

基于基色调的镜头边界检测方法

智敏^{1,2} 蔡安妮²

摘要 视频结构化组织是建立视频检索和浏览系统的基础,而镜头边界检测是视频结构化的第一步。在本文中,我们提出了基于基色调的镜头边界检测方法。该方法首先利用 I 帧的比特数信息,减少参与检测的 I 帧;然后在考虑基色调的基础上得到局部的自适应阈值;最后用双阈值方法检测镜头边界。实验证明,该方法可以对长视频序列进行较好的镜头突变和渐变检测,并且减少了计算量,同时能够排除大物体运动和摇镜头对镜头检测的影响。

关键词 镜头边界,基色调,自适应阈值,比特数
中图分类号 TP391

Shot Boundary Detection with Main Color

Zhi Min^{1,2} Cai An-Ni²

Abstract Shot boundary detection is one of the critical techniques in video browsing and indexing systems. In this paper, we propose a main color based shot boundary detection algorithm on DC images. We use the bit-rate information to decrease the number of I frames that take part in the shot change detection. Then we obtain local adaptive thresholds based on the main color of a segment of the video sequence. Finally, we detect shot change based on twin-comparison method. Experimental results show that our algorithm can detect shot boundary from a long video sequence with reduced computation, and at the same time solve the problems that are caused by object moving and camera panning in shot change detection.

Key words Shot boundary, adaptive threshold, bit-rate, