于过滤器的方法,他们设计了 4 个过滤器,分别针对 4 个方向即水平、垂直和两个对角来检测此方向上是否有突变。但是,该方法只能针对一定的语言和字体,扩展性不大。

基于以上原因,为提高检测效率和准确率,本文创新性地将字幕帧和字幕区域分开检测,提出了一种两阶段的字幕检测和提取算法。第一阶段进行字幕帧检测:首先,根据帧间差算法进行运动检测,对字幕进行初步判断,并得到新的二值化图像序列;然后,根据普通字幕和滚动字幕的动态特征对新序列进行二次筛选,得到字幕帧。第二阶段对字幕帧进行字幕区域的检测和提取:首先,利用 Sobel 边缘检测算法初检文字区域;然后,根据高度特征剔除其背景,根据宽高比区分出纵向字幕和横向字幕,得到字幕帧中的所有字幕,即静止字幕、普通字幕、滚动字幕。本文算法的流程如图 1 所示。

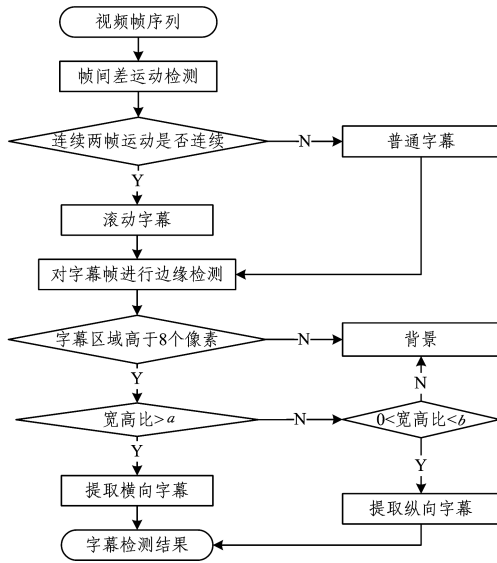


图 1 所提算法的流程图

Fig. 1 Flowchart of proposed algorithm

## 2 字幕帧检测

此阶段的主要目标是将含有字幕的帧提取出来,主要用到了帧间差算法和字幕动态特征相结合的思想。首先,需要对整体视频进行阈值化处理<sup>[5]</sup>,并且建立阈值化后的图像与原始彩色图像的映射。然后,通过帧间差算法对视频进行运动检测,得到各类字幕及一些背景噪声<sup>[6]</sup>所在的帧序列。最后,需要对字幕帧进行进一步的判断,即在处理过的帧序列上,将出现字幕的帧的帧与前一帧进行比较:

1)如果连续两帧都有字幕出现,则定位滚动字幕出现的区域,并判断滚动字幕,一直比较到后一帧无字幕出现时此帧

便为滚动字幕结束的帧,此时需要对滚动字幕出现的每一帧进行提取。

2)如果出现字幕的帧不连续,则将其判断为普通字幕,此时只需提取普通字幕出现的第一帧。

最后,提取已定位好的字幕帧所对应的原始帧,并将其存储为彩色图像格式。

### 2.1 查找字幕事件

帧间差算法是一种通过对视频图像序列的连续两帧图像进行差分运算来获取运动目标轮廓的方法。当监控场景中出现目标运动时,相邻两帧图像之间会出现较为明显的差别,将两帧相减以求得图像对应位置像素值差<sup>[7]</sup>的绝对值,判断其是否大于某一阈值,进而分析视频或图像序列的物体运动特性,如式(1)所示:

$$D(x,y) = \begin{cases} 1, & \text{if } |I(t) - I(t-1)| > T \\ 0, & \text{others} \end{cases} \quad (1)$$

其中, $D(x,y)$ 为连续两帧图像之间的差分图像, $I(t)$ 和 $I(t-1)$ 分别为  $t$  和  $t-1$  时刻的图像, $T$  为对差分图像进行二值化时选取的阈值, $D(x,y)=1$  表示前景, $D(x,y)=0$  表示背景。

