

视频镜头边界检测的统一策略

耿玉亮 须 德

(北京交通大学计算机与信息技术学院计算机研究所, 北京 100044)

摘要 随着计算机和影视技术的发展,用来有效地组织、索引和检索海量视频数据的基于内容的视频检索系统吸引了众多学者的兴趣,而自动镜头边界检测是实现基于内容的视频检索的一个重要步骤。因此,为了有效地进行视频镜头边界检测,在介绍现有的镜头边界检测基本方法的基础上,提出了一种镜头边界检测的统一策略。由于它是在综合考虑视频的颜色特征和时间特征的基础上来建立基于幅值映射的1维视频特征曲线,并利用小波分析对信号奇异点的检测能力和对噪声具有较强的鲁棒性,首先实现了镜头边界的预检测;同时利用1维特征曲线在镜头边界区间中表现出的良好统计特性实现了对渐变镜头的进一步分类,进而实现了在统一检测框架下的镜头边界检测。实验结果表明,该算法能够有效地检测出视频镜头边界。

关键词 视频 镜头边界检测 小波分析 奇异点检测

中图分类号: TN941.1 **文献标识码:** A **文章编号:** 1006-8961(2005)05-0650-06

A Unified Framework for Shot Boundary Detection

GENG Yu-liang, XU De

(Institute of Computer Science, School of Computer & Information Technology, Beijing Jiaotong University, Beijing 100044)

Abstract With the development of computer and TV technology, the content-based video retrieval system becomes more and more important to organize, index and retrieve the massive video information in many application domains. Automatic shot boundary detection is a crucial issue in content-based video retrieval and has attracted attention of many researchers in recent years. However, few literatures are reported to make a unified method to detect the types of shot boundaries, such as cut, fade in/out and dissolve. In this paper, a unified framework is proposed to identify the various shot boundary types. Firstly, each video frame is represented as a point in HS color space, and the video sequence is mapped into a normalized feature curve by using the L_2 -norm. The feature curve can hold shot boundary characteristic very well. Then the detection of singularities with wavelet is employed to identify the potential shot boundaries in the feature curve. At last, the shot boundaries are further classified according to the characteristic of normalized feature curve. The experimental results compared with the conventional schemes achieve satisfying precision and recall of detecting shot boundaries in this scheme.

Keywords video, shot boundary detection, wavelet transformation, singularity detection

1 引言

随着多媒体存储压缩技术和计算机网络的发展,数字视频已越来越广泛地存在于人们的生活空间中,而视频点播、数字图书馆等视频服务也开始走进人们的生活,由于人们接触到的视频数据以前所未有的速度增长,因此,面对海量的视频信息,如何方便、快捷、准确地检索到所需要的信息,一直是人们关注的热点。在这种环境下,有关基于内容的视

频检索技术的研究也就应运而生。所谓基于内容的视频检索是指借助于视频数据的底层特征(如颜色、纹理、形状和运动等)从数据库中查找具有特定内容的视频片段,它区别于传统的基于关键字的检索手段,因它已融合了图像理解、模式识别等技术。

由于视频信号是由连续的图像组成的,不便于直接管理和检索,因此需要将其分割成相对独立的视频片段。由视频信号的组成可知,视频一般由多段镜头拼接而成,而镜头内部各帧图像是连续变化的。由于在目前研究的基于内容的视频检索系统

收稿日期:2004-02-23;改回日期:2004-09-28

第一作者简介:耿玉亮(1975~),男,1997年获北方交通大学学士学位,2000年开始在北方交通大学计算机应用专业攻读硕士学位,现为该专业博士研究生。主要研究方向为多媒体信息处理、基于内容的视频检索和机器学习。E-mail:gengyuliang@sohu.com

中,一般都是先将视频分割成独立的镜头,然后通过选取每个镜头的代表帧来表示该镜头,这样视频的特征提取、运动分析和检索等过程就可在代表帧上完成^[1],因此,镜头边界检测是基于内容的视频检索中十分重要的问题。

在镜头边界检测方面,目前已经研究出了许多检测算法^[2-6],但其中只有少数算法用于处理压缩视频,而大部分算法都是用于处理非压缩视频。在非压缩域中,算法可分为基于像素比较、基于块比较、基于直方图比较、基于聚类 and 基于模型等检测方法;在压缩域中,主要利用 DCT (discrete cosine transform) 系数、DC (DC coefficient) 图像、宏块编码方式和运动矢量等特征来进行镜头边界检测。

