

基于互信息量和自回归模型的镜头分割方法

李强军,李启南

(兰州交通大学 电子与信息工程学院,甘肃 兰州 730070)

摘要:随着互联网的急速发展,盗版,不健康,暴力等视频在网络上肆意流窜,如何快速、有效、准确地对视频数据进行管理,已然成为迫切需要解决的问题。在视频处理过程中,首先是对镜头进行分割,然后再进行视频帧的分析处理。然而许多视频内容的复杂性比较高,不一定能得到比较好的分割结果。鉴于此,提出一种基于互信息量和自回归模型的自适应阈值镜头分割算法。该算法首先以非均匀分块加权 HSV 直方图为基础,通过计算两帧的互信息量求出两帧的相似度值,然后建立自回归模型产生自适应阈值进行镜头分割,最终实现突变镜头的突变检测和渐变镜头的渐变检测,并采用时间窗口进一步降低检测误差。以优酷网上随机抽取下载的真实视频为测试对象,实验结果表明,该算法可适用于不同类型的视频镜头分割,具有很好的检测效果。

关键词:镜头分割;自适应阈值;互信息量;自回归模型

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2019)01-0035-05

doi:10.3969/j.issn.1673-629X.2019.01.008

A Lens Segmentation Method Based on Mutual Information and Self-regression Model

LI Qiang-jun, LI Qi-nan

(School of Electrical and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: With the rapid development of Internet, piracy, unhealthy, violence and other videos flows freely on the Internet. How to manage video data quickly, effectively and accurately has become an urgent problem to be solved. In the video processing, the lens is first segmented, and then the analysis and processing of video frames are carried out. However, the complexity of many video contents is relatively high, so it is not necessary to get a better segmentation result. In view of this, we propose an adaptive threshold lens segmentation algorithm based on mutual information and self-regression model. Firstly based on the non-uniform block weighted HSV histogram, through calculating the mutual information of two frames the similarity of two frames is computed. And then a regression model is established to generate adaptive threshold for lens segmentation. Finally, the mutation detection of mutation lens and gradient detection of gradient lens are realized, and the time window is used to further reduce the detection error. The real video downloaded randomly on Youku is selected as the test object. The experiment shows that the proposed algorithm is applicable to different video lens segmentation with better detection effect.

Key words: lens segmentation; adaptive threshold; mutual information; autoregressive model

0 引言

视频作为信息的一种载体,普遍化程度越来越高,在各个领域的应用也越来越广泛,同时也伴随着非法盗版,不健康,暴力等视频数据在网络上的肆意流窜。面对海量视频数据,如何快速、有效、准确地查找出需要的视频资源,已然成为一个迫切需要解决的问题。在视频分析处理过程中,首先是对镜头进行分割,然后进行视频帧的分析、提取、检索等处理。然而后续处理

的效果很大程度上都受镜头分割情况的影响,好的镜头分割对于内容帧的分析处理是极其重要的。正因为如此,视频镜头的分析算法受到了越来越多的关注和研究^[1]。

近年来,研究人员提出了许多镜头分割方法。Yeo等^[2]提出一种通过 MPEG 压缩视频的 DC 序列对视频镜头边界进行检测的算法,但是算法本身的应用范围相对较窄,对视频的变化要求较高。韩冰等^[3]提

收稿日期:2018-02-02

修回日期:2018-06-14

网络出版时间:2018-11-15

基金项目:教育部人文社会科学研究规划基金(13YJA870013);甘肃省档案科技基金(2014-04);甘肃省自然科学基金(1506RJZA072)

作者简介:李强军(1990-),男,硕士研究生,通讯作者,研究方向为信息安全技术;李启南,副教授,研究方向为信息安全技术。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20181114.1554.004.html>