由于字幕也属于视频中运动的目标,此次利用帧间差对视频中的字幕进行了运动检测,可以粗略判断出字幕的出现事件。但是,在帧间差的初步判断中包含了所有可能的字幕以及一些背景,因此需要对字幕帧进行进一步的判断。

### 2.2 字幕帧类型的判定

对字幕进行进一步的判断时需要区分普通字幕和滚动字幕。

普通字幕的特点是:突然出现,一段时间内存在于屏幕上,字幕内容不发生改变,之后消失。针对此特点,对帧间差算法处理后得到的新的两帧进行比较,如果第一帧检测出运动,第二帧未检测出运动,即符合以上特点,说明该运动物体是普通字幕。此时,只需要提取检测出运动帧的对应原始帧,即为普通字幕出现的第一帧。

如图 2 所示,左侧为帧间差算法处理后得到的连续 3 帧中的第一帧,是纯黑色图片,表示没有运动出现;中间的第二帧是带有字幕边缘的图片,说明有字幕出现;右侧的第三帧为纯黑色图片,说明没有运动出现。由此,可以判断普通字幕出现的事件,即连续 3 帧中的第二帧对应的原始帧为普通字幕出现帧。



图 2 普通字幕连续 3 帧的检测结果

Fig. 2 Three consecutive detection results for ordinary caption

滚动字幕的特点是:突然出现,一段时间内存在于屏幕上,字幕内容发生改变,之后消失,且位置较为规律,可以预判。针对此特点,可将滚动字幕定位在屏幕距左右边缘  $0-X$ ,距上下边缘  $0-Y$  的区域, $X$  为屏幕宽度的  $1/5$ , $Y$  为屏幕高度的  $1/5$ 。将通过帧间差算法处理后得到的新的两帧进行比较,在此区域内如果第一帧检测出运动,第二帧也检测出运动,说明该运动物体是滚动字幕。之后需要判断滚动字幕的

消失帧,即将前后两帧不断进行比较,直到检测不到运动的帧,即为滚动字幕消失的帧。对于滚动字幕,我们需要对滚动字幕出现的每一帧都进行提取。

如图3所示,左侧为帧间差算法处理后得到的连续3帧中的第一帧,是纯黑色图片,表示没有运动出现;中间的第二帧是带有字幕边缘的图片,说明有字幕出现;同理,右侧的第三帧有字幕出现。由此,可判断滚动字幕的出现事件,即连续3帧中的第二帧对应的原始帧为滚动字幕出现帧。



图3 滚动字幕连续3帧的检测结果

Fig. 3 Three consecutive detection results for scrolling caption

综上所述,本文进行字幕帧检测的具体流程如下:

- 1)对原始帧进行阈值化得到二值化图像;
- 2)利用帧间差算法找到新二值化图像序列,并将其与原始帧进行映射,从而找出对应关系;
- 3)根据动态特征区分有字幕帧和无字幕帧;
- 4)将新二值化图像序列的连续两帧进行比较,连续两帧运动初步定义为滚动字幕,非连续两帧运动为普通字幕;
- 5)记录普通字幕的第一帧;
- 6)对初步定义为滚动字幕的帧进行定位,并剔除复杂背景干扰;
- 7)对滚动字幕出现帧进行连续两帧比较,找出滚动字幕结束帧,并记录所有帧;
- 8)提取对应的原始帧。

经过以上8个步骤,最终得到了含有字幕的帧。

通过对普通字幕和滚动字幕的判断,排除了大部分的背景干扰,得到了明确含有字幕的帧,减少了字幕提取所需检测的帧数,方便了下一步的字幕提取。

### 3 对关键帧字幕的提取

#### 3.1 利用 Sobel 算法进行边缘检测

首先对取得的关键帧进行二值化,然后对二值化图片进行边缘检测。对多种边缘算法(Roberts, Sobel, Prewitt, Canny 和 Log 等)进行实验比较,结果表明,Roberts, Sobel 和 Prewitt<sup>[8]</sup>算子的算法简单,但检测精度不高;Canny 和 Log 算子的算法复杂,但检测精度较高。因此,我们最后选择使用 Sobel 算法对关键帧进行初步处理。