由于镜头边界类型的多样性以及镜头渐变时间的不确定性,在进行镜头边缘检测时,使用一个统一的视频特征和算法比较困难,因此,在实际的镜头检测中,常先检测突变,再检测渐变,在渐变的检测中,也常先检测某些类型的渐变,再检测其他类型的渐变。由于这样检测较复杂,因此,近年来一些学者提出了镜头边界检测的统一方法^[4-6]。

Zhang 最早提出了使用双阈值进行突变和渐变检测的算法。该算法虽实现简单,但其效率需依赖于全局阈值的选取,而且没有区分渐变类型^[4]。

文献[5]提出了分层的多分辨率镜头边界检测方法,该方法是先通过对每帧图像进行 Haar 小波分解,然后使用低频分量的颜色直方图差来区分突变,接着使用高频分量的边缘特征 (edge count, edge spectrum average, EC, ESA) 来检测淡入淡出 (fade), 并用双色差 (double chromatic difference, DCD) 来检验叠化 (dissolve), 最后使用空间平移运动矢量 (spatial translation motion vector, STMV) 来实现擦除 (wipe) 的检测。文献[5]基于对渐变产生式模型的分析,建立了一种对 3 种重要的渐变类型进行层次化分类的方法,但该算法对噪声比较敏感,同时对渐变的规范程度有一定要求。

文献[6]提出了基于小波的镜头边界检测框架,即首先提取视频数据的 64 维颜色特征和 64 维运动特征曲线,并利用小波变换对高维曲线进行奇异点分析来检测和定位候选镜头边界,然后使用支持向量机 (support vector machine, SVM) 将候选边界分为渐变、突变和错误检测点。文献[6]虽提出了一种新颖的镜头边界检测的方法,但是算法本身不仅计算量较大 (包括基于 SVM 的多类分类器的训

练), 而且对渐变的分类未能给出具体的方法。

基于以上分析,并综合考虑算法的效率和有效性,本文提出了一种基于小波分析的镜头边界检测的统一策略,即,首先,通过对高维视频特征空间中的特征点进行分析来建立视频序列与基于特征点幅值的 1 维特征曲线的映射,实验表明,1 维特征曲线在表征镜头边界特征上具有良好的特性;其次,由于在信号处理中,小波分析能可靠地识别信号的奇异性,并且对噪声有较强的鲁棒性^[7],因此本文将小波分析应用到 1 维特征曲线,用于实现对镜头突变和渐变的检测,并利用 1 维特征曲线在镜头边界区间中表现出的良好的统计特性来实现对渐变镜头的进一步分类。实验表明,本文算法不仅能有效地实现镜头边界的检测,并具有较高的查全率和查准率。

2 镜头边界检测的统一策略

2.1 视频特征曲线

在视频结构分析中,视频内容通常由其低级语义特征 (如颜色、纹理、运动和形状) 来表征,而且视频帧序列按照所选的特征可以表示为多维特征空间中的一条曲线,而每个视频帧即是特征曲线中的一个点,其维数即是所选特征的维数。这种特征曲线在一定程度上表现了视频的语义特征,如果两帧内容较为相似,则特征点的空间距离就比较接近,反之亦然。

由于选择不同的特征,同一视频可以表现为不同的曲线,因此,特征的选取直接影响到算法的效率。在实验中,可采用与人类颜色视觉感知比较接近的 HSV 颜色模型来建立基于颜色直方图特征的视频特征曲线,由于亮度分量受光照影响比较大,因此,在计算直方图时,只使用色度和饱和度分量,并对色度和饱和度分量分别量化为 16 个值和 4 个值来得到 2 维 HS 颜色特征直方图,共 $16 \times 4 = 64$ 个直方条 (bin), 每个直方条的取值范围为 $[0, c]$, c 为视频帧的像素总数。每个视频帧可表示为一个 64 维的特征矢量

