

# 基于台标识别的网络视频分类系统

刘 纓<sup>1,2</sup>, 尹 珩<sup>1</sup>, 戴文娟<sup>3</sup>

- (1. 北京市网信办, 北京 100062;  
2. 北京理工大学 管理与经济学院, 北京 100062;  
3. 讯飞智元信息科技有限公司, 安徽 合肥 230088)

**摘要:**为了快速准确地清理互联网平台上的非法视频,减少人力物力的投入,在研究非法视频特征的基础上,提出一套基于台标识别的网络视频分类系统。在分析不同来源非法视频台标特性的基础上,根据不同来源的非法视频所携带的台标特性不同,系统采用基于模板匹配和特征识别相结合的方法进行台标识别,以提升识别的准确率。同时采用多种图像增强技术降低分辨率、清晰度、背景等不确定因素对识别结果的影响,提升对非法视频的召回率以及召回的准确率。实验结果表明,该系统可通过筛选出确认安全视频和疑似非法视频来有效减少非法视频筛选中的人力投入。另外,采用FFMPEG解码库解码视频抽取关键帧,提高了系统处理效率,平均单个视频处理时间缩短30%。

**关键词:**公共安全;网络视频;台标识别;模板匹配;ORB特征识别

中图分类号:TP302

文献标识码:A

文章编号:1673-629X(2016)10-0087-05

doi:10.3969/j.issn.1673-629X.2016.10.019

## A Classification System for Internet Videos Based on TV Logo Recognition

LIU Ying<sup>1,2</sup>, YIN Hang<sup>1</sup>, DAI Wen-juan<sup>3</sup>

- (1. Cyberspace Administration of Beijing, Beijing 100062, China;  
2. School of Management and Economics, Beijing Institute of Technology, Beijing 100062, China;  
3. Iflytek Intelligent System Co., Ltd., Hefei 230088, China)

**Abstract:** In order to clean up the illegal videos on the internet rapidly and accurately and reduce the input of manpower and material resources, a video classification system based on TV logo recognition has been proposed. On the basis of analyzing illegal videos TV logo with different producer, the system adopts a method combined pattern matching and feature recognition for TV logo recognition to improve accuracy according to different TV logo feature of illegal video. At the same time, a variety of image enhancement techniques are adopted to decrease the influence of uncertain factors like resolution, definition and background on the identification results, improving the recall rate of illegal video and recall accuracy. Experimental results show that this system could effectively reduce the human input and material resource by screening out the security video and suspected illegal video. In addition, FFMPEG is used to decode and draw key frame, which can improve the efficiency at 30%.

**Key words:** public safety; internet videos; TV logo recognition; template matching; ORB

### 1 概述

随着互联网技术的发展,越来越多的视频通过网络传播,其中充斥着大量危害公共安全的不健康视频,而在海量的视频库中要将这些不安全视频挑出需要耗费大量的人力物力。而台标作为视频的一个重要标识,包含了视频来源、视频类别和节目取向等重要信息,近年来台标识别在视频检索的研究中受到了越来越

越多的重视<sup>[1-5]</sup>,也提出了一些针对视频台标检测的算法。将台标作为检索依据,根据视频来源可将网络视频分为三类:带有省市级卫视台标的绝对安全的视频、带有非法组织电视台台标的危险视频,以及不包含前两种台标、需要采用其他方法进一步判别的视频。依据台标将网络视频分类可以极大减少在网络视频审查中的人力物力,保障互联网视频的健康、安全。

收稿日期:2015-05-29

修回日期:2015-09-17

网络出版时间:2016-09-18

基金项目:北京市科委项目(Z141100006014002)

作者简介:刘 纓(1973-),女,博士,研究方向为科学与管理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160918.1707.008.html>

视频台标通常由图像、图像型字母、字母、汉字组合而成,如图 1 所示。就电视台标而言,很多台标具有很大的局部相似性,比如 CCTV 系列的所有台标、卫视台与对应的二级电视台台标,而很多非法组织的标识含有复杂图像和字母,如图 1(b)。为了分析视频来源,不仅要分辨出台标是否是合法的,还要准确识别出具体台标的名称。