#### 3.2 确定字幕区域

视频字幕具有如下特点:

- 1)视频文字与背景有明显的对比;
- 2)视频文字便于识别且有一定尺寸;
- 3)文字有大量的水平边缘和垂直边缘。

基于以上特点,在一定阈值范围内可将字幕与背景区分开,从而确定字幕区域。首先,对图片从下到上进行逐行的像素扫描,并依据同一行的每两个像素点之间的差值设立一个临时值,将这张图片的每个临时值的平方保存到一个二维长数组  $D$  中。由于字幕的边缘与背景边缘的区别较大,我们设定一个全局阈值  $TH$ (经多次实验取得),当  $D$  中的数值大于  $TH$  时,则在另一个二维数组  $A$  中保存该位置的像素,否则在

数组  $A$  中设定该位置的值为 0。

#### 3.3 特征判断

使用高度和宽高比两种标准对得到的字幕区域进行筛选。宽高比特征在文献[3]的基础上加以改进;高度特征由文献[9]提出。

##### 1)高度

为了进行更清晰的显示,视频中的英文字幕符必须高于 8 个像素,中文字幕符必须高于 20 个像素,边界窗的高度必须高于 10 个像素。这样,很多高度不满足条件的区域就会被过滤掉,从而可以更加精确地提取到正确的字幕帧。

##### 2)宽高比

宽高比即边界窗宽度和高度的比例。现在很多视频字幕检测都会忽略掉竖排字幕,因此我们根据宽高比设置一个判断,从而有效地检测出横排和竖排两种字幕。对于视频中水平方向的字幕信息,边界窗的宽高比越大,则越可能是字幕区域,定义水平字幕的宽高比  $w_1 > 1.3$ ;对于视频中竖直方向的字幕信息,边界窗的宽高比越小,则越可能是字幕区域,定义竖直字幕的宽高比  $0 < w_2 < 0.7$ 。

#### 3.4 输出结果

经过上文步骤的处理,可以得到字幕的位置及该位置的像素值,然后经过处理将其还原成彩色图片,并输出字幕图片。如果有多条字幕,则会得到多张字幕图片。由于是在整张字幕帧中进行字幕的提取,因此在字幕帧的检测过程中未判断出来的静止字幕也可以得到提取,如图4所示,使用该方法可将静止字幕“秒拍”提取出来。



图4 静止字幕的提取结果

Fig. 4 Static caption extraction results

### 4 实验结果

本文采用随机采集的数据集,对新闻、电影、电视剧、电视节目等各种不同时长、格式及类型的总计约 90 分钟的视频文件进行了实验。视频分辨率为 480P 至 1080P。从测试视频中共截取了 1761 张字幕帧并对其进行字幕提取,最终得到 3645 条字幕。

在实验中,将差分图像二值化时选取的阈值  $T$  设定为 10,将确定字幕区域时选取的全局阈值  $TH$  设定为 126。

选用 3 个指标来定量分析所提算法对视频字幕检测的效果:查全率、查准率和 F 值。各指标的定义分别如下:

$$\begin{aligned} \text{查全率} &= \text{正确检测的字幕数} / \text{视频中字幕总数} \times 100\% \\ \text{查准率} &= \text{正确检测的字幕数} / \text{检测出的区域总数} \times 100\% \\ \text{F 值} &= \text{查准率} \times \text{查全率} \times 2 / (\text{查准率} + \text{查全率}) \times 100\% \end{aligned}$$