$$\mathbf{h} = (x_1, x_2, \dots, x_{64}) \quad (1)$$

其中, x_i 是直方图中的第 i 个直方条的值。视频序列 \mathbf{V} 可表示为 64 维的特征曲线

$$\mathbf{V} = \{\mathbf{h}_t = (x_1(t), x_2(t), \dots, x_{64}(t))\} \quad (2)$$

其中, t 是视频帧 f 在视频序列中的序号。

镜头边界检测即是识别特征曲线 \mathbf{V} 中的奇异点,并进一步判断其边界类型。虽然在高维空间中,利用小波分析可以实现信号的奇异点检测,但由于

计算复杂度较高^[6],因此本文提出先将高维曲线映射到1维空间,再进行小波分析,这种降维处理既能保持信号的奇异性,又能降低计算复杂度。

由于同一镜头中的相邻帧间内容较为相似,因此相邻特征点具有相近的幅值,本文使用第 t 帧视频的特征矢量 \mathbf{h}_t 的幅值,将高维曲线映射到以下1维空间来得到以下1维特征曲线:

$$L(t) = \|\mathbf{h}_t\|_2 = \sum_{i=1}^{64} (x_i(t))^2 \quad t = 1, \dots, N \quad (3)$$

其中, N 为视频序列的帧数。

由直方图定义可知, $\sum_{i=1}^{64} x_i(t)$ 等于一个视频帧的像素总数 c ,而对于给定一个视频片段, c 为常数,求该视频片段的1维特征曲线 $L(t)$ 中的最大值点等价于

$$\begin{aligned} \max \sum_{i=1}^{64} (x_i(t))^2 \\ \text{s. t. } \sum_{i=1}^{64} x_i(t) = c \quad 0 \leq x_i(t) \leq c \end{aligned} \quad (4)$$

显然,对给定第 t 帧视频,当 $x_i(t) = c$, $\sum_{j \neq i}^{64} |x_j(t)| = 0$ 时,则 $L(t)$ 达到最大值,即 $\sum_{i=1}^{64} (x_i(t))^2 = c^2$,对式(3)进行归一化,得




$$L_{\text{normalized}}(t) = \sum_{i=1}^{64} (x_i(t))^2 / c^2 \quad t = 1, \dots, N \quad (5)$$

由于镜头是由一个具有时间或空间连续性的视频帧序列组成,且通常由摄像机的一次连续拍摄得到^[8],因此,同一镜头内的帧间内容相近,即具有相近的幅值;而不同镜头内的帧间内容则差别较大,即帧间幅值差别较大,其表现为1维特征曲线在镜头边界处存在明显奇异性,而在镜头内部则较为平滑,但对于不同的镜头边界类型,其1维特征曲线的特征不同(如表1所示)。

对于突变,由于镜头边界变化发生在相邻两帧间,因此,1维归一化特征曲线 $L_{\text{normalized}}(t)$ 表现为垂直的阶跃;对于淡入淡出,前一镜头视频帧逐渐趋于单色,而后一镜头的视频帧则由单色逐渐显现。这样当视频帧趋于单色时,由于直方图中相应直方条的统计值趋于视频帧的像素总数 c ,因此,1维归一化特征曲线 $L_{\text{normalized}}(t)$ 在单色帧处呈现一个峰值(见式(4)说明);对于叠化,前一镜头的视频帧逐渐消隐,同时后一镜头的视频帧逐渐显现,其1维归一化特征曲线 $L_{\text{normalized}}(t)$ 在镜头边界处表现为一个上

表1 镜头边界特征曲线

Tab. 1 The feature curves of different shot

boundary types		
镜头类型	1维特征曲线	说明
突变		在突变中, $L_{\text{normalized}}(t)$ 呈现为一个垂直的跳变
叠化		$L_{\text{normalized}}(t)$ 表现为上升或下降坡度
淡入淡出		$L_{\text{normalized}}(t)$ 表现为一个尖峰,并在单色或接近单色时达到峰值

升或下降坡度。需要说明的是,在极端情况下,镜头边界起止图像(后一镜头的起始视频帧和前一镜头的结束视频帧)的直方图只发生重新排列,如起始视频帧特征矢量记为 $\mathbf{h}_{\text{in}} = (c_1, c_2, \dots, c_n)$,结束视频帧特征矢量记为 $\mathbf{h}_{\text{out}} = (c^{(1)}, c^{(2)}, \dots, c^{(n)})$, $(c^{(1)}, c^{(2)}, \dots, c^{(n)})$ 为 (c_1, c_2, \dots, c_n) 的新排列,此时,起止帧虽然内容相差较大,但具有相同的幅值,致使产生漏检,但是在实验中,通过对纪录片、生活片、动作片等大量视频数据进行的实验分析发现,这种情况的发生几率很低,即对镜头边界检测算法的影响很小。