出用粗糙集和模糊聚类的方法检测视频镜头边界,虽然加入了聚类的方法在一定程度上可以提高检测效率,但是具体的算法优化空间不大。巢娟等^[4]提出了基于多阈值检测的算法,通过设定一个高阈值和一个低阈值,将较高的阈值用于切变检测,较低的阈值用于渐变检测,该算法复杂度较低,但是对噪声、光线的剧烈变化以及镜头或物体的运动非常敏感。刘嘉琦等^[5]利用基于多模态特征融合的分割算法,对视频中的音频、画面、主题、文本等进行综合考虑并结合视频的结构特征进行镜头分割,该算法准确率较高,但是高的准确率依赖于声音、图像、文本等各个方面的综合分析,实现起来比较复杂,并且运算量大。Mohanta 等^[6]提出利用神经网络学习来获得镜头边界检测模型的算法,但是神经网络算法本身复杂度较高,而且神经网络算法需要的训练集本身要求也较大,对于短镜头的视频非常容易造成过学习。

视频由许多连续显示的镜头构成,而镜头的连续显示主要是通过连续切换的方式实现,切换可分为突变切换和渐变切换两种形式。突变切换在视频镜头的切换过程中,表现为一个镜头的最后一帧结束以后直接切换到下一个镜头的第一帧,这种切换方式的相邻两个镜头不存在交叉问题,并且它们的帧间差比较大,没有时间上的延迟,切换速度快,镜头变化明显;渐变切换相对来说比较复杂,在进行相邻镜头间的切换时,尾部出现内容的淡化变换,存在局部的交叉,通过时间上的延迟,渐变到下一个镜头,渐变类型常见的有淡入、淡出、溶解、扫换等方式^[7]。

文中提出一种基于自回归模型和互信息量的镜头分割方法,即选用 HSV 直方图特征向量,首先在 HSV 直方图的基础上计算两帧的互信息量并转换成相似度值,然后通过计算的相似度值建立自回归模型求取判异决策值,从而获得自适应阈值,最后结合产生的阈值和设定的帧时间窗口确定镜头的边界分割。

1 改进的镜头分割方法

文中采用文献[8]中的非均匀分块加权 HSV 直方图法。按照黄金分割比将整个视频的帧 T 的长和宽划分成 3×3 的不等小子块,然后计算每块小子块的信息量,再给每块小子块赋予不同的权值,最终对帧中的所有小块采取加权平均,从而计算出一帧的信息量。加权矩阵如式 1:

$$T = \begin{bmatrix} T_1 & T_2 & T_3 \\ T_4 & T_5 & T_6 \\ T_7 & T_8 & T_9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 1 \end{bmatrix} \quad (1)$$

1.1 帧间相似度计算

一般对于两个随机变量,可以通过计算它们的信

息量,比对信息量的相似性来衡量它们的相似度。在图像中,信息点被定义为图像中的像素点,信息量则通过信息点计算得出,计算两帧相似度时,通过两帧相互包含对方信息点的多少,求取相互包含的信息量就可以进行帧间差的度量。当两帧图像的帧差较大时,两帧图像内容改变比较大,对应的互信息量则较小;当两帧图像的帧差较小时,两帧图像内容比较相似,对应的互信息量则较大^[9]。设视频中两帧为 f_a 、 f_{a+1} ,式 2 定义了帧 f_a 、 f_{a+1} 在 T 子块的互信息量。

$$I_{a,a+1}^{T_i} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{a,a+1}^{T_i}(i,j) \log \frac{P_{a,a+1}^{T_i}(i,j)}{P_a^{T_i}(i)P_{a+1}^{T_i}(j)} \quad (2)$$

其中, N 表示灰度级数; $P_a^{T_i}(i)$ 、 $P_{a+1}^{T_i}(j)$ 是帧 f_a 、 f_{a+1} 在灰度块 T_k 的边缘概率密度,表示帧 f_a 、 f_{a+1} 在 T_k 块 HSV 直方图统计中像素数为 i,j 的概率; $P_{a,a+1}^{T_i}(i,j)$ 是帧 f_a 、 f_{a+1} 在块 T 中的联合概率密度,表示帧 f_a 到帧 f_{a+1} 的变化中,对应 T_k 子块 HSV 空间的像素数从 i 变换到 j 的概率。 $P_a^{T_i}(i)$ 、 $P_{a+1}^{T_i}(j)$ 均为零时,计算的互信息量存在无意义性,此时设置互信息量为零。

通过式 2 的计算,统计 9 个子块的互信息量,用式 3 计算帧 f_a 、 f_{a+1} 之间的分块加权平均互信息量。

$$I_{a,a+1} = \frac{\sum_{T=1}^9 T_{T_i} \times I_{a,a+1}^{T_i}}{\sum_{T=1}^9 T_{T_i}} \quad (3)$$

