

一种快速有效的视频镜头边界检测方法

耿玉亮, 须 德, 冯松鹤

(北京交通大学计算机研究所, 北京 100044)

摘 要: 本文在分析现有的镜头边界检测方法的基础上, 提出了一种层次化镜头边界检测方法. 首先我们利用小波变换对信号奇异点的检测能力和对噪声具有较强的鲁棒性, 实现了镜头边界的预检测. 然后我们通过对候选镜头边界的虚警分析, 有效地降低了闪光灯和快速运动对镜头边界检测产生的影响, 并实现了镜头突变检测. 在渐变类型分类中, 我们分别实现了淡入淡出、扫换和溶解等三种重要镜头边界类型的检测. 实验结果表明, 本文方法能够快速、有效地实现视频镜头边界类型的检测, 对运动和光线变化有较强的鲁棒性.

关键词: 镜头边界检测; 小波变换; 奇异点检测; 渐变; 视频

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0372-2112 (2006) 12-2272-06

A Rapid and Effective Method for Shot Boundary Detection

GENG Yu-liang, XU De, FENG Song-he

(Institute of Computer Science and Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: A rapid and effective shot boundary detection algorithm is proposed. First, the video sequence is mapped into a temporal evolution curve (TEC) by using the L_2 norm of HS histogram. Then singularity detection with wavelet is employed to identify the potential shot boundaries in TEC curve. Next then, we utilize intensity of motion intensity and block based edge direction histogram to distinguish shot boundaries from the false positives caused by rapid motion and complex flashlight scenes. At last, gradual transition is further classified into dissolve, wipe and fade in/ fade out. The experimental results compared with other methods achieve satisfying precision and recall of detecting shot boundaries in this algorithm.

Key words: shot boundary detection; wavelet transformation; singularity detection; gradual transition; video

1 引言

随着计算机和影视技术的发展, 人们接触到的视频数据以前所未有的速度增长. 面对海量的视频信息, 用来有效地分析、组织和检索海量视频数据的基于内容的视频检索技术 (Content based Video Retrieval) 吸引了众多研究者的兴趣. 在目前研究的基于内容的视频检索系统中, 一般都是先将视频分割成独立的镜头, 然后在镜头的基础上提取关键帧、建立视频摘要和索引结构, 因此, 镜头边界检测是视频结构化分析和检索的重要基础.

镜头作为视频内容的基本结构单元, 它是由摄像机一次连续拍摄的帧序列组成. 镜头边界类型有两种: 突变 (Cut) 和渐变 (Gradual Transition), 其中, 渐变又包括淡入淡出 (Fade in/out), 溶解 (Dissolve), 扫换 (Wipe) 等编辑效果. 镜头边界检测的任务就是有效地识别镜头边界类型及其位置, 其基本思想就是通过比较视频序列的帧间差异来寻找镜头边界变化的规律, 因此, 在镜头边界检测算法中, 特征度量和阈值选取是算法研究的两个主要方面^[1~7]. 当然, 如何消除噪声, 特别是光

线变化、摄像机或物体运动对算法性能的影响, 也是算法考虑的重要问题^[2~5].

在特征度量方面, 由于颜色直方图对运动和遮挡具有较强的鲁棒性, 因此被广泛地应用于帧间差异的度量^[1,4]. 另外, 边缘和运动特征在镜头边界检测中也发挥着重要作用^[2,3]. 文献[2]利用边缘方向在光线变化时变化较小的特点, 实现了复杂闪光灯的检测, 从而减小了光线变化对镜头边界检测的影响. 文献[3]运用仿射运动模型的平移分量建立视频序列的全局运动轨迹, 并利用离散曲线演化估计轨迹中的主要转折点, 从而获得镜头边界. 利用运动信息进行镜头边界检测, 可以有效地抑制物体或摄像机运动对镜头边界检测的影响. 目前, 一些科研院所开发了各自视频分析原型系统, 如: IBM Almaden 研究中心的 VideoAnnEx 系统, 它通过计算基于颜色直方图和运动直方图的帧间差来实现镜头边界检测, 并利用启发式规则来减少闪光灯和运动等噪声的影响^[4]. 为了提高算法的运行速度, 许多研究是基于压缩域特征来实现镜头边界的检测^[5], 并取得了较好的效果.

