

基于镜头的鲁棒视频广告检测

张 亮¹,朱振峰¹,赵 耀¹,卢汉清²

(1. 北京交通大学 信息科学研究所,北京 100044;2. 中国科学院 自动化所模式识别国家重点实验室,北京 100080)

摘 要:随着多媒体技术的发展,自动检测出数字视频节目里面嵌入的广告是很具挑战性的研究.然而,由于嵌入的广告的制作方式和表现手法的多样性,很多自动检测模型的实验结果往往不甚理想.为了提高检测系统的鲁棒性,提出了 3 阶段广告检测系统.首先,提出了基于区域特征重要性的镜头检测算法(RBFI, region-based feature importance detection),实现视频播放中突变镜头和消隐镜头的检测,同时从每个镜头提取出一些统计特征用来标识镜头.然后,利用 SVM 的优异分类特性实现镜头分类.最后为了能得到精确的广告视频段,利用广告视频在内容和时间上的连续性来消除错分的镜头,然后将广告镜头整合成广告视频段.本系统在 30 个电视节目的片段上进行验证,实验结果表明此广告检测系统具有实用性.

关键词:视频广告检测;镜头检测;支持向量机

中图分类号:TP391 文献标识码:A 文章编号:1673-4785(2007)02-0083-06

Video commercial detection based on the robustness of shot

ZHANG Liang¹, ZHU Zhen-feng¹, ZHAO Yao¹, LU Han-qing²

(1. Institute of Information Science, Beijing Jiaotong Univesity, Beijing 100044, China; 2. Key Laboratory of Pattern Identification, Institute of Automation under the Chinese Academy of Sciences, Beijing 100080, China)

Abstract :Automatic detection of commercials embedded in digital video materials is a challenging task with the development of retrieval of multimedia data. However, because of the diversity of production modes and expression methods of commercials, the performances of current detection systems are inadequate. In this paper, a three-phase system for commercial detection is suggested to improve robustness of the detection system. Firstly, two Region-based Feature Importance Detection schemes are proposed to detect cut shots and dissolved shots respectively and some statistical features are also extracted to mark shots. Secondly, an SVM classifier is applied to classify these shots. Finally, in order to obtain commercial segments more accurately, a statistical comparison of time and content of commercials is used to eliminate falsely cut shots. Finally, these commercial shots are integrated into commercial video segment. Test results on 30 TV video segments show the effectiveness of the suggested system.

Key words :video commercial detection; shot detection; support vector machine

随着多媒体技术的发展,电视广告在日常生活扮演着日益重要的角色.广告自动检测系统以及相关技术的发展带来主要包括 2 个方面的应用.一方面,对普通电视节目感兴趣的人可以在视频节目里快速的定位广告并去除广告,从而可以提高视频节目的存储效率.另一方面,对于广告感兴趣的人可

以利用自动检测系统快速获得广告片段,从而可以节省他们分析广告制作技巧的时间.为了实现上述应用,人们提出了很多的广告检测策略.早期的研究通常集中在基于台标的检测^[1-2]和 black/ silent 帧方面的检测^[3].然而,由于现在的电视台在播放广告的时候通常不隐去台标,导致基于台标的检测方法失败,并且现在的台标也变得越来越复杂,检测起来十分不易.另外,在普通节目片段和广告片段转换时并不总是存在 black/ silent 帧,甚至 black/ silent 帧也可以为了某种剪辑需要随机的插入,这直接导致

收稿日期:2006-10-26.
基金项目:国家自然科学基金资助项目(60373028、90604032、60602030);教育部博士点专项基金资助项目(20030004016).