图 1 常见台标

由于背景的影响,同一段视频不同帧的清晰度也会有比较大的差距,如图 1 所示。为了应对实际应用中的各种不理想情况,文中在结合图像配准和多种图像增强方法<sup>[6-8]</sup>(如方差法去背景、区域扩展、多幅图叠加等)降低背景的影响的基础上,采用模板匹配和特征点匹配相结合的方法识别台标,同时采用 FFMPEG 关键帧抽取技术降低系统的时间消耗。该系统在保证时间效率的基础上可有效分离卫视台标与二级子台的台标、识别低分辨率视频中的非健康台标。

## 2 台标识别系统总体设计

根据目前的已有数据显示,中国大陆电视台的台标均位于屏幕的左上角,部分港澳台的电视台台标位于屏幕的右上角,而非法组织的标志通常分布在视频的四个角,根据已有视频资源,统计各非法台标的分布规律,从而配置各个区域的需要识别的台标模板。

### 2.1 系统结构

系统总体结构如图 2 所示。该系统包含模板资源制作、资源加载、台标识别与结果处理等模块,系统功能除了将网络视频分为三类集合(即正常视频集合、危险视频集合以及需要进一步甄别的视频集合)以外还承担收集台标模板的功能。对于客户端不能识别的台标将会上传至云端,由后台工作人员审核处理后,由云端更新资源。对于系统没有召回的视频,将继续采用人脸检索、视频拷贝等视频检索的方法继续识别。

### 2.2 资源制作

资源模板应尽可能应对由于视频信号、清晰度、视频比例等的差异导致的台标位置、形态偏差,同时降低非目标台标带来的干扰,因此需要收集目标台标的各种版本以及干扰台标。该系统模板的获取主要分两个途径:一个是在前期的台标模板通过选取清晰稳定的

视频,将其缩放至指定尺寸后,截取台标所在区域,并将背景替换为纯色,以减少背景对识别的影响;另一个是在后期系统上线后会传回大量不能识别的图片,经后台工作人员审核后加入到模板库中。

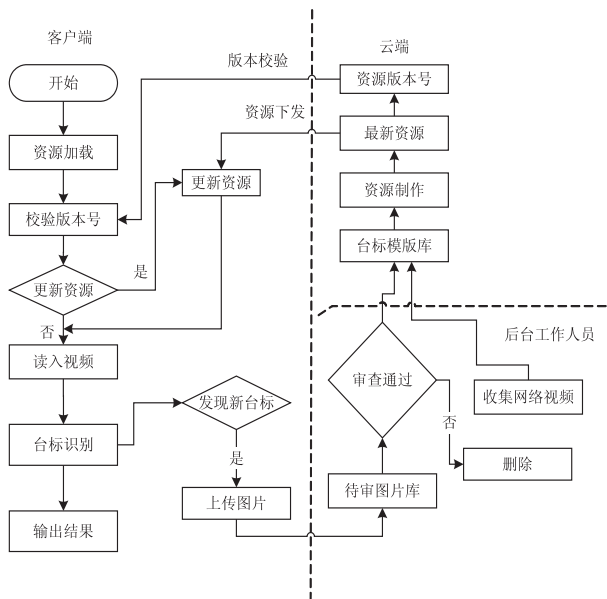


图 2 系统总体设计图

根据先验知识,大陆地区省市级电视台的台标均位于视频的左上角,对样本库中的 4 000 余条非法视频危险台标进行统计,包含 250 余种非法台标,视频中非法台标的分布规律统计结果显示,约 60% 的非法视频台标位于右下角,29% 的非法视频台标位于右上角,另外分别有 9% 位于左上角和 2% 位于左下角。

统计电视台标和非法台标的分布规律,根据分布规律统计结果和各台标本身性质配置各监测区域以及识别算法对应的模板特征。

## 3 系统台标识别算法

根据模板特性制作台标的梯度模板、边缘模板和特征资源库,采用多种图像切割算法<sup>[9]</sup>切割观测视频监测区域图像,计算其梯度模板、边缘模板和特征点集合,与可信任电视台标、非法组织的复杂台标模板库分别比对,完成视频的分类。