通过上面的计算,得到 $I_{a,a+1}$,即帧 f_a 、 f_{a+1} 的互信息量。接着使用上面的结果计算出帧 f_a 、 f_{a+1} 的相似度值,利用文献[10]中的定义计算出两帧 f_a 、 f_{a+1} 的相似度值,表示为:

$$S_{a,a+1} = I_{a,a+1}^T (1 - \text{Dif}_{a,a+1}) \quad (4)$$

其中, $\text{Dif}_{a,a+1}$ 是帧 f_a 、 f_{a+1} 的非均匀分块 HSV 颜色直方图的特征差。

1.2 自适应阈值选择

自回归模型 (autoregressive model) 是用自身做回归变量的过程,即利用前期若干时刻的随机变量的线性组合来描述以后某时刻随机变量的线性回归过程,它是时间序列中的一种常见形式,一般表示为:

$$X_i = \beta_{i-1}X_{i-1} + \beta_{i-2}X_{i-2} + \cdots + \beta_{i-p}X_{i-p} + \varepsilon_i \quad (5)$$

其中, X_i 为模型变量; $\beta_{i-1}, \beta_{i-2}, \cdots, \beta_{i-p}$ 为回归系数; ε_i 为随机误差; p 为阶数。

在视频帧序列的变化过程中,镜头切换除了突变过程,其余的可以看作是时间序列帧的一种渐变过程。尽管相邻帧相似度值序列从整体上进行观察时是不平稳的,但在局部上可以看作是统计学上近似平稳^[11]。

文中把这个近似平稳的局部作为滑动窗口,选取当前镜头内按时间顺序排列的邻帧相似度值,作为序列样本观测值 S_1, S_2, \cdots, S_n ,阶数为 p 的自回归模型如

下所示:

$$Y = \begin{bmatrix} S_{p+1} \\ S_{p+2} \\ \vdots \\ S_n \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_{p+n} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad (6)$$

$$X = \begin{bmatrix} S_p & S_{p-1} & \cdots & S_1 \\ S_{p+1} & S_p & \cdots & S_2 \\ \vdots & \vdots & \ddots & \vdots \\ S_{n-1} & S_{n-2} & \cdots & S_{n-p} \end{bmatrix} \quad (7)$$

则有:

$$Y = X\beta + \epsilon \quad (8)$$

由最小二乘法估计回归系数,用式9表示为:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (X^T X)^{-1} X^T Y \quad (9)$$

通常情况下一组镜头的帧数是不少于20帧的,这里取 $n=20$ 。文中采用二阶自回归模型, p 设定为2。接下来再对滑动窗口零均值化,设 \bar{S} 为 S_1, S_2, \dots, S_n 的平均值,表示为:

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n S_n \quad (10)$$

令

$$\chi_i = S_i - \bar{S}, i = 1, 2, \dots, n + 1 \quad (11)$$

其中, $\chi_1, \chi_2, \dots, \chi_{n+1}$ 是零均值化后的序列, χ_{n+1} 为滑动窗口后续的零均值化后的相似度值。

由文献[12]得,时间顺序排列的邻帧相似度值序列样本二阶自回归模型表示为:

$$\chi_i = \beta_1 \chi_{i-1} + \beta_2 \chi_{i-2} + e_i \quad (12)$$

其中, β_1, β_2 为回归系数; e_i 为白噪声。从而可得方差 σ_e^2 ,表示为:

$$\sigma_e^2 = \frac{1}{n-1} \sum_{i=3}^{n+1} e_i^2 = \frac{1}{n-1} \sum_{i=3}^{n+1} (\chi_i - \hat{\beta}_1 \chi_{i-1} - \hat{\beta}_2 \chi_{i-2})^2 \quad (13)$$

计算判异决策值 λ ,表示为:

$$\lambda = \frac{e_{n+1}}{\sigma_e} \quad (14)$$

最后确定阈值。如果统计量 λ 大于或等于阈值,则说明镜头未进行突变切换,反之则出现了突变切换。对于由计算得到的统计量序列 $\{\lambda_i\}$,计算正统计量的平均值及标准差。

$$\bar{\lambda} = \frac{1}{n} \sum_{i=1}^n \lambda_i \quad (15)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^N (\lambda_i - \bar{\lambda})^2} \quad (16)$$