基于 black/ silent 帧检测方法的失败. 为了避免上述问题, 人们又提出了基于镜头的检测方法^[4-5]. 此方法通过从镜头中提取一些能够表示广告片段的特征, 然后利用这些特征将电视镜头分成普通节目镜头和广告镜头. 但是这些方法通常只是简单的分类, 而没有考虑如何消除错分广告镜头的影响, 同时也没有考虑如何合并广告镜头得到广告片段的问题. 而且, 现有的很多方法都没有考虑广告内容的一致性, 从而很难改善检测的效果. 另外, 上述方法即使在检测切变镜头方面取得很好的效果, 但当检测消隐或者淡入淡出镜头时就会遇到问题. 除了上述的检测方法, 基于数据库的广告检测方法也被采用^[6], 数据库存储预先定义的广告视频段的特征, 然后利用此数据库识别嵌入在电视节目里面的广告段. 然而, 此方法的缺点是必需一个足够大的数据库来存储已知的广告特征, 因此不能检测出数据库中预先没有存储的广告视频段.

为了避免上述问题, 提出了一种 3 阶段的鲁棒性的广告检测系统. 首先, 本文提出了基于区域的区域特征重要性的镜头检测算法 (region-based feature importance detection, RBFID), 实现突变和消隐镜头的检测, 进而实现对电视节目的镜头分割; 同时提取出一些有效的镜头统计特征. 然后, 利用 SVM 实现镜头分类, 把它们分成普通节目镜头和广告镜头. 最后通过引入后处理机制, 实现对错分镜头的纠错, 并将广告镜头整合成广告视频段. 实验表明本文提出的广告检测系统能够很准确地检测出广告视频段. 图 1 是所提出的广告检测系统的工作流程.

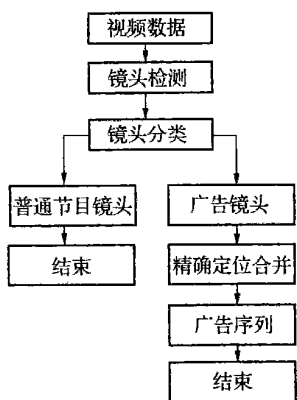


图 1 视频广告检测系统流程图

Fig. 1 The general diagram of the proposed commercial detection system

1 广告场景中的镜头检测

为在较短时间内传递尽可能多的信息, 广告视

频通常包含更多的镜头变换以吸引观众. 此外广告视频还显示独特的场景转换特征, 通常包括切变镜头和消隐镜头的转换. 因此在广告视频和普通视频里面的镜头变化的方式和频率是明显不同的.

1.1 切变镜头检测

切变镜头是视频分析中非常有效的特征, 它在视频序列中产生一个视频内容上的中断. 和普通的节目视频相比, 广告视频中出现切变镜头的频率更高. 由于在大部分的场景转换的过程中都是体现切变镜头的特征, 所以一般用检测效果很好的颜色或灰度直方图差值检测切变镜头^[7]. 但是这些检测方法通常忽略了广告视频的制作技巧. 也就是说, 在广告视频段中, 比较重要的信息通常都位于视频帧的中间部分. 因此, 如果连续 2 帧中间区域的内容发生突然变化而同时背景却保持不变的话, 应该认为它很有可能发生了切变. 基于这种理论, 提出了基于区域的切变镜头检测方法.

如图 2 所示, 每一帧首先被分成多个区域, 用以描述广告制作中的表现手法, 即在广告视频帧中不同的区域在传递信息过程中的重要性不同. $H(j, R_i, m)$ 表示在第 m 帧中的第 i 个区域的 B 阶的直方图, j 表示相应的阶数. 因此, 连续 2 帧间 (第 m 帧和第 $m-1$ 帧) 场景切换的帧差 FD_m 的计算如下:

$$FD_m = \sum_{i=1}^n w_i \sum_{j=0}^B \frac{|H(j, R_i, m) - H(j, R_i, m-1)|^2}{\max\{H(j, R_i, m), H(j, R_i, m-1)\}} \quad (1)$$

式中: w_i 表示归一化的权值, 用来标识相应区域 i 的重要性, 并且根据区域重要性的不同, 权值大小排序如下: $w_n < w_{n-1} < \dots < w_i < w_1$, n 表示区域标识. 在系统实际运行中, 本文采用 3 区域的划分, 并有 $w_1 = 0.5, w_2 = 0.3, w_3 = 0.2$.