从上面分析可以看出,用幅值映射建立1维特征曲线的方法比全局直方图比较^[2]中计算帧间差(特征点)的方法有以下优点:①特征点幅值保持了视频帧间的独立性,实验结果显示,1维特征曲线在镜头边界处显示出较好的边界特征;②特征点幅值序列由于保留了视频帧的绝对信息,而不是差值信息,因此对噪音有较强的抵御能力;③特征点幅值序列对渐变过程中的不规范程度有较强的鲁棒性。表1给出了主要边界类型在1维特征曲线中的表现形式。

2.2 镜头边界检测算法

从上述分析可以看到,归一化1维特征曲线 $L_{\text{normalized}}(t)$ 在镜头边界处发生显著奇变,而在信号处理中,由于小波分析能可靠地识别信号的奇异性,并且对噪音有较强的鲁棒性,因此本文将小波分析应用到1维特征曲线,用来实现对镜头突变和渐变进行检测,并利用1维特征曲线在镜头边界表现出的良好的统计特性来实现对渐变镜头的进一步分类。

设 $\theta(t)$ 为平滑函数,且满足条件

$$\begin{aligned} \int_{-\infty}^{+\infty} \theta(t) dt = 1 \\ \lim_{t \rightarrow \pm\infty} \theta(t) = 0 \end{aligned} \quad (6)$$

$\theta(t)$ 的一阶导数定义为 $\Psi(t) = d\theta(t)/dt$,因为

$\int_{-\infty}^{+\infty} \Psi(t) dt = 0$, 所以 $\Psi(t)$ 是小波函数, 如果引入符号 $\zeta_a(t) = \frac{1}{a} \zeta(t/a)$, 则 $f(t)$ 关于小波函数 $\Psi(t)$ 在尺度 a 和位置 t 上的规范小波变换定义为

$$\begin{aligned} W_a f(t) &= f(t) * \Psi_a(t) = f(t) * \left(a \frac{d\theta_a(t)}{dt} \right) \\ &= a \frac{d}{dt} (f * \theta_a)(t) \end{aligned} \quad (7)$$

当 $\theta(t)$ 取 Gauss 函数时, $d\theta(t)/dt$ 实际上形成了 Canny 小波, Mallet 证明了 Canny 边缘检测等价于寻找小波变换的模的局部极大值^[7]。实验中, 本文使用 Gauss 核函数来实现镜头边界的检测。

从式(7)可以看出, 当尺度 a 很大时, 由于信号 $f(t)$ 与平滑函数 $\theta_a(t)$ 的卷积可消去信号中较小的变化, 所以能检测出比较大的变化点, 这样就可以通过计算 $|W_a f(t)|$ 的局部极大值来确定候选镜头边界。

镜头边界检测算法步骤:

(1) 先通过计算视频序列中每一帧的 HS 颜色直方图来建立归一化 1 维视频特征曲线 $L_{\text{normalized}}(t)$, 如图 1(a) 曲线所示 (简称为图 1(a)), 然后计算归一化特征曲线 $L_{\text{normalized}}(t)$ 的小波变换 $W_a f(t)$, 即可由 $|W_a f(t)|$ 得到归一化特征曲线 $L_{\text{normalized}}(t)$ 在低分辨率空间中的小波变换模局部极大值序列 $\{M_i\}$, 如图 1(b) 所示, 其中, M_i 为二元组 (P_{M_i}, Z_{M_i}) , P_{M_i} 为 M_i 在 $L_{\text{normalized}}(t)$ 中的帧序, Z_{M_i} 为极值点 M_i 的模值, 在本文的算法中, 是采用局部阈值的方法来确定候选镜头边界, 若设滑动窗口宽度为 w , 则极值点 M_i 的局部阈值 T_i 为

$$T_i = \frac{K}{N_i} \sum_j Z_{M_j}, j \in [P_{M_i} - w/2, P_{M_i} + w/2] \quad (8)$$

即, T_i 是以极值点 M_i 为中心的滑动窗口中所有极值点的平均值的 K 倍, 其中, N_i 为位于滑动窗口内的极值点的数目, K 为经验值, 实验中, 取 $K=3$ 。当 $Z_{M_i} > T_i$ 时, M_i 为有效镜头边界, 记为 $B_i^{\text{potential}}$, 否则为噪声点, 这样即得到一个候选镜头边界集合 $\{B_i^{\text{potential}}\}$ 。实验证明, 使用局部阈值的方法能够有效地滤除噪声产生的局部极大值点。