#### 4.1 实验数据统计

将本文方法与单独使用帧间差判断字幕的方法<sup>[2]</sup>和单独使用边缘检测判断字幕的方法<sup>[10]</sup>进行比较,结果如表 1 所列。

表 1 字幕检测率的比较

Table 1 Comparison of caption detection rate (单位:%)

视频种类	本文字幕 F 值	运动检测字幕 F 值 <sup>[2]</sup>	边缘检测字幕 F 值 <sup>[10]</sup>
新闻	95.79	85.45	88.18
电视剧、电影	99.23	87.38	90.95
电视节目	91.92	83.67	87.46

本文对字幕帧和字幕提取的查全率和查准率等进行了整体统计,如表 2 和表 3 所列。对不同情况下字幕提取的查全率和查准率分别进行了数据统计,结果如表 4 所列。

表 2 字幕帧提取情况

Table 2 Caption frame extraction

总字幕帧数	提取字幕帧数	遗漏字幕帧数	错误提取帧数
1920	1761	172	13

表 3 字幕提取情况

Table 3 Caption extraction

字幕总数	提取字幕数	遗漏字幕数	错误字幕数
3713	3645	68	26

表 4 不同情况下字幕的查全率和查准率

Table 4 Detection recall and precision of caption in different situations

字幕检测情况	查全率	查准率
单个普通字幕	99.44	99.56
单个滚动字幕	98.13	98.45
外文字幕	99.44	99.56
多个普通字幕同时存在	98.03	98.98
多个滚动字幕同时存在	97.64	98.53
3 种字幕同时存在	98.02	98.67

#### 4.2 字幕提取结果

本文对新闻、电影、电视节目 3 类不同视频进行了实验。其中,新闻视频中一般含有静止字幕、普通字幕、滚动字幕 3 种,但是字幕的位置比较固定,一般在屏幕的下方或者上方,边界也很清晰;电影字幕通常只含有普通字幕,位置也固定在屏幕下方或上方,边界清晰,但是背景比较复杂;电视节目也会存在静止字幕、普通字幕、滚动字幕 3 种,但是字幕位置不固定,艺术字的字幕边界也无法很好区分,且背景复杂,字幕提取的难度最大。图 5 给出了不同情况下的字幕提取结果。

如图 5(a)所示,对于含有多种字幕的新闻视频,所提方法可以很好地将其检测出来,几乎没有遗漏;但是对于“泰州 24℃”这个字幕,由于左右两边的边缘都比较规则,图形也比较小,因此中间的图形没有被检测出来。

如图 5(b)所示,对于有复杂背景的视频,所提方法也可以将字幕检测出来;但是对边缘比较不规则的艺术字则不能得到很理想的结果,一些不规则的边缘被去掉了。

如图 5(c)所示,对于字幕颜色和背景部分颜色比较相近的视频,所提方法可以将字幕检测出来。从得到的结果来看,背景干扰对字幕提取基本没有产生影响。

如图 5(d)所示,对于含有多国语言的视频,因为不同语种的语言基本都符合字幕的基本特征,有较明显的边缘,与背景有较大的区分,所以所提方法能将各个国家语言的字幕都检测出来。

如图 5(e)所示,对于含有多种类型字幕的电视节目视频,所提方法可以将字幕检测出来。这种视频同时含有横向字幕和纵向字幕,且字幕位置和大小不固定。通过对纵向字幕的特殊分析,所提方法能够完整地将纵向字幕截取出来,其他各种大小和颜色的字幕也都被贴合边缘地截取了出来。



(a) 新闻视频



(b) 电视节目视频



(c) 双语电视剧视频



(d) 德语视频



(e) 日语视频

图 5 不同情况下的字幕提取结果

Fig. 5 Caption extraction results under different conditions

经过字幕事件判断、边缘判断和视频字幕抽取 3 个步骤,最后提取出含有字幕帧的图片。实验结果表明,所提方法能较好地判断出字幕位置,截取出的部分也是贴合字幕边缘的矩形框,并且对不同语言的字幕具有广泛的适用性,对不同种类的字幕均有较好的提取效果。其平均查全率和查准率分别为 98.13%,99.27%。