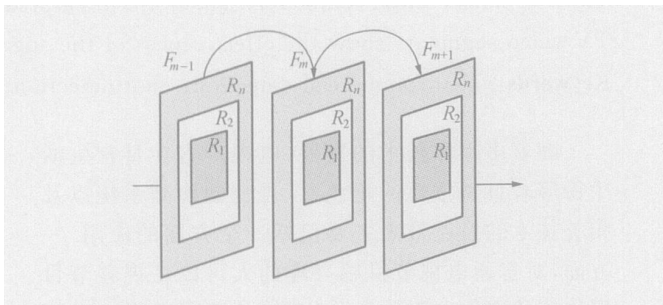


图 2 基于重要区域的切变镜头检测算法示意图

Fig. 2 Sketch map of hard cut shots detection based on important regions

假设一段电视节目里面的大部分场景的切换都是平滑的, 也就是说大部分的 FD_m 值 (除了发生切

变的位置)都是相对集中地落在一个固定的区间内的.由此依据 FD_m 的统计特性,可以近似地利用 Rayleigh 分布来描述:

$$R[FD] = \frac{ED^2}{2g} \exp\left[-\frac{ED^2}{2g}\right]. \tag{2}$$

式中: g 表示高斯噪声的方差,而且 $FD_m > 0$. 为了估计式(2)内的 g ,本文采用了计算复杂度较低的中值估计算法,即有

$$\hat{g} = \frac{M}{2} \sqrt{\frac{2}{\ln(2)}}. \tag{3}$$

式中: M 是 $R[FD_m]$ 的中值,并有

$$\int_0^M R[FD_m] dFD_m = 1/2. \tag{4}$$

图 3 表明了电视节目 FD_m 的实际分布和经估计得到的 Rayleigh 分布.从图中可以看出,估计得到的 Rayleigh 分布可以较好地拟合实际分布.

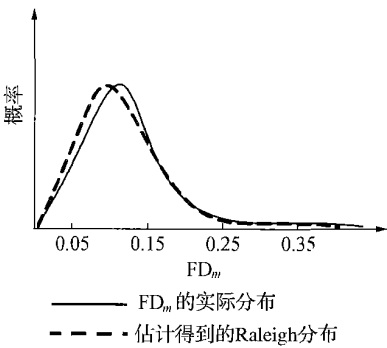


图 3 FD_m 的实际分布和估计得到的 Rayleigh 分布
Fig. 3 The real distribution and estimated Rayleigh distribution of FD_m

为了确定相邻的 2 帧是否发生切变,本文利用 FD_m 的估计分布的置信区间来计算得到阈值 t_{cut} .因此,如果第 i 帧的 FD_m 大于阈值 t_{cut} ,就认为在第 i 帧发生切变.

1.2 消隐镜头检测

除了切变镜头外,在广告制作手法中还往往采用另外一种很常见的技巧:消隐镜头.尤其在化妆品的广告中,消隐镜头更加普遍.消隐镜头是指前一个镜头的后几帧在灰度上逐渐变暗同时后一个镜头的前几帧在灰度上逐渐变亮,类似于淡出、淡入,不过在消隐中的淡出与淡入在时间上有一定的重叠.和切变镜头的检测比起来,消隐镜头的检测更加困难.

为能准确地检测到消隐镜头,本文提出了基于局部特征时变一致性的检测算法.对于消隐镜头,在一定时间内连续的帧中,平均灰度值的变化是连续的,并且这种单调性的变化通常持续一段时间.因此,本文利用这种变化的累积值来实现消隐镜头的

检测.如图 4 所示, $R_{m,i}$ 表示第 m 帧中的第 i 个区域.