(2) 检测镜头突变

在确定镜头边界位置后, 可再进一步检测镜头的边界类型, 由表 1 可知, 由于镜头突变在 1 维特征曲线上表现为一个阶跃, 而且持续时间只有两帧, 因此使用一个简单有效的阈值比较, 即可将突变从其他渐变类型中区分出来, 而不用计算视频曲线在奇异点的 Lipschitz 指数^[6]。

检测时, 首先定位镜头边界的起止点。对于每个候选的镜头边界 $B_i^{\text{potential}}$, 在低分辨率空间向左右搜索偏离其位移最小的极小值点 S_i^{left} 和 S_i^{right} , 并将其作为镜头边界的起止点, 即将 S_i^{left} 作为第 i 个镜头边界区间的起始帧序号, 将 S_i^{right} 作为镜头边界区间的结束帧序号。

在 $[S_i^{\text{left}}, S_i^{\text{right}}]$ 帧序列区间, 计算归一化的 1 维特征曲线 $L_{\text{normalized}}(t)$ 的差值序列

$$\begin{aligned} D(k, k+1) &= |L_{\text{normalized}}(k) - L_{\text{normalized}}(k+1)| \\ k &\in [S_i^{\text{left}}, S_i^{\text{right}} - 1] \end{aligned} \quad (9)$$

其中, k 为视频帧在视频序列中的序号, 定义突变的判定阈值为

$$T_i^{\text{cut}} = K_{\text{cut}} \times \frac{\left[\sum_{k=S_i^{\text{left}}}^{S_i^{\text{right}}-1} D(k, k+1) \right] - \max(D(k, k+1))}{S_i^{\text{right}} - S_i^{\text{left}}} \quad (10)$$

即, 判定阈值为舍弃镜头边界区间中最大差值 $\max(D(k, k+1))$ 后的差值序列的平均值。当 $\max(D(k, k+1)) > T_i^{\text{cut}}$ 时, 镜头边界点 $B_i^{\text{potential}}$ 为一个突变, 镜头突变位置为第 k 帧, K_{cut} 为经验值, 这样, 即可得到镜头突变集合 $\{B_i^{\text{cut}}\}$ 。

(3) 检测镜头渐变

在确定了突变镜头集后, 就可对剩余镜头边界点 $\{B_i^{\text{potential}}\} - \{B_i^{\text{cut}}\}$ 进行进一步检测。叠化镜头边界在 1 维特征曲线中近似表现为上升或下降的坡度, 其对应于小波变换的模值曲线中一个单峰 (见图 1(b) 中 D 点), 而淡入淡出边界类型在 1 维特征曲线中近似表现为单峰, 其对应于小波变换的模值曲线中的一个双峰 (见图 1(b) 中 F 点)。

在进行镜头渐变检测时, 首先在小波变换的模极大值曲线的低分辨率空间, 通过对镜头边界的起止点进行校验来区分叠化和淡入淡出类型, 即当前一镜头边界的结束帧与后一镜头的起始帧重合时, 即 $S_{i+1}^{\text{left}} = B_i^{\text{right}}$ 时, 则将两个镜头边界 $B_i^{\text{potential}}$ 和 $B_{i+1}^{\text{potential}}$ 合并为候选淡入淡出镜头边界, 这样就得到候选淡入淡出边界集合 $\{B_i^{\text{fade}}\}$; 对存在 3 个以上连续边界的镜头边界点, 则合并其边界点, 并将其作为未知边界类型集合 $\{B_i^{\text{unknown}}\}$; 将其他镜头边界点集合作为叠化边界集合 $\{B_i^{\text{dissolve}}\}$ 。由于强光 (如闪光灯、爆炸, 见图 1 中 E 点为一爆炸场景) 产生的现象 (指特征曲线相应位置的形状特征) 与淡入淡出边界相似, 故需要进一步区分。由于强光持续时间较

短,而且具有较高的峰值,因此具有较大的变化率。在实验中,本文通过计算 1 维特征曲线 $L_{\text{normalized}}(t)$ 在镜头变化区间 $[S_i^{\text{left}}, S_i^{\text{right}}]$ 的变化率 r_i 来区分强光和淡入淡出边界

$$r_i = \frac{\max(L_{\text{normalized}}(k)) - (L_{\text{normalized}}(S_i^{\text{right}}) + L_{\text{normalized}}(S_i^{\text{left}}))/2}{S_i^{\text{right}} - S_i^{\text{left}}} \quad k \in [S_i^{\text{left}}, S_i^{\text{right}}] \quad (11)$$