与 Zhao 等提出的一种基于 Harris 角点的算法<sup>[3]</sup>相比,所提算法的时间复杂度降低约  $1-p$ ,其中  $p$  等于字幕帧总数除以视频总帧数。证明如下:

假设  $N$  为该视频总帧数,  $T_1$  为本文算法的时间复杂度,  $T_2$  为对比算法的时间复杂度,  $k_1$  为检测某帧是否有字幕的时间复杂度,为  $O(1)$ ,  $k_2$  为对某帧进行字幕提取的时间复杂度,为  $O(1280 * 720)$ ,因为标准的 720P 视频的分辨率为  $1280 * 720$ 。综上,可得式(2)、式(3):

$$T_1 = k_1 * N + p * k_2 * N \tag{2}$$

$$T_2 = k_2 * N \tag{3}$$

由式(2)和式(3)推导出式(4):

$$\frac{T_2 - T_1}{T_2} = \frac{k_2 * N - k_1 * N + p * k_2 * N}{k_2 * N} = 1 - p - \frac{k_1}{k_2} \tag{4}$$

其中,  $\frac{k_1}{k_2} = \frac{O(1)}{O(1780 * 720)} \approx 0$ 。

因此,与对比算法相比,本文算法的时间复杂度降低约  $1-p$ 。

通过对实验选取的总计约 90 分钟的视频文件进行检测,可以得出本文算法在区分了滚动字幕和普通字幕的前提下

- 刘洲洲,王福豹.改进的离散混合蛙跳算法压缩感知信号重构及应用[J].吉林大学学报(工学版),2016,46(4):1261-1268.
- [5] TANG D Y, YANG J, DONG S B, et al. A lévy flight-Based shuffled frog-leaping algorithm and its applications for continuous optimization problems[J]. Applied Soft Computing, 2016, 49:641-662.
- [6] ZHAO F, ZHANG G Z. Shuffled frog leaping algorithm based on new search strategy [J]. Computer Applications and Software, 2015, 32(8):224-228. (in Chinese)  
赵芳,张桂珠.基于新搜索策略的混合蛙跳算法[J].计算机应用与软件,2015,32(8):224-228.
- [7] ZHANG Q, LI P C. Adaptive grouping chaotic cloud model shuffled frog leaping algorithm for continuous space optimization problems [J]. Control and Decision, 2015, 30(5):923-928. (in Chinese)  
张强,李盼池.自适应分组混沌云模型蛙跳算法求解连续空间优化问题[J].控制与决策,2015,30(5):923-928.
- [8] KAUR P, MEHTA S. Resource provisioning and work flow scheduling in clouds using augmented shuffled frog leaping algorithm[J]. Journal of Parallel & Distributed Computing, 2016, 101:41-50.
- [9] HORNG M H, LIOU R J. Multilevel minimum cross entropy threshold selection based on the firefly algorithm[J]. Expert Systems with Application, 2011, 38(12):14805-14811.
- [10] ZHANG X M, TU Q, KANG Q, et al. Grey wolf optimization algorithm with double-hunting modes and its application to multi-threshold image segmentation [J]. Journal of Shanxi University (Natural Science Edition), 2016, 39(3):378-385. (in Chinese)  
张新明,涂强,康强,等.双模狩猎的灰狼优化算法在多阈值图像分割中应用[J].山西大学学报(自然科学版),2016,39(3):378-385.
- [11] SAHOO P, WILKINS C, YEAGER J. Threshold selection using Renyi's entropy[J]. Pattern Recognition, 1997, 30(1):71-84.
- [12] ZHANG X M, YIN X X, TU Q. High-dimensional multilevel thresholding based on BBO with dynamic migration and salt & pepper mutation [J]. Optics and Precision Engineering, 2015, 23(10):2943-2951. (in Chinese)  
张新明,尹欣欣,涂强.动态迁移和椒盐变异融合生物地理学优化算法的高维多阈值分割[J].光学精密工程,2015,23(10):2943-2951.
- [13] SUGANTHAN P N, HANSEN N, LIANG J J, et al. Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization; Technical Report, KanGAL Report # 2005005[R]. Singapore: Kanpur Genetic Algorithms Laboratory, Nanyang Technological University, 2005.
- [14] LIANG J J, QU B Y, SUGANTHAN P N, et al. Problem definitions and evaluation criteria for the CEC 2015 competition on learning-based real-parameter single objective optimization; Technical Report 201411A[R]. Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China And Technical Report, Nanyang Technological University, Singapore, 2014.