在阈值选取方面, 固定阈值虽然实现简单, 但不能很好的

适应视频类型和内容的变化^[1]. 如果阈值过低, 查全率虽然有所增加, 但是, 误检率也随之增加; 而如果阈值过高, 误检率虽然降低, 但查全率也随之降低, 因此, 在镜头边界检测的判别阶段, 最优阈值的自动选择是十分重要的^[6,7]. 文献^[6]通过分析镜头长度的概率分布, 提出了一种根据已逝时间和已检测镜头边界来动态选取阈值的方法. 文献^[7]则利用最大熵阈值选择方法来确定镜头边界检测中的最优阈值.

从以上分析可以看出, 在基于内容的视频检索系统中, 镜头边界检测得到了广泛关注, 并取得了一系列的成果, 但是, 如何消除光线变化和快速运动场景对镜头边界检测的影响依然是一个亟待解决的难题. 实现渐变类型的精确分类, 建立一个通用、鲁棒的检测算法仍然需要深入地研究. 本文针对上述研究问题, 提出了一种快速有效的镜头边界检测方法. 实验结果表明, 该方法能有效地实现镜头边界的检测, 具有较高的查全率和查准率.

2 快速有效的视频镜头边界检测方法

图 1 给出了本文提出的视频镜头边界检测方法的系统框图. 如图 1 所示, 首先, 我们利用基于小波变换的奇异点检测实现镜头边界的预检测, 从而得到视频序列的候选镜头边界; 然后, 在消除闪光灯场景产生的虚警信息的基础上, 实现了镜头突变检测; 最后, 在镜头渐变检测中, 通过对快速运动场景产生的虚警信息的分析, 有效地降低了快速运动场景对镜头边界检测产生的影响, 进而实现了淡入淡出、扫换和溶解等三种重要镜头边界类型的检测.

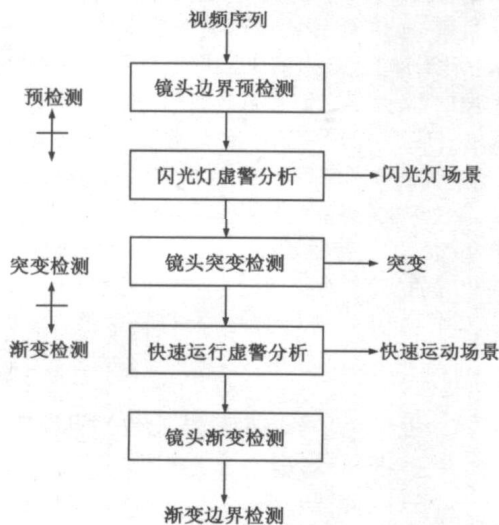


图 1 镜头边界检测方法的系统框图

2.1 基于小波变换的镜头边界预检测

视频序列可以由底层特征表示为相应特征空间中的一条曲线, 我们进一步改进已有的视频序列时间变化曲线 (TEC)^[8], 使它更能有效地表达视频的语义特征. 首先, 我们采用与人类视觉感知系统比较接近的 HSV 颜色空间来提取视频帧的颜色直方图. 由于亮度分量 (Value) 受光线变化影响较大, 而且亮度较低时, 人眼对色度 (Hue) 和饱和度 (Saturation) 分辨能力降低, 因此, 在计算颜色直方图时, 我们只使用

色度和饱和度分量, 并利用亮度对色度和饱和度进行非均匀量化, 这样, 得到颜色直方图 $HS(i, j)$:

$$HS(i, j) = \begin{cases} \sum \delta(v(x, y) = 0) / (M \times N), & i = 0, j = 0 \\ \sum \delta(h(x, y) = i, s(x, y) = j, v(x, y) = 1) / (M \times N) \end{cases} \quad (1)$$

其中, $h(x, y)$, $s(x, y)$, $v(x, y)$ 分别为色度、饱和度和亮度的量化分量. 对于亮度分量, 当像素 (x, y) 的亮度值小于 0.25 时, $v(x, y) = 0$, 否则, $v(x, y) = 1$. 对于色度和饱和度分量, 则分别均匀量化为 16 和 4 个量化区间 ($i = 0, \dots, 15, j = 0, \dots, 3$), M, N 为视频帧的尺寸. $\delta(X)$ 是判真函数, 当 X 为真时, $\delta(X) = 1$, 否则, $\delta(X) = 0$.