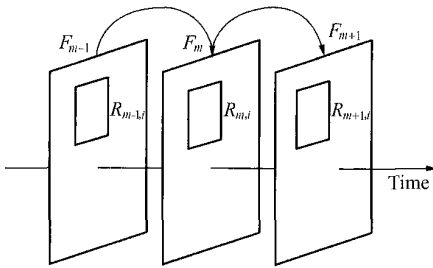


图 4 基于局部特征时变一致性的消隐镜头检测方法
Fig. 4 Sketch map of dissolve shot detection based on local coherent temporal change

设 V_i^m 表示第 m 帧中的第 i 个区域 $R_{m,i}$ 的平均灰度值.为了叙述方便, S_i^m 表示第 m 帧中的第 i 个区域的灰度变化趋势. T_i^m 表示在第 m 帧之前的相应的第 i 个区域具有单调灰度变化模式的连续帧的数目. N_h^m 表示在第 m 帧的所有区域的统计特性,即属于区间 $[n_1, n_2]$ 内的 T_i^m 的数目.基于局部特征时变一致性的检测算法流程如下:

- 1) 计算样本区域 V_i^m 的平均灰度值.
- 2) 初始化: $T_i^1 = 0, S_i^1 = +1$.
- 3) 如式(5)、(6)更新第 n 帧中的每一个区域的 S_i^m 和 T_i^m :

$$S_i^m = \begin{cases} 1 & V_i^m - V_i^{m-1} > 0, \\ -1 & V_i^m - V_i^{m-1} < 0. \end{cases} \tag{5}$$

$$T_i^m = T_i^{m-1} + \frac{1 + S_i^{m-1} \cdot S_i^m}{2}. \tag{6}$$

- 4) 计算 N_h^m :
- $$N_h^m = \# \{ T_i^m \mid T_i^m \in [n_1, n_2] \}. \tag{7}$$

式中: $\#$ 表示属于区间 $[n_1, n_2]$ 的数目.

图 5 给出了在一段测试视频的实验结果.从中可以看出,此段广告视频包含 3 段消隐镜头,分别位于 2 950 ~ 2 980 帧,3 320 ~ 3 350 帧和 3 390 ~ 3 410 帧内,并且和实际的广告视频段相吻合.在此 3 段镜头内, N_h^m 值的变化趋势是先增加然后减小,这和普通节目内部的 N_h^m 的变化趋势有明显的区别.

2 基于 SVM 的镜头分类

V. Vapnik^[8]提出的支持向量机(support vector machine, SVM)方法是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的,根据有限的样本信息在模型的复杂性和学习能力之间寻求

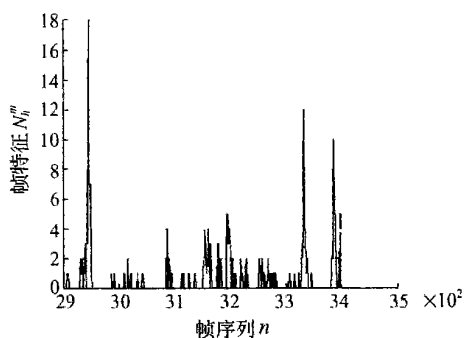


图5 消隐镜头的统计图

Fig. 5 Effectiveness of the statistical values N_n^m for characterizing the dissolve cuts

最佳折衷以期获得最好的推广能力. 支持向量机理论因其坚实的理论基础和诸多良好特性在近年获得了广泛的关注. 其原理是首先通过非线性变换将输入空间变换到一个高维空间, 然后在这个新空间中求取最优线性分类面. 经过优化求解, 该最优分类超平面由式(8)给出:

$$S(x) = \sum_{i=1}^n y_i \phi(x_i) \cdot \phi(x) + b. \quad (8)$$

式中: $\phi(\cdot)$ 是一个非线性的映射函数, $X = \{x_i \in R^d\}_{i=1, \dots, n}$ 和 $y_i = \{-1, 1\}$ 分别表示训练集和相应的类标识, b 是偏移量. 其中任何一个对应于一个非零值 y_i 的样本 x_i 被称为支持向量. 一般来说, 采用不同的核函数将对最终求得的最优分类超平面产生影响, 在本文中采用高斯核函数 $K(x, y) = \phi(x) \cdot \phi(y) = \exp\left(-\frac{\|x - y\|^2}{2}\right)$. 因此, 对于一个测试样本, 其最终判别输出为

$$L(x) = (S(x)) = \begin{cases} +1 & \text{if } S(x) > 0, \\ -1 & \text{if } S(x) < 0. \end{cases} \quad (9)$$

式中: (\cdot) 为一指示函数.