根据统计分布理论及渐变切换过程中相似度值波动较小的特点,可得到统计量 λ 的阈值为: $k = \bar{\lambda} + 2\sigma$ 。镜头进行切割完成后,下一组镜头内通过上述方法重新计算自适应阈值。

2 突变检测

镜头突变在视频切换中比较特殊,由于前后内容未发生交叉,如图1的突变给人的主观感受就像是一种画面的跳变,它在切换过程中不存在时间上的延迟,对比切换前后,变化非常明确,检测也比较容易。

计算镜头内的自适应阈值 k 和第 $i+n, i+n+1$ 帧的相似度值 $S(f_{i+n}, f_{i+n+1})$,将其与 k 进行比较。如果 $S(f_{i+n}, f_{i+n+1}) < k$,表示位置 f_{i+n+1} 为突变切换。由于受闪光帧的影响,相似度值也会发生突变,可能会持续数帧,因此可以适当设置预测序列的时间窗口,增加步长 i 来进行比较,使得相隔 $2i+1$ 帧后恰好跳过闪光持续的帧,跳过闪光持续帧的下一帧相似度值为 $S(f_{i+n}, f_{2i+n+1})$,通过计算 $S(f_{i+n}, f_{2i+n+1})$ 来检测突变状态,如果存在后续数据有 $S(f_{i+n}, f_{2i+n+1}) < k$,则认为发生了一次突变切换。

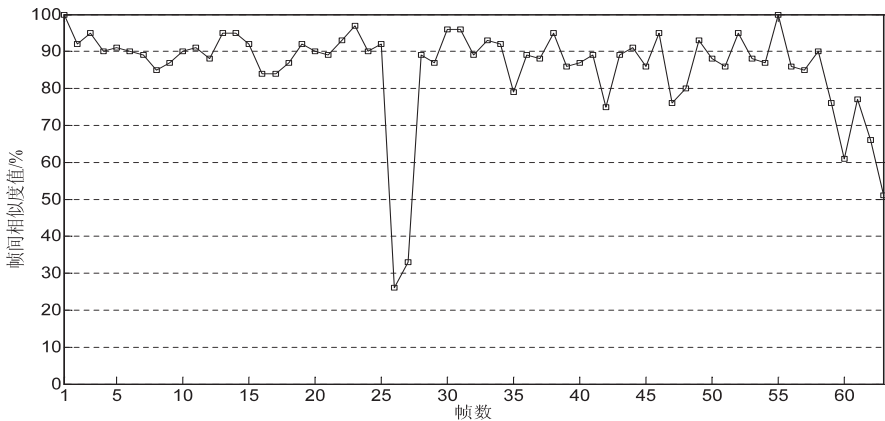


图1 突变

万方数据

3 渐变检测

淡入和淡出作为渐变的主要形式都有一个共同特点,就是在其变化过程中,都存在一个画面淡化的过程,因此,可以通过对视频帧相似度值的缓慢变化特性来检测渐变切换的位置^[13]。图 2 淡入时视频段帧间相似度值缓慢增大,图 3 淡出时视频段帧间相似度值缓慢减小。溶解时视频段帧间相似度的变化为图 2 和图 3 的综合,表现为帧间相似度先减小后增大或者先增大后减小。