由颜色直方图的定义, 我们可以将视频序列看作该颜色量化空间中一条有时间标记的曲线, 每个视频帧为该曲线中的一点, 并由其颜色直方图表征. 该曲线在一定程度上表现了视频的语义特征: 在同一镜头内, 视频帧内容相似, 曲线较为平滑, 而在镜头边界处, 视频帧内容存在较大的差异, 曲线发生明显的奇变. 因此, 我们可以通过对曲线的奇异点分析实现视频镜头边界的检测. 由于在多维空间中计算复杂度较高, 我们将高维曲线映射到低维空间, 再进行奇异点分析. 实验中, 我们使用幅值映射得到视频序列的时间变化 TEC 曲线:

$$f(t) = \|HS_t\|_2 = \sum_{i,j} (HS_t(i, j))^2, \quad t = 1, 2, \dots \quad (2)$$

其中, HS_t 是第 t 个视频帧的颜色直方图. 这种降维处理既保持了信号的奇异性, 又降低了计算复杂度. 对于不同的镜头边界类型及事件, TEC 曲线表现出不同的特征:

对于突变, 曲线表现为一个阶跃, 如图 2(a) 所示; 对于渐变、大物体快速出现或消失以及摄像机快速转移到一个新的场景时, 曲线表现为上升或下降的坡度, 如图 2(b) 所示; 当闪光灯出现时, 由于其持续时间较短, 因此, 曲线表现为一个向上或向下的尖峰, 如图 2(c) 所示; 对于淡入淡出, 曲线表现为一个平台, 平台的宽度由单色持续的时间决定, 其坡度为其淡入/淡出的速度, 如图 2(d) 所示. 由于淡入淡出连接的两个镜头内容可能相差较大, 因此平台两侧并不在同一个水平线上, 它可以看作两个图 2(b) 组成; 对于连续多闪光灯场景, 曲线表现为多个尖峰, 如图 2(e) 所示. 从以上分析可以看出, 通过视频序列的 TEC 曲线, 可以较为全面地刻画视频的镜头边界特性.

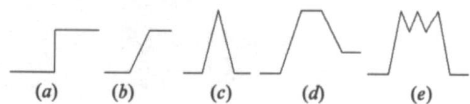


图 2 视频序列的时间变化曲线的镜头边界及事件特征

在信号处理中, 小波变换能可靠地识别信号的奇异性, 并且对噪声具有较强的鲁棒性^[9], 因此, 我们利用小波变换来分析 TEC 曲线的奇异性, 从而实现候选镜头边界的检测.

设 $\theta(t)$ 为平滑函数, 且满足条件,

$$\int_{-\infty}^{+\infty} \theta(t) dt = 1, \quad \lim_{t \rightarrow \pm\infty} \theta(t) = 0 \quad (3)$$

$\theta(t)$ 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, 如果引入符号 $\zeta_a(t) = \frac{1}{a}\zeta(t/a)$, 则 $f(t)$ 关于小波 $\Psi(t)$ 在尺度 a 和位置 t 上的规范小波变换定义为

$$\begin{aligned} W_f(t) &= f(t) * \Psi_a(t) = f(t) * \left[a \frac{d\theta_a(t)}{dt} \right] \\ &= a \frac{d}{dt} (f * \theta_a)(t) \end{aligned} \quad (4)$$

当 $\theta(t)$ 取 Gauss 函数时, $d\theta(t)/dt$ 实际上形成了 Canny 小波, Mallat 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. 实验中本文使用 Gauss 函数作为平滑函数来实现候选镜头边界的检测. 从式(4)可以看出, 当尺度 a 增大时, 信号 $f(t)$ 与平滑函数 $\theta_a(t)$ 的卷积可以消去信号中较小的变化, 因此, 能检测出比较大的变化点, 这样就可以通过计算小波变换模值曲线 $|W_f(t)|$ 的局部极大值来确定候选镜头边界.

对于 TEC 曲线在低分辨率空间中的小波变换模局部极大值点 M_i , 我们计算以 M_i 为中心的滑动窗口 w 内所有局部极大值点的均值作为阈值来确定候选镜头边界, 如式(5)所示, K 为常数 ($K = 2$), N_i 为当前滑动窗口内局部极大值点的数目. 图 6(b) 中虚线为自适应局部阈值曲线.

$$Th_i = \frac{K}{N} \sum_{M_j \in w} M_j \quad (5)$$

如果 $M_i > Th_i$ 成立, 则 M_i 为候选镜头边界 CSB_i , 其边界区间由位于该极大值点两侧, 并距离其位移最小的极小值点 (L_i, R_i) 确定. 由于在低分辨率空间缺乏足够的定位精度, 因此, 我们在高分辨率空间进一步确定候选镜头边界的边界区间. 对于相邻候选镜头边界 CSB_i, CSB_{i+1} , 如条件 $|L_{i+1} - R_i| < 10$ 成立, 则合并两个候选镜头边界, 这样候选镜头边界在小波变换模值曲线 $|W_f(t)|$ 中表现为单峰、双峰和多峰三种形式.