为了实现有效的镜头分类, 提取了一些鲁棒性的包括镜头切换频率、帧差的平均值和方差以及帧的主颜色空间占有率等在内的特征来对镜头进行描述. 另外, 在支持向量机的学习训练中, 训练集由 800 个镜头组成, 其中包括 380 个广告镜头.

3 广告镜头的重定界及合并

通过 SVM 分类, 不可避免地会出现错分的镜头. 为此本文还采用包括基于广告时间连续性的滑动窗口累积算法来精确定位广告边界以及合并不同的广告视频镜头的后处理方案, 而得到广告序列. 由于广告镜头内容上都是连续的而且它们成组地出现, 因此如果一个非广告镜头存在于连续的广告镜

头中间, 那么这个镜头极有可能是被错分的, 即它很可能是一个广告镜头.

假设 $\text{shot}_i (1 \leq i \leq n)$ 代表第 i 镜头. C_i 代表镜头的当前标识, shot_i 用来确定这个镜头是否为广告镜头. n 表示镜头的数目. 滑动窗口累积算法如下所示:

初始化 C_i :

$$C_i = \begin{cases} +1 & \text{如果第 } i \text{ 个镜头属于广告镜头,} \\ -1 & \text{其他.} \end{cases}$$

(10)

1) 利用包含 5 个镜头的滑动窗口来更新 C_i . 假设 $W = \{w_j | w_j = 1, \text{ if } -2 \leq j \leq +2\}$ 表示一个滑动窗口. $C_j (-2 \leq j \leq +2)$ 更新算法如下:

$$C_j = \begin{cases} C_j + 1 & \text{if } \sum_{k=-2}^2 C_{i+k} W_k \geq 3, \\ C_j - 1 & \text{else.} \end{cases} \quad (11)$$

2) 通过判断 C_i 来确定是否为广告镜头. 如果 shot_i 等于 +1, 镜头 shot_i 就认为是广告镜头. 否则, shot_i 就是普通节目镜头:

$$\text{shot}_i = \begin{cases} +1 & \text{if } C_i > 0, \\ 0 & \text{else.} \end{cases} \quad (12)$$

3) 如果在一次循环中存在多于 2 个镜头从广告镜头转变成普通节目镜头, 或者相反的情况出现, 则回到第一步从新开始. 如果小于 2 个, 则算法结束.

为了得到完整的广告视频段, 不同的广告镜头应该合并成数个相应的广告视频段. 合并算法应该遵循下列原则:

1) 如果多于 5 个广告镜头是连续的, 那么它们应该合并成一个广告视频段. 如果小于 5 个, 则定位距离最近的广告段. 如果在它们和这个广告段之间存在少于 3 个镜头, 则它们应该并入这个广告视频段.

2) 如果在两段连续的广告视频段之间存在少于 3 个的镜头, 那么这 2 个广告段和它们中间的这些镜头应该合并成一个完整的广告视频段.

3) 重复上面的合并步骤, 直到没有新的广告视频段出现.

通过上述过程的后处理, 就可以消除偶然的分类错误并可以得到广告视频.

4 实验结果

在实验中, 选取了国内不同电视台的 30 段普通电视节目剪辑. 首先, 在这些电视节目上进行镜头检测的实验. 实验结果包括本文的 RBFID 镜头检测算法和文献[7]中的检测算法的结果. 图 6 所示为 FD 的计算分布图. 其中图 6(a) 图所示为文献[7]镜头

检测算法计算的结果,图 6(b)图所示为 RBFID 镜头检测算法的计算结果.图中 x 轴表示帧序列, y 轴表示连续 2 帧之间的差值.

从 RBFID 算法计算出的分布图 6 (b) 可以看出,其 FD_m 的分布结果更符合实际的实验视频.在广告阶段,RBFID 可以检测出更多的切变镜头,其中一些是在文献[7]中没有检测出来的.这说明,本文的镜头检测算法在广告阶段能检测出更加密集的广告视频镜头.因此,根据本文算法得到的普通节目和广告段的镜头频率差值更加明显,从而能够更好地区分广告视频段和普通节目段.