### 3.1 梯度直方图法

由于台标背景变化大,通过提取台标位置的图像特征时,背景特征掺杂其中,容易造成分类器识别结果错误。

为提高抗背景干扰能力,去除无关杂点的影响,该系统将传统的统计区域内梯度方向直方图转换为统计匹配点的梯度方向直方图,并将传统模板匹配的 0-1 匹配转换为梯度方向的 0-7 匹配。抵抗背景干扰的同时增强子台的可区分性,图 3 为安徽卫视及其各子台台标的梯度方向图。

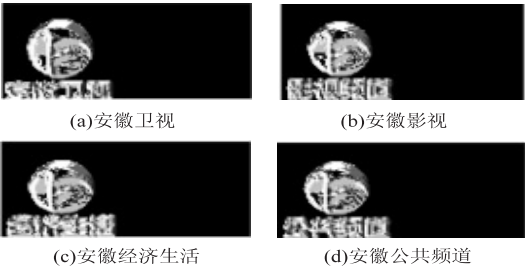


图 3 台标梯度方向模板

图像  $I(x,y)$  可看作二维离散函数,图像梯度和梯度方向的定义如下:

$$\begin{cases} dx = I(x+1,y) - I(x-1,y) \\ dy = I(x,y+1) - I(x,y-1) \\ ori(x,y) = \arctan\left(\frac{dy}{dx}\right) \end{cases} \quad (1)$$

其中,  $dy$  表示像素点  $I(x,y)$  垂直方向相邻点色度值的差;  $dx$  表示像素点  $I(x,y)$  水平方向相邻点色度值的差;  $ori(x,y)$  的大小范围为  $[0,2\pi]$ ,将台标梯度方向按  $[15\pi/8,2\pi] \cap [0,\pi/8]$ 、 $[\pi/8,3\pi/8]$ 、 $[3\pi/8,5\pi/8]$ 、 $[5\pi/8,7\pi/8]$ 、 $[7\pi/8,9\pi/8]$ 、 $[9\pi/8,11\pi/8]$ 、 $[11\pi/8,13\pi/8]$ 、 $[13\pi/8,15\pi/8]$  划分,分别标注为 0~7,见图 4。

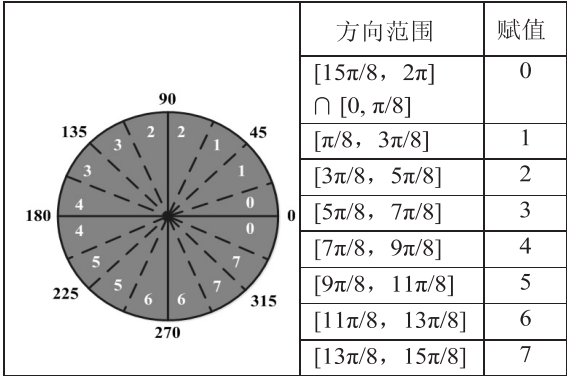


图 4 梯度方向划分

如果观测图与模板图对应的像素点梯度方向在同一个方向区域即标注结果相同,则认为该点匹配成功。

统计区域内梯度方向匹配成功的像素点数得到匹配直方图  $p$ ,与之对应的模板直方图记为  $q$ ,两个直方图的 Bhattacharyya 距离为:

$$D_B(p,q) = -\ln(B_c(p,q)) \quad (2)$$

其中,  $B_c(p,q) = \sum_{x \in X} \sqrt{p(x)q(x)}$  是 Bhattacharyya 系数。

选取最优匹配作为结果,若识别得分低于阈值,则上传视频台标至云端,待后台人员审核后上传到待审图片库中的新台标,加以标注后加入到台标模板序列中。

3.2 边缘模板匹配

图像边缘是指图像局部强度变化显著的部分,是

图像分割和形状特征等图像分析的重要基础,常用的检测算子有 Roberts 算子、Sobel 算子、Prewitt 算子、Log 算子等。Canny 算子可有效抑制噪声,采用双阈值精确确定边缘的位置,Canny 边缘计算流程为:

(1) 像素点的梯度计算。

图像函数为  $I(x,y)$ ,  $x,y$  是像素坐标,像素点梯度为:

$$\begin{aligned} \text{GradX} &= (I(x-1,y-1) - I(x-1,y+1)) + \\ &\quad 2(I(x,y-1) - I(x,y+1)) + \\ &\quad (I(x+1,y-1) - I(x+1,y+1)) \\ \text{GradY} &= (I(x+1,y-1) - I(x-1,y-1)) + \\ &\quad 2(I(x+1,y) - I(x-1,y)) + \\ &\quad (I(x+1,y+1) - I(x-1,y+1)) \\ \text{Mag} &= \sqrt{|\text{GradX}|^2 + |\text{GradY}|^2} \end{aligned} \quad (3)$$

(2) 边缘点判定。

采用双阈值,梯度大于高阈值的点确定为边缘点,梯度小于低阈值或者梯度取到极大值的点确定为非极值点,介于高低阈值之前的点通过边缘的连续性判断是否是边缘点,得到边缘二值图,统计边缘点数  $N_1$ 。

$$\text{Edge} = \begin{cases} 2 & \text{Mag} < \text{LowThre or Mag is maximum} \\ 1 & \text{Mag} > \text{HighThre} \\ 0 & \text{Mag} < \text{HighThre \& Mag} > \text{LowThre} \end{cases} \quad (4)$$

台标边缘模板包含边缘点个数  $N_2$ ,将观测图像的边缘二值图与台标边缘模板做与运算,设与运算后得到边缘点数为  $N$ 。匹配得分为:

$$\text{Score} = 0.5 * (N/N_1 + N/N_2) \quad (5)$$

选得分最高作为结果,若结果得分低于阈值,则上传截图至云端。

3.3 特征识别

非法视频的清晰度、台标位置、台标形态等都非常不稳定,部分非法视频是由手机录制后再二次编辑,分辨率很低,采用模板对台标位置与比例进行暴力匹配非常耗时,且效果也不理想。特征将图像映射为局部特征向量集,将图像的匹配转化为特征点间的匹配,可以抵抗平移、缩放和旋转,对光照变化、仿射及投影变换也有一定不变性,而且提取特征点数量较多,以量取胜,可抵抗一定的背景变化。

常用特征识别算法包括 Sift<sup>[10]</sup>、Surf<sup>[11]</sup>、Orb<sup>[12]</sup>等,其中 Orb 算法是最具效率的一个,效率比 Sift 算法高两个数量级。Orb 算法实际是具有方向性的 fast 角点检测<sup>[13]</sup>加上可旋转的 Brief 特征点描述<sup>[14]</sup>的方法组合,以改进 Brief 不具有旋转不变性和不具有多尺度等特性。



FAST 算法检测关键点,通过对比像素周围的  $N$  个点的灰度值,如果有连续  $M$  个点的灰度值与之不同,则认为该点是关键点。在原 ORB 的方案中是通过矩(moment)来计算特征点的主方向,而在该系统台标识别的应用中,由于网络视频常见的变形时挤压、拉伸和缩放,并不存在图像旋转,并且在经过实验对比后确认忽略关键点的方向更有利于提高检测效果。

在图像上选择一个局部区域,用  $p$  表示,它的大小是  $S * S$  像素,在  $p$  上面提取 Brief 特征。通过高斯分布选取点对,并用特征点的主方向校正点对的坐标,定义 tao 测试:

$$\tau(p;x,y)=\begin{cases}1 & p(x)<p(y) \\ 0 & p(x)\geq p(y)\end{cases}\tag{6}$$

其中,  $x,y$  是  $p$  内的两个像素位置,是形如  $[u,v]$  的二维坐标;  $p(x)$  和  $p(y)$  是  $x$  和  $y$  的灰度。

得到的二进制向量如下:

$$f_n(p)=\sum_{1\leq i\leq n}2^{i-1}\tau(p;x_i,y_i)\tag{7}$$

由于背景的影响,orb 算法会有一些错配点,在特征匹配前通过图像增强降低背景的干扰至关重要。

4 系统优化

4.1 解码与抽帧策略

网络视频,特别是非法组织制定的非法视频都经常是在原有视频的基础上经过多次编辑后的结果,经过删帧、插帧、转码等一系列操作之后,台标变得不稳定,比如某些时间段不含有台标或者台标不够清晰,部分帧背景嘈杂易被误识为图像台标,另外对于电视视频来说,不够清晰的二级子台会被误识为卫视台召回。如果这种不够质量的帧被抽取作为测试样本,会严重影响识别结果。