2.2 闪光灯场景的虚警分析

闪光灯场景主要有两种形式: 单一闪光灯和连续闪光灯, 如图 2 中(c)、(e)所示, 对应的 $|W_f(t)|$ 曲线表现为双峰和多峰. 对于第一种形式, 由于其持续时间短(通常是一帧), 因此, 我们可以通过对 TEC 曲线进行 3×1 窗口的中值滤波来消除该类闪光灯产生的虚警. 对于第二种形式, 采用我们提出的基于分块边缘方向直方图的方法来识别连续闪光灯^[10].

2.3 镜头突变检测

如图 2(a) 所示, TEC 曲线在突变位置表现为一个阶跃, 而 $|W_f(t)|$ 曲线则表现为一个单峰. 因此, 对于给定候选镜头边界 CSB_i , 及其边界区间 $[L_i, R_i]$, 如果它在 $|W_f(t)|$ 曲线中表现为一个单峰, 我们计算

$$N_{\max} = \sum_{t \in [L_i, R_i-1]} \delta(D_{ff}(t) > \mu + \sigma) \quad (6)$$

$$D_{ff}(t) = |f(t+1) - f(t)| \quad t \in [L_i, R_i-1]$$

如果 $N_{\max} = 1$, 则该 CSB_i 为突变镜头边界, 边界位置为 $\arg \max(D_{ff}(t))$, 其中, μ, σ 为差值序列 D_{ff} 的均值和方差.

2.4 镜头渐变检测

从 2.1 节的分析可以看出, 摄像机快速运动, 大物体的快

速出现或消失(统称为快速运动场景)与镜头渐变边界具有相似的 TEC 曲线, 并且它们的 $|W_f(t)|$ 曲线均表现为单峰, 如图 2(b) 所示, 因此, 在进行镜头渐变检测之前, 首先需要识别由快速运动场景产生的虚警信息.

2.4.1 快速运动场景的虚警分析

我们注意到, 在镜头渐变过程中, 视频内容通常处于静止或慢速运动状态, 运动强度(Motion Intensity)较小, 而快速运动场景则具有较大的运动强度. 因此, 我们可以通过计算候选镜头边界区间中视频片段的运动强度来识别运动场景产生的虚警. 由于运动矢量幅度的均值和方差是一种简单而有效的运动强度的度量方法^[11], 而运动矢量幅度的均值能够同时表征摄像机和局部物体运动产生的运动强度, 因此, 我们在虚警分析中采用运动幅度的均值来识别快速运动场景. 对于候选镜头边界 CSB_i , 我们计算其边界区间 $[L_i, R_i]$ 中视频序列的运动强度, 如式(7)所示:

$$avrMI = \sum_{t \in [L_i, R_i]} MI(t) / (R_i - L_i + 1) \quad (7)$$

其中, $MI(t)$ 为第 t 帧运动矢量幅度的均值, 运动矢量幅值均匀量化为 5 个等级. 如果 $avrMI > TH_M$, 则 CSB_i 为快速运动场景产生的虚警, 否则为渐变边界, 实验中, 阈值 $TH_M = 3$.

2.4.2 镜头渐变检测

在镜头突变检测和虚警分析之后, 我们进一步实现三种主要镜头渐变类型, 即: 淡入淡出、扫换和溶解边界类型的分类.

淡入淡出是影视剪辑中一种重要的镜头边界类型, 淡入是指镜头的首帧缓慢地从全单色(通常为黑色)屏幕中出现, 而淡出是指镜头的末帧缓慢地消失为全单色(通常为黑色). 如图 2(d) 所示, 淡入、淡出的 TEC 曲线分别表现为一个上升和下降的坡度, 又由于它们常常成对出现, 因此, 淡入淡出在 TEC 曲线中表现为一个平台, 平台的宽度由单色帧持续的时间决定. 这里我们分别实现淡入、淡出边界的检测. 按式(2)给出的 TEC 曲线定义, 当视频帧趋于单色时, $f(t)$ 趋于 1, 这是其他渐变类型所不具备的, 因此, 我们利用如下公式实现淡入淡出镜头边界的检测. 对于给定候选镜头边界 CSB_i , 及其边界区间 $[L_i, R_i]$, 如果式(8.1)成立, 则 CSB_i 为淡入镜头边界.