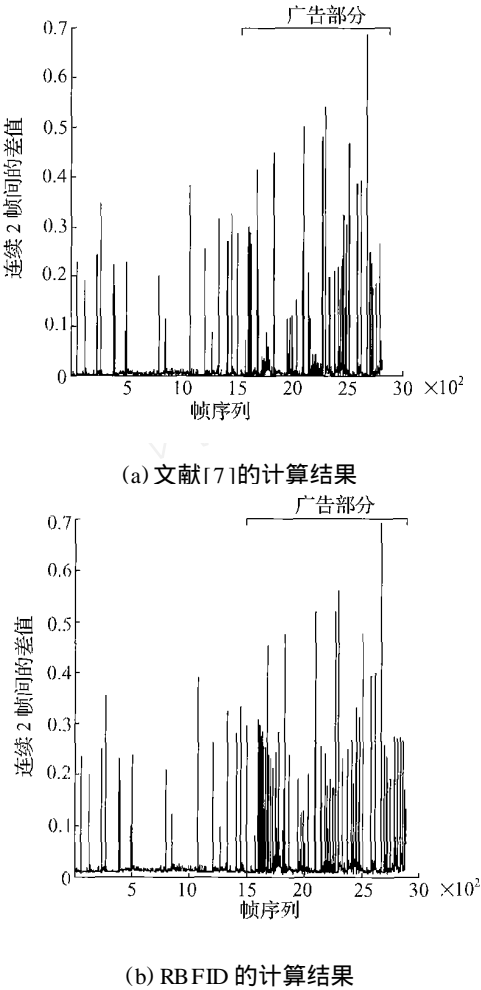


图 6 FD_m 分布图

Fig. 6 FD_m between frames calculated by different algorithms

本系统还对共包括 820 个切变镜头的 10 段电视节目片断进行了切变镜头检测.检测结果如表 1 所示,其中 RBFID 共检测到 800 个切变镜头,其中正确的是 789 个.而文献[7]检测出了 758 个镜头,其中正确的为 693 个.

表 2 为本系统在 2 段包括切变镜头和消隐镜头的视频段上的实验结果.第 1 段视频包含 80 个切变镜头和 2 个消隐镜头,其中 2 个消隐镜头和 75 个切

变镜头被准确地检测到.第 2 段视频包含 49 个切变镜头和 5 个消隐镜头,其中消隐镜头 5 个和 43 个切变镜头被正确的检测出来.

表 1 切变镜头的检测结果				
Table 1 Results of hard cut detection				
	阈值	切变	查全率/ %	准确率/ %
Algorithm[7]	0.10	758	84.51	91.42
RBFID	0.18	800	96.22	97.12

表 2 切变和消隐镜头的检测结果				
Table 2 Results of shot detection of hard cut and dissolve				
	消隐	切变	查全率/ %	准确率/ %
clip 1	2	75	93.90	96.25
clip 2	5	43	91.84	93.75

从表 1 和表 2 可以看出,本文算法在切变镜头和消隐镜头的查全率和准确率上都达到很好的检测效果.表 3 所示为本文算法在 4 段电视节目上的广告镜头的检测结果,其中包括没有经过重定界算法和经过重定界算法的广告镜头分类结果.从试验中可以看出,广告镜头重定界算法可以有效地消除广告边界的错分镜头,提高镜头分类的准确率.从表 3 可以看出,在采用重定界算法前、后的正确率分别为 93.77 %和 97.38 %.