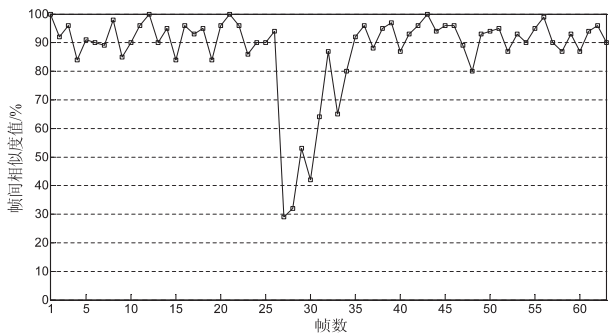


图 2 淡 入

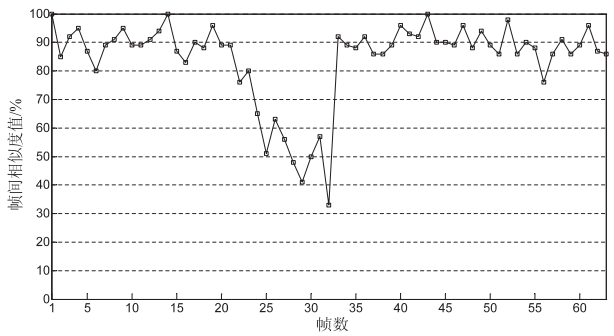


图 3 淡 出

镜头渐变切换检测的方法如下描述:
计算镜头内的自适应阈值 k 和第 $i+n$ 、 $i+n+1$ 帧的相似度值 $S(f_{i+n}, f_{i+n+1})$, 如果总有 $S(f_{i+n}, f_{i+n+1}) \geq k$,

则设置时间窗口增加步长 i , 并计算帧 f_{n+1} 与 f_{2i+n+1} 的相似度值。这个过程中跨过了渐变帧的连续变换时出现的符合阈值的缓慢变化, 当存在 $S(f_{n+1}, f_{2i+n+1}) < k$ 时, 则认为存在渐变, 并检测渐变的镜头边界帧。具体如下:

(1) 计算 f_{n+1} 到 f_{2i+n+1} 相邻帧的相似度值集合 $\{S(f_{n+1}, f_{i+n+1})\}$, 接着计算其均值 u 。

$$u = \frac{\sum_{i=1}^n S_{n+i}}{n}$$

(17)

(2) 计算相似度值的方差 σ^2 。

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (S(f_{i+n}, f_{i+n+1}) - u)^2$$

(18)

(3) 进行归一化处理并求归一化方差 σ_1^2 , 计算 σ_1^2 的导数值 $\sigma_1^{2'}$, 如果 $\sigma_1^{2'} < 0$, 则发生淡出, 如果 $\sigma_1^{2'} > 0$, 则发生淡入。

(4) 计算帧 f_{n+1} 与 f_{2i+n+1} 相似度值 $S(f_{n+1}, f_{i+n+1})$ 距离 k 偏移最大的值, 并求得偏离最大值的位置为 $j+n+1$ 处, 则认为 $j+n+1$ 处发生了一次渐变切换, 可认为 $j+n+1$ 处为该镜头的边界。

4 实验结果及分析

文中采用对媒体信息检索的通用指标—查全率和准确率^[14]来评估视频镜头分割的效果。查全率为正确检出数与视频实际镜头总数的比值, 查准率为正确检出数与检出的镜头总数之间的比值。实验开发平台为 (Java Development Kit) 和 Eclipse, 选用文献 [15-16] 的算法验证文中改进算法的有效性。实验视频均来自优酷网上随机抽取下载, 抽取的视频均为不同题材类型, 视频的帧数也不相同。实验结果如表 1~3 所示。

表 1 文中算法结果

视频名称	帧数	镜头数	检测数	正确数	漏检数	误检数	查全率/%	准确率/%
中央华府广告	370	11	11	11	0	0	100	100
Dota2 简介视频	1 518	33	44	31	2	11	93.94	70.45
法国香水广告	1 074	37	40	36	1	4	97.3	90
战狼 2 预告片	2 623	71	87	68	3	18	95.77	78.16
中央空调广告	486	17	15	15	2	0	88.24	100
Mate10 广告	1 657	42	63	40	2	23	95.24	63.49

表 2 文献[15]算法(规范化灰度分布帧差)结果

视频名称	帧数	镜头数	检测数	正确数	漏检数	误检数	查全率/%	准确率/%
中央华府广告	370	11	6	6	5	0	54.54	100
Dota2 简介视频	1 518	33	24	19	14	5	57.57	79.16
法国香水广告	1 074	37	51	30	7	24	81.08	58.82
战狼 2 预告片	2 623	71	96	66	5	30	92.95	68.75
中央空调广告	486	17	17	15	2	2	88.24	88.24
华为数据	1 657	42	60	37	5	13	88.09	61.66