$$\max_{t \in [L_i, R_i]} (f(t)) > 0.95, \quad \sum_{t \in [L_i, L_i + \varepsilon]} f(t) / \varepsilon > 0.9 \quad (8.1)$$

其中, $\varepsilon = (R_i - L_i) / 4$. 如果式(8.2)成立, 则 CSB_i 为淡出镜头边界.

$$\max_{t \in [L_i, R_i]} (f(t)) > 0.95, \quad \sum_{t \in [R_i - \varepsilon, R_i]} f(t) / \varepsilon > 0.9 \quad (8.2)$$

图 3 给出了淡入淡出镜头边界的一个实例, 上图为 TEC 曲线, 下图为 $|W_f(t)|$ 曲线及检测结果(星号表示边界中心, 灰色方块为检测到的镜头边界的起止位置, 虚线为局部阈值), 视频序列为 UGS09_001 的前 300 帧, 3 个镜头边界分别为淡入、淡出和淡入.

扫换是一种复杂镜头边界类型, 一般有两种扫换方式, 其一是指后一镜头的首帧, 如图 4(a) 所示, 逐渐滑入屏幕而取代前一镜头的末帧, 如图 4(b) 所示; 其二是后一镜头的首帧逐区域地显示而取代前一镜头的末帧, 如图 4(c) 所示, 图

中 1, 2, 3, 4 表示视频帧中 4 个区域的编号. 按照扫描方向又可以分为水平、垂直、对角以及扩散等多种类型. 我们以水平扫描为例, 并参考文献 [12] 中的扫描定义, 给出图 3 中 (b), (c) 中两种扫描方式的公式化定义, 如式 (9.1), (9.2) 所示. 现有扫描检测算法一般仅考虑式 (9.2) 所示情况^[12, 13].

$$g(x, y, t) = \begin{cases} S_1(x, y), & (x, y) \in R_w \\ S_2(Width - x, y), & (x, y) \notin R_w \end{cases} \quad (9.1)$$

$$g(x, y, t) = \begin{cases} S_1(x, y), & (x, y) \in R_w \\ S_2(x, y), & (x, y) \notin R_w \end{cases} \quad (9.2)$$

其中, $g(x, y, t)$ 为时刻 t , 像素 (x, y) 的灰度; $S_1(x, y)$ 为前一镜头末帧的灰度值; $S_2(x, y)$ 为后一镜头首帧的灰度值; $Width$ 为视频帧的宽度; R_w 为前一镜头末帧中未消失的区域.

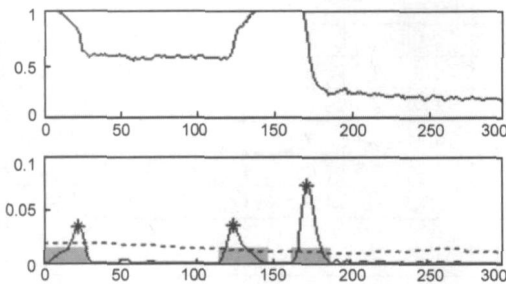


图 3 淡入/淡出镜头边界的特征曲线

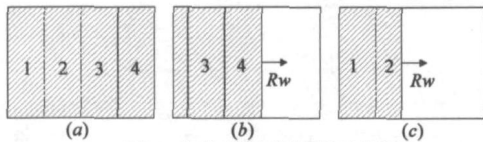


图 4 扫描镜头边界示意图

这里, 提出一种更为有效的、通用的扫描检测方法. 对于给定候选镜头边界 CSB_i , 及其边界区间 $[L_i, R_i]$, 我们计算 $g(x, y, L_i)$ 与其后续视频帧 $g(x, y, t)$ 的帧间像素差, 二值化后, 分别统计二值图像中每行的 0 游程长度, 如果 0 游程长度大于视频帧的宽度 $2/3$, 记为长游程, 这样, 得到二值图像的长游程数目 $LongRun(L_i, t), t \in (L_i, R_i)$. 同理, 我们计算 $g(x, y, R_i)$ 与其前面视频帧 $g(x, y, t)$ 的帧间像素差, 并得到长游程数目 $LongRun(R_i, t), t \in (L_i, R_i)$. 为了减少物体运动的影响, 在计算帧间差之前, 先用 3×3 窗口的均值滤波器对帧像素进行平滑处理. 候选镜头边界的长游程曲线能够有效地表

征式具有不同扫描方式和方向的扫描镜头边界, 而且显著地区别于其他镜头边界类型, 如淡入淡出、溶解等. 图 5(c), (f) 分别给出了扫描和溶解两种镜头边界类型的长游程曲线, 实线为曲线 $LongRun(L_i, t)$, 虚线为曲线 $LongRun(R_i, t)$. 数据来自视频序列 BOR10_005 中的第 1 个和第 2 个镜头边界, 图 5(a) 和 (b) 为扫描镜头边界中的两帧, 图 5(d) 和 (e) 为溶解镜头边界中的两帧.