(上接第 53 页)

除了不含字幕的帧,字幕检测效率约提高 11%。

**结束语** 本文提出的两阶段的视频字幕检测和提取算法能准确判断出字幕位置且贴合字幕边缘进行截取,有较高的查全率、查准率以及较广泛的实用性,适用于不同语言的字幕。所提算法能解决复杂背景的问题,对字幕的位置、大小也没有要求,但是对字幕边缘有一定的要求,即需要较为规整的字幕边缘;否则会将不规则的字幕如艺术字的边缘截取掉或者检测不出该字幕,从而导致字幕残缺过多信息而无法提取的问题,具有一定的局限性。如何进一步研究不规则边缘字幕的特征,更加完整地截取边缘不规则字幕,将是我们下一步的研究重点。

### 参 考 文 献

- [1] WANG G H, WANG Z, YANG Y M, et al. Detection and positioning of video captions based on ICA algorithm [J]. Journal of Xi'an Shiyong University (Natural Science Edition), 2011, 26(3):100-103. (in Chinese)  
王国红,王喆,杨永民,等.基于 ICA 算法的视频字幕检测与定位[J].西安石油大学学报(自然科学版),2011,26(3):100-103.
- [2] WANG R R, JIN W J, WU L D. A new algorithm for detecting video caption by using multi-frame combination [J]. Journal of Computer Research and Development, 2005, 42(7):1191-1197. (in Chinese)  
王蓉蓉,金万军,吴立德.一种新的利用多帧结合检测视频标题文字的算法[J].计算机研究与发展,2005,42(7):1191-1197.
- [3] ZHAO X, LIN K H, HU Y X, et al. Caption form corners: A novel approach to detect caption and caption in videos [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(2):243-255.
- [4] SATO T, KANADE T, HUGHES E K, et al. Video OCR: indexing digital news libraries by recognition of superimposed captions[J]. Multimedia Systems, 1999, 7(5):385-395.
- [5] ANGADI S A, KODABAGI M M. Text region extraction from low resolution natural scene images using texture features[C]// 2010 IEEE 2nd International Advance Computing Conference (IACC). 2010.
- [6] OUYANG P R, ZHANG W J, GUPTA M M. An adaptive switching learning control method for trajectory tracking of robot manipulators[J]. Mechatronics, 2006, 16(1):51-61.
- [7] SU Y X, PARRA-VEGA V. Global asymptotic saturated output feedback control of robot manipulators[C]// Proceedings of the 7th World Congress on Intelligent Control and Automation. 2008.
- [8] ZHAO X, LIN K H, FU Y, et al. Text from corners: A novel approach to detect text and caption in videos[J]. IEEE Transactions on Image Processing, 2011, 20(3):790-799.
- [9] LYU M R, SONG J Q, CAI M. A comprehensive method for multilingual video caption detection, localization and extraction [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(2):243-255.
- [10] LANG Y, ZHENG D. An Improved Sobel Edge Detection Operator[C]// IEEE International Conference on Computer Science and Information Technology. IEEE, 2010:67-71.