设 T_{illum} 为强光阈值,当 $r_i < T_{\text{illum}}$ 时,则镜头边界 $B_i^{\text{p-fade}}$ 为淡入淡出边界,这样得到的淡入淡出边界集合 $\{B_i^{\text{fade}}\}$ 和强光集合(即强光产生的极大值点集合) $\{B_i^{\text{illum}}\}$ 的关系为 $\{B_i^{\text{illum}}\} = \{B_i^{\text{p-fade}}\} - \{B_i^{\text{fade}}\}$,由于强光对其附近极值点的阈值计算(步骤 1)可能会产生较大的影响,因此需要重新修正强光产生的极大值点 B_i^{illum} 所在区间 $[P_{M_i} - w/2, P_{M_i} + w/2]$ 中极大值点的阈值 T_i (P_{M_i} 为 B_i^{illum} 在视频序列中的帧序)。镜头类型判断时,首先,在小波变换模局部极大值点序列 $\{M_i\}$ 中,去除强光产生的极大值点 B_i^{illum} ;然后,运用式(8)重新计算 B_i^{illum} 所在区间中所有极大值点的局部阈值 T_i ,并对新镜头边界点按步骤 2、步骤 3 依次判断其镜头类型。

3 实验

为验证本文算法效果,选取了 4 段视频对本文算法、双阈值算法、HMSTD 算法进行了镜头边界检测对比实验。

对镜头边界检测结果的评价方法一般使用查全率(recall) R_{recall} 和查准率(precision) $R_{\text{precision}}$ 这两个参数,它们的定义如下:

$$R_{\text{recall}} = N/(N + N_1), R_{\text{precision}} = N/(N + N_2)$$

其中, N 为正确检出数, N_1 为漏检数, N_2 为误检数。

本文也采用了这种评价标准。

实验中选取了 4 段视频,包括纪录片、动作片和生活片,一共有 16 000 余帧图像,包含 100 多个镜头。视频帧大小均为 352×240 , 24 位真彩色,其中 2 段视频数据来自 2001 TREC video retrieval test collection^[9] 中的中速运动视频。

在镜头检测中,滑动窗口大小的选择是非常重要的,一般设置滑动窗口的大小为 300 帧,平滑函数为 Gaussian 函数,小波尺度 $a=4$ 。实验结果显示,本文提出的用于镜头边界检测的统一策略取得了较好的效果。图 1 显示的是部分帧间差的实验数据,包含有强光 E(explosion)、镜头突变 C(cut)、叠化 D(dissolve)、淡入淡出 F(fade)等情况,其中,图 1(a)为归一化的 1 维视频特征曲线,图 1(b)为检测到的镜头边界(*所示),图 1(c)为排除强光 E 后,修正后得到镜头边界(新增边界 C)。表 2 给出了实验结果。

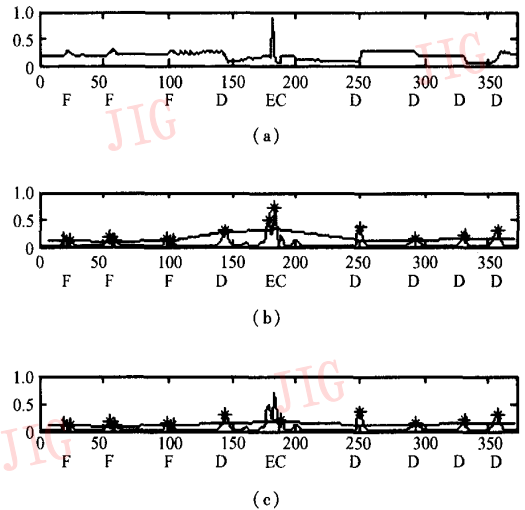


图 1 《碧血剑》片段

Fig. 1 Experimental result on sword stained with royal blood

表 2 镜头边界检测统一策略实验结果

Tab. 2 Experimental results of our algorithm

视频片段	视频描述	帧数	突变	渐变		漏检	误检	强光	镜头检测结果统计	
				叠化	淡入淡出				查准率 $R_{\text{precision}}$ (%)	查全率 R_{recall} (%)
Challenge at Glen Canyon (part)	35 突变	6 327	33	1	1	2	3	0	92.1	94.5
	1 叠化									
how water won the west (part)	17 叠化	3 375	0	14	3	3	2	0	89.4	85
	3 淡入淡出									
《碧血剑》片段	1 突变	386	1	4	3	1	1	1 (爆炸)	88.9	88.9
	5 叠化									
	3 淡入淡出									
Act3 of family album U. S. A II	43 突变	5 900	41	1	0	3	1	0	97.8	93.3
	2 叠化									