对于候选镜头边界 CSB_i , 如果式 (10.1) 成立, 则 CSB_i 为扫描镜头边界.

$$\sum_{t \in [L_i, L_i + \varepsilon]} LongRun(L_i, t) / \varepsilon > Height / 2 \quad \text{或者} \quad (10.1)$$

$$\sum_{t \in [R_i, R_i - \varepsilon]} LongRun(R_i, t) / \varepsilon > Height / 2$$

其中, $\varepsilon = (R_i - L_i) / 4$, $Height$ 为视频帧的高度. 如果式 (10.2) 成立, 则 CSB_i 为溶解镜头边界.

$$\sum_{t \in [L_i, L_i + \varepsilon]} LongRun(L_i, t) / \varepsilon < Height / 2 \quad \text{并且} \quad (10.2)$$

$$\sum_{t \in [R_i, R_i - \varepsilon]} LongRun(R_i, t) / \varepsilon < Height / 2$$

这样, 我们得到镜头渐变效果的精确分类.

3 实验研究

为了验证本文提出的镜头边界检测方法的有效性, 选取了 2 段 BBC 新闻视频 (BBC221203、BBC260705), 3 段 www.operr-video.org 提供的中速运动视频 (BOR10_005、UGS09_001 和 DynamicA1956) 以及 1 段电影视频 (Star Wars II, 剪辑自《星球大战 II》CD1, 包括 9 种扫描类型), 视频帧为每秒 30 帧. 视频中包含显著的全局运动和局部运动, 以及连续的闪光灯场景. 图 6 给出了部分视频数据的实验结果. 图 6(a) 为视频序列的时间变化曲线 (TEC 曲线), 其中, 灰色方块表示视频序列中真实镜头边界的起止位置, 包括突变 (C)、溶解 (D)、扫描 (W)、淡入 (FI) 和淡出 (FO). 图 6(b) 给出了小波变换模值曲线 (曲线) 及检测结果, 其中, 星号表示检测到的候选镜头边界, 虚线为局部阈值曲线, 红色线段为检测到的突变镜头边界, 灰色方块为检测到的渐变镜头边界的起止位置. 由于视频序列在视频帧 380-500 区间中存在快速的近景人物运动, 而在视频帧 1300-1500 区间中存在连续 12 个闪光灯, 因此, 在镜头边界预检测阶段分别产生了 1 个运动虚警和 2 个闪光灯虚警. 我们在闪光灯和快速运动场景的虚警分析中, 有效地消除了该虚警信息, 从而得到了准确的镜头边界.