为了降低抽帧带来的影响,该系统对视频分段做多次识别,返回统计结果,可有效降低虚警发生的概率,且可以规避由于台标不稳定而造成的漏警。系统同时会缓存多张图片,多次检测势必会造成效率的负担,于是采用 FFMPEG 解码库对视频解码并按照设定的抽帧策略抽取关键帧,可极大提高视频处理效率。

4.2 图像增强

为了保证截取的图像识别区域包含整个台标,因此截取的区域远大于台标实际所占的面积,而这也带来了新的问题:嘈杂的背景对于台标识别产生巨大的影响。为了对抗视频背景的影响,系统采用一系列的图像增强技术,包括多张图像叠加去背景、方差法去背景和区域扩展等。

去背景的思想是基于同一段视频中台标的位置是相对固定,而背景是一直在变化这一前提。在视频

中以 10 s 为间隔抽取三帧图像,截取台标区域位置,分别计算边缘二值图,边缘图叠加是多幅边缘图按照权重叠加,叠加过程如图 5 所示,经过叠加之后背景部分的强度明显降低。

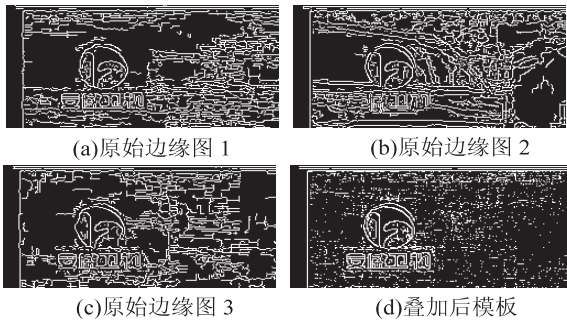


图 5 多幅图权重叠加去背景效果展示

方差法去背景是依据视频中台标具有时空不变性这一特点,截取目标区域,遍历区域图像中的每一个像素,计算每个像素点的方差,判断当前点是否属于背景点,另外在方差法去背景的基础上进一步采用区域扩展等图像增强方法去除背景的影响。

5 系统效果分析

模板库说明:模板库中包含了 CCTV1 ~ CCTV15、各省级卫视及二级子台共 174 个电视台台标,加上 232 个非法组织的图形台标,模板库中共 434 个梯度模板、103 个边缘模板和 1 189 个 orb 特征模板。

测试集说明:测试集合分为三部分,CCTV 和省级电视台的视频、带有非法台标的危险视频及本次测试的干扰集,其中负样本包含了非省级电视台的视频、带有视频网站 logo 的视频、不含有台标的网络视频共计 16 596 个。

测试机器配置 CPU: Interl Xeon E5 - 2620 @ 2 GHz、64 GB 内存,采用单线程测试,16 596 条视频采用 OpenCV 解码和 FFMPEG 解码的耗时分别为 77 337 s 和 53 938 s,平均单个视频处理时间由 4.66 s 缩短至 3.25 s,效率优化 30.26%。对 FFMPEG 解码技术采用 24 线程测试上述集合,共耗时 6 432 s,平均单个视频耗时 0.38 s,完全可以达到项目需求。

由于电视台标和非法电视台台标所用识别算法不同,且互为干扰,分别统计算法优化前和算法优化后两类台标识别的指标,如表 1 和表 2 所示。

表 1 电视台标测试结果

	优化前		优化后	
	判正(T)	判负(F)	判正(T)	判负(F)
正样本	2 568	558	2 982	144
负样本	83	13 387	32	13 438
总计	2 651	13 945	3 014	13 582

表 2 非法台标测试结果

	优化前		优化后	
	判正(T)	判负(F)	判正(T)	判负(F)
正样本	2 216	1 098	2 876	438
负样本	121	1 3161	38	1 3244
总计	2 337	14 259	2 914	13 682