表3 文献[16]算法(基于直方图的切变镜头自动检测)结果

视频名称	帧数	镜头数	检测数	正确数	漏检数	误检数	查全率/%	准确率/%
中央华府广告	370	11	9	7	4	2	63.63	77.77
Dota2 简介视频	1 518	33	40	24	9	16	72.77	60
法国香水广告	1 074	37	28	23	14	5	62.16	82.14
战狼2 预告片	2 623	71	102	61	10	41	85.91	59.8
中央空调广告	486	17	16	15	2	1	88.23	93.75
Mate10 广告	1 657	42	70	36	6	34	85.71	51.42

对比表1~3可以看出,在查全率上文中方法的表现更为优异,对于不同类型的视频都具有较好的稳定性。类似于“Dota2 简介视频”中的内容运动较剧烈,同时存在闪光灯的影响和许多的淡变切换,使其检测过程中存在一些误检,干扰相对比较大。总体来说,文中方法在视频镜头边界检测过程的检测效果比较明显,具有一定的有效性。

5 结束语

通过视频帧的 HSV 直方图互信息量计算出的相似度值,计算出镜头内的自适应阈值并结合时间窗口,在剔除了闪光灯的影响下,进行镜头的突变检测和渐变检测。其中阈值是采用自回归模型计算而来,体现了自适应性。实验结果表明,该方法对镜头边界检测具有良好的检测能力。

参考文献:

[1] 陈逸韬,宫宁生,王淑敏. 基于分块直方图帧差变化率的镜头分割算法研究[J]. 无线互联科技,2016(12):103-105.

[2] LU H B,ZHANG Y J,YAO Y R. Robust gradual scene change detection [C]//International conference on image processing. Kobe,Japan;IEEE,2002:304-308.

[3] 韩冰,高新波,姬红兵. 基于粗糙集和模糊聚类的新闻视频镜头边界检测方法[J]. 中国图象图形学报,2007,12(3):522-528.

[4] 巢娟,孙钺锋,蒋兴浩. 基于双重检测模型的视频镜头分割算法[J]. 上海交通大学学报,2011,45(10):1542-1546.

[5] 刘嘉琦,封化民,闫建鹏. 基于多模态特征融合的新闻故事单元分割[J]. 计算机工程,2012,38(24):161-165.

[6] MOHANTA P P,SAHA S K,CHANDA B. A model-based shot boundary detection technique using frame transition parameters[J]. IEEE Transactions on Multimedia,2012,14(1):223-233.

[7] SANTOS A C S,PEDRINI H. Shot boundary detection for video temporal segmentation based on the weber local descriptor[C]//2017 IEEE international conference on systems,man,and cybernetics. [s.l.];IEEE,2017:1310-1315.

[8] 张炜,殷杰,王士林,等. 结合颜色和空间信息的自适应视频镜头分割[J]. 计算机应用与软件,2012,29(4):111-113.

[9] CASELLA G,FIENBERG S,OLKIN I. Time series analysis and its applications-with r examples[J]. Journal of the American Statistical Association,2006,97(458):656-657.

[10] 刘高军,杨丽. 改进的互信息量相似度曲线关键帧提取研究[J]. 计算机应用与软件,2014,31(2):153-156.

[11] 熊伟,吴春明,姜明. 使用线性分割和序列合并的视频镜头分割方法[J]. 电子学报,2014,42(4):640-645.

[12] 徐慧娟,周世健,鲁铁定. 自回归 AR 模型整体最小二乘分析[J]. 江西科学,2011,29(5):543-545.

[13] 奚晓晔,严利民,杜斌. 帧间差值分布和渐变模型的视频镜头分割方法[J]. 电子测量与仪器学报,2016,30(11):1765-1773.

[14] RANATHUNGA L,ZAINUDDIN R,ABDULLAH N A. Conventional video shot segmentation to semantic shot segmentation[C]//6th IEEE international conference on industrial and information systems. [s.l.];IEEE,2011:186-191.

[15] 沈博超,周军. 视频突变检测的规范化灰度分布帧差方法[J]. 计算机工程,2009,35(3):242-244.

[16] 刘典,刘文萍. 一种基于直方图的切变镜头自动检测算法[J]. 北方工业大学学报,2007,19(3):16-20.