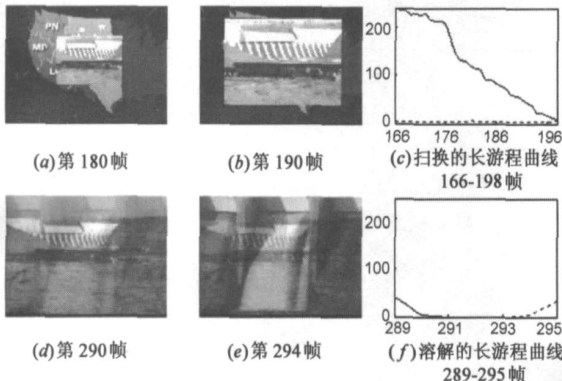


图 5 扫描和溶解镜头边界类型的长游程曲线

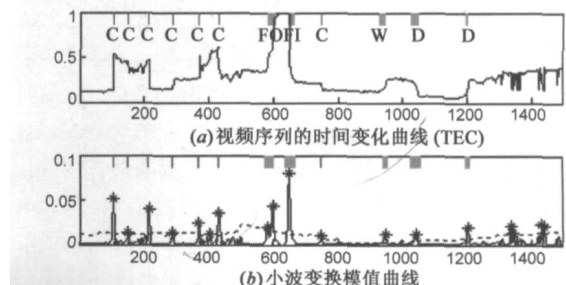


图 6 视频序列的时间变化曲线及小波变换模值曲线

在实验中,我们采用查全率(Recall)和查准率(Precision)来评价算法的性能.表1给出了实验的统计结果,其中包括对突变(C)、溶解(D)、扫换(W)、淡入淡出(F)等4种主要镜头边界类型的检测结果的统计.滑动窗口的大小为300帧,小波尺度=3.实验结果显示,本文提出的方法对复杂闪光灯、快速的摄像机和物体运动场景具有较强的鲁棒性,对突变具有较高的查全率和查准率,并能有效的识别3种主要的渐变镜头边界类型.这主要是因为:(1)在特征选择上,视频序列的时间变化曲线能有效地表征不同的镜头边界类型、闪光灯及快速运动等场景;(2)基于小波变换的镜头边界的预检测方法以及自

适应局部阈值的选择策略能有效地检测出潜在的镜头边界,而且在一定程度上抑制了闪光灯和运动产生的噪声;(3)闪光灯和快速运动场景的虚警分析进一步降低了误检率;(4)渐变检测方法实现了淡入淡出、扫换和溶解的准确分类.当然,算法也存在一些缺点,需要我们进一步改进.如在视频UGS09_001中,由于大多数场景是灰色图像,而我们的方法是以HS颜色特征为基础,这导致镜头间特征相近,从而使镜头边界检测的查准率降低.这需要在兼顾算法运行效率的同时,提取更为有效的特征或建立有效的虚警检测方法.

表1 镜头边界检测算法的实验结果

视频文件	帧数	实际边界数				漏检数				误检数				查准率(%)		查全率(%)	
		C	D	F	W	C	D	F	W	C	D	F	W	突变	渐变	突变	渐变
BBC221203	19644	87	7	0	0	7	1	0	0	9	2	0	0	89.9	75	92	85.8
BBC260705	10740	34	3	1	0	2	0	0	0	4	2	0	0	88.9	66.7	94.1	100
BOR10_005	3576	0	20	0	2	0	1	0	0	0	1	0	0	—	95.5	—	95.5
UGS09_001	3650	3	12	3	0	0	1	0	0	2	4	0	0	60	77.8	100	93.3
DynamicA1956	25830	44	2	16	0	2	0	0	0	6	3	0	0	87.5	85.7	95.5	100
Star Wars II	4473	16	0	0	17	0	0	0	3	1	3	0	1	94.1	77.8	100	82.4
合计	67913	184	44	20	19	11	3	0	3	22	15	0	1	88.7	82.8	94	92.8

表2 三种镜头边界检测算法检测性能比较

镜头边界类型	视频文件	帧数	双阈值算法(%)		VideoAnnEx(%)		本文算法(%)	
			查准率	查全率	查准率	查全率	查准率	查全率
突变	BBC221203	19644	66.3	70.1	94.2	74.7	89.9	92
	BBC260705	10740	74.3	76.5	100	61.8	88.9	94.1
	DynamicA1956	25830	78.3	81.8	94.7	81.8	87.5	95.5
	合计	56214	71.1	74.5	95.3	73.9	89	93.3
渐变	BOR10_005	3576	73.3	100	91.7	100	95.5	95.5
	UGS09_001	3650	75	80	88.9	53.3	77.8	93.3
	DynamicA1956	25830	73.7	77.8	93.8	83.3	85.7	100
	合计	33056	73.8	87.3	91.8	81.8	86.9	96.4

表2给出了本文方法与双阈值算法^[1]和VideoAnnEx系统中镜头边界检测算法^[4]的对比实验结果.从实验结果可以看出,双阈值算法虽然实现简单,但是,在渐变检测中具有较高的误检率,而且阈值的选择对算法性能影响较大.文献[4]算法虽然对运动和闪光灯场景具有较高的鲁棒性,但同时取得

了较低的查全率.从视频结构分析的角度来看,由于镜头边界检测是场景分析的基础,在场景分析中可以有效地合并过分割的镜头,因此,查全率相对于查准率更为重要.本文方法即取得了较好的查准率,同时也取得了较高的查全率,并且实现了三种重要渐变边界类型的准确分类.