由表 1 计算可得,优化前召回率和正确率分别为 82.15% 和 96.87%,而优化后召回率和正确率分别为 95.40% 和 98.94%;由表 2 计算可得,优化前召回率和正确率分别为 66.87% 和 94.82%,而优化后召回率和正确率分别为 86.78% 和 98.70%。

对于电视台标,虚警主要来自于卫视台对应的二级子台的视频,相似性比较大,导致虚警的产生。为了保证正确率,系统保持了比较高的阈值,代价是降低了召回率。对于非法台标来说,导致召回率不够理想的原因是视频样本本身质量和清晰度参差不齐,部分非法视频的分辨率只有 128 \* 88,另外部分非法视频使得剪切拼接严重,台标出现的位置以及时间很不稳定,导致去背景的方法失效。

6 结束语

台标标志着视频的来源和类别属性,是视频分类的一个重要依据,由于网络视频经过了多次编辑,造成视频台标与台标模板之间存在较大差异。文中在多种图像增强技术去除嘈杂背景影响的基础上,采用梯度、边缘模板识别与特征匹配的方法检测不同类别的台标,实现卫视台标与对应二级台标的分离以及低分辨率条件下非法台标的识别,达到可观的召回率和正确率。采用 FFMPEG 抽取关键帧缩短了视频解码抽帧的时间,使系统的效率满足需求。在后续的工作中为了进一步提高系统对低分辨率视频、台标位置偏离及文字型台标的处理能力,将继续扩大系统检测区域并降低背景对识别结果的影响。

参考文献:

[1] 包秀国,张冬明,张曦珊,等. 基于选择性 Gabor 滤波器组的网络视频台标识别[J]. 计算机辅助设计与图形学学报,

2014,26(2):248-257.  
[2] 金 阳,程江华,任 通,等. 一种基于二值图角点匹配的台标识别方法[J]. 电视技术,2012,36(17):15-19.  
[3] Wang J Q,Duan L Y,Li Z L,et al. A robust method for TV logo tracking in video streams[C]//Proceedings of IEEE international conference on multimedia and expo. Los Alamitos: IEEE Computer Society Press,2006:1041-1044.  
[4] Zhang L,Xia T,Zhang Y,et al. Hollow TV logo detection [C]//Proceedings of the 18th IEEE international conference on image processing. Los Alamitos: IEEE Computer Society Press,2011:3581-3584.  
[5] 侯胜伟,滕奇志,高明亮,等. 台标的自动检测与识别[J]. 四川大学学报:自然科学版,2013,50(3):522-528.  
[6] 汪文英,张冬明,张勇东,等. 利用仿射变换的快速空间关系验证[J]. 计算机辅助设计与图形学学报,2010,22(4):625-631.  
[7] Lettner M,Kleber F,Sablatnig R,et al. Contrast enhancement in multispectral images by emphasizing text regions[C]//Proc of the eighth IAPR workshop on document analysis systems. [s. l. ]:[s. n. ],2008:225-231.  
[8] Gorelick L,Boykov Y,Veksler O,et al. Submodularization for binary pairwise energies[C]//Proc of IEEE conference on computer vision and pattern recognition. [s. l. ]:IEEE,2014:1-8.  
[9] 张 伟,蒋 宏,任 章. 自适应多阈值图像分割算法[J]. 自动化技术与应用,2007,26(8):71-73.  
[10] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision,2004,60(2):91-110.  
[11] Bay H,Tuytelaars T,van Gool L. Surf:speeded up robust features[C]//Proc of European conference on computer vision. [s. l. ]:[s. n. ],2006.  
[12] Rublee E,Rabaud V. ORB:an efficient alternative to SIFT or SURF[C]//Proc of IEEE international conference on computer vision. [s. l. ]:IEEE,2011:2564-2571.  
[13] Rosten E,Drummond T. Machine learning for high-speed corner detection[C]//Proc of European conference on computer vision. [s. l. ]:[s. n. ],2006.  
[14] Calonder M,Lepetit V,Strecha C,et al. Brief:binary robust independent elementary features[C]//Proc of European conference on computer vision,[s. l. ]:[s. n. ],2010.