表3、表4给出了本文算法与几种传统方法实验结果的比较。由表3、表4可见,双阈值算法^[4]虽计算简单,但是对于渐变有较高的误检,而且不能区分渐变类型;HMSTD^[5]算法是建立在渐变产生式的基础上,由于其对渐变的规范程度有较高的要求,因此具有较低的查全率,而且计算复杂度较高。从实验结果可以看出,本文提出方法在综合性能上能取得较好的效果。

表3 本文算法与传统算法实验结果比较
(视频 Challenge at Glen Canyon)

Tab. 3 Experimental results compared with the conventional schemes (video data from Challenge at Glen Canyon)

算法	视频描述	查准率($N/(N+N_1)$), 查全率($N/(N+N_2)$)		
		突变	渐变	
			叠化	淡入淡出
双阈值算法 ^[4]	6 327 帧	34/35, 34/35	2/4, 2/2	
HMSTD 算法 ^[5]	35 突变 1 叠化	34/35, 34/35	1/1, 1/1	1/1, 1/1
本文算法	1 淡入淡出	33/35, 33/35	1/2, 1/1	1/1, 1/1

表4 本文算法与传统算法实验结果比较
(视频 How water won the west)

Tab. 4 Experimental results compared with the conventional schemes (video data from How water won the west)

算法	视频描述	查准率($N/(N+N_1)$), 查全率($N/(N+N_2)$)		
		突变	渐变	
			叠化	淡入淡出
双阈值算法 ^[4]	3 375 帧	0/0, 0/0	12/23, 12/20	
HMSTD 算法 ^[5]	17 叠化	0/0, 0/0	11/14, 11/17	2/2, 2/3
本文算法	3 淡入淡出	0/0, 0/0	14/16, 14/17	3/3, 3/3

4 结 论

镜头边界检测不仅是实现基于内容的视频检索的一个重要步骤,而且一直是视频检索领域的研究热点,本文提出了一种进行镜头边界检测的统一策略,由于它是在综合考虑视频的颜色特征和时间特征的基础上来建立视频帧序列与基于特征点幅值的1维特征曲线的映射,并利用小波分析对信号奇异点检测的能力和对噪声点具有的较强的鲁棒性,实现了镜头边界类型的统一检测。实验结果表明,本文算法不仅能有效地检测出镜头边界类型,而且具有较高的查全率和查准率,但在渐变类型的精确分类上,还需要进一步研究。

参考文献 (References)

- 1 Elmagarmid A K, Jiang H, Helal A, et al. Video database systems [M]. Boston MA, USA: Kluwer Academic Publishers, 1997.
- 2 Irena Koprinska, Sergio Carrato. Temporal video segmentation: A survey[J]. Signal Processing: Image Communication, 2001, 16(5): 477 ~ 500.
- 3 Zhang H J. Content-based video analysis, retrieval and browsing [R]. Beijing: Microsoft Research Asia, 2001.
- 4 Zhang H J, Kankanhalli A, Smoliar S W. Automatic partitioning of full-motion video[J]. Multimedia Systems, 1993, 1(1): 10 ~ 28.
- 5 Hong Heather Yu. A hierarchical multiresolution video shot transition detection scheme [J]. Computer Vision and Image Understanding, 1999, 75(1-2): 196 ~ 213.
- 6 Chua Tat-seng, Feng Huamin, Chandrashekhara A. An unified framework for shot boundary detection via active learning [A]. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing [C], Hong Kong, 2003: 845 ~ 848.
- 7 Peng Yuhua. Wavelet transform and its application in engineering [M]. Beijing: Science Press, 1999: 165 ~ 173. [彭玉华. 小波变换与工程应用 [M]. 北京: 科学出版社, 1999: 165 ~ 173.]
- 8 Film analysis guide [EB/OL]. <http://classes.yale.edu/film-analysis>.
- 9 TREC video retrieval test collection [EB/OL]. <http://www.open-video.org>, 2001.