表3 三种镜头边界检测算法的时间性能比较

视频文件	帧数	分辨率	镜头边界数(突变/渐变)	双阈值算法(s)	VideoAnnEx(s)	本文算法(s)
BBC221203	19644	320×240	87/7	14	45	35
BBC260705	10740	320×240	34/4	8	26	21
DynamicA1956	25830	368×480	44/18	43	138	149
BOR10_005	3576	352×240	0/22	3	9	12
UGS09_001	3650	352×240	3/15	3	10	12

最后,我们给出三种算法在不同视频序列下的时间性能比较,如表3所示,所有实验在VC++ 6.0,1.7G/496M RAM实验平台下进行.从实验结果可以看出,由于双阈值算法仅仅通过计算颜色直方图帧间差来实现镜头边界的检测,因此具有较低的时间复杂度;本文方法和文献[4]算法具有较高的时间复杂度,这主要是因为这两种算法为提高镜头边界检测的查准率,均在系统中增加了运动特征的分析,从而降低了运行效

率.而文献[4]在提高查准率的同时也显著地降低了查全率.综上所述,本文提出的方法在综合性能上取得较好的效果.

4 结论

本文提出了一种层次化镜头边界检测方法,首先利用小波变换实现了镜头边界的预检测,然后,通过对候选镜头边界的虚警分析,有效地降低了闪光灯和快速运动对镜头边界检

测产生的影响,并实现了镜头突变检测.在渐变类型分类中,通过分析候选镜头边界的时间变化曲线和长游程曲线,实现了淡入淡出、扫换和溶解等三种重要渐变类型的准确分类.实验结果表明,该方法能有效地实现镜头边界的检测,具有较高的查全率和查准率.

参考文献:

- [1] H J Zhang, A Kankanhalli, S W Smoliar. Automatic partitioning of full motion video[J]. *Multimedia Systems*, 1993, 1(1): 10–28.
- [2] W J Heng, K N Ngan. High accuracy flashlight scene determination for shot boundary detection[J]. *Signal Processing: Image Communication*, 2003, 18(3): 203–219.
- [3] S Tretasanatavom, J Heuer, et al. Temporal video segmentation using global motion estimation and discrete curve evolution [A]. *IEEE ICIP 2004*[C]. Singapore: IEEE Press, 2004. 385–388.
- [4] A Amir, M Berg, et al. IBM research TRECVID 2003 video retrieval system[A]. *TRECVID Workshop*[C]. USA: Washington D C, 2003.
- [5] W S Chau, O C Au, et al. Efficient and reliable scene change detection in MPEG compressed video[A]. *Visual Communications and Image Processing 2005*[C]. Beijing, China: SPIE, 2005. 5960: 1542–1549.
- [6] S Youm, W Kim. Dynamic threshold method for scene change detection[A]. *IEEE ICME 2003*[C]. Baltimore, Maryland: IEEE Press, 2003. 337–340.
- [7] K W Sze, K M Lam, G Qiu. Scene cut detection using the colored pattern appearance model[A]. *IEEE ICIP 2003*[C]. Barcelona, Spain: IEEE Press, 2003. 2: 1017–1020.
- [8] 耿玉亮, 须德. 视频镜头边界检测的统一策略[J]. *中国图像图形学报*, 2005, 10(5): 650–655.
- Geng Yr liang, Xu De. A Unified Framework For Shot Boundary Detection[J]. *Journal of Image and Graphics*, 2005, 10(5): 650–655. (in Chinese)

- [9] 彭玉华. 小波变换与工程应用[M]. 北京: 科学出版社, 1999: 165–173.
- PENG Yr hua. Wavelet transform and its application in engineering[M]. Beijing: Science Press, 1999: 165–173. (in Chinese)
- [10] Y L Geng, D Xu. A solution to illumination variation problem in shot detection[A]. *IEEE TENCON 2004*[C]. Chiang Mai, Thailand: IEEE Press, 2004. 1: 81–84.
- [11] K A Peker, A Divakaran. Automatic measurement of intensity of motion activity of video segments[A]. *SPIE on Storage and Retrieval for Media Databases 2001*[C]. San Jose, CA: SPIE, 2001. 341–351.
- [12] H H Yu, W Wolf. A hierarchical multiresolution video shot transition detection scheme[J]. *Computer Vision and Image Understanding*, 1999, 75(1/2): 196–213.
- [12] W A C Fernando. Wipe scene change detection and classification in video sequences[J]. *Journal of Electronic Imaging*, 2004, 13(2): 362–375.

作者简介:



耿玉亮 男, 1975 年出生于黑龙江双鸭山, 2002 年起在北京交通大学攻读博士学位, 主要研究方向为视频内容分析。
E-mail: gengyuliang@sohu.com



须德 男, 1944 年出生于江苏常州, 北京交通大学教授, 博士生导师, 主要研究方向为数据库系统及应用和多媒体信息处理。
E-mail: xd@computer.njtu.edu.cn