表 3 广告镜头的分类结果			
Table 3 Classification results of commercial shots			
重定界	镜头数	广告镜头	准确率/ %
without	305	286	93.77
with	305	297	97.38

图 7 所示是基于本文算法搭建的视频广告检测原型系统.界面左侧的上半部分是电视节目视频播放窗口,下半部分是功能键,包括视频文件的打开、

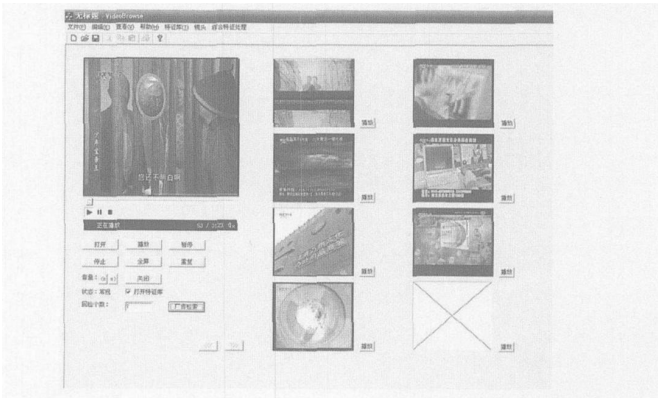


图 7 视频广告检测原型系统

Fig. 7 The robust commercial detection prototype system

播放、暂停和广告检索等等;右侧是检索到的广告视频段的起始帧,在每一个小窗口下面有一个键可以用来播放此广告视频段。

5 结束语

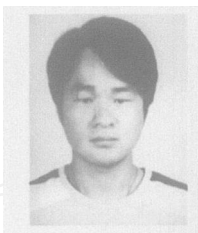
本文提出了一个鲁棒的广告检测系统,首先利用2个基于区域特征的镜头检测算法将视频段分割成镜头,同时提取一些基于镜头的鲁棒性特征参数。然后利用SVM分类器将镜头分成广告节目镜头和普通节目镜头。最后,通过后期处理重新定位广告镜头的边界以去除一些错分的镜头,并将广告镜头合并成相应的广告视频段。为了进一步提高系统对广告视频检测的准确性,可以引入另外一些鲁棒性的特征,例如音频特征和文本特征。而且在后处理中,还应该尝试采用基于广告内容连续性和一致性的镜头分割从而能更好的定位广告边界。

参考文献:

- [1] THOMAS HARGROVE url: <http://thomashargrove.com/logo-detection/>, 2005 - 08 - 28.
- [2] HAUPTMANN A, WITBROCKM, Story segmentation and detection of commercials in broadcast news video [A]. Advances in Digital Libraries Conference [C]. Santa Barbara, CA, 1998.
- [3] DAVID A. SADLER. Automatic TV advertisement detection from MPEG bitstream [A]. International Conference on Enterprise Information Systems [C]. Setubal, Portugal, 2001.
- [4] HUA Xiansheng, LU Lie, ZHANG Hongjiang. Robust learning-based TV commercial detection [A]. International Conference on Multimedia and Expo (ICME '05) [C]. Amsterdam, Netherlands, 2005.

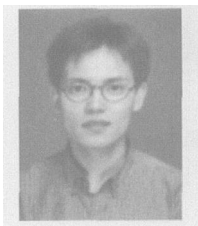
- [5] PINAR DUYGULU. Comparison and combination of two novel commercial detection Methods [A]. International Conference on Multimedia and Expo (ICME '04) [C]. Taipei, China, 2004.
- [6] RAINER L. On the detection and recognition of television commercials [A]. Proc of IEEE Int'l Conf. On Multimedia Computing and Systems [C]. Ottawa, Canada, 1997.
- [7] MIENE A, DAMMEYER A, HERMES T H, et al. Advanced and adaptive shot boundary detection [J]. Proc. of ECDL, 2001 (8) 39 - 43.
- [8] VAPNIK V. The Nature of Statistical Learning Theory [M]. New York Springer, 1995.

作者简介:



张亮,男,1981年生,硕士研究生,主要研究方向为基于内容的视频检索。

E-mail: zhangliangwin @ hotmail.com.



朱振峰,男,1974年生,讲师,主要研究方向为目标检测与识别、图像检索、视频分析以及机器学习。



赵耀,男,1967年生,教授,博士生导师,北京交通大学特聘教授,北京交通大学信息科学研究所所长。主持和参加了24项科研课题,发表论文90余篇,目前50余篇被SCI、EI检索,出版著作2部,申请发明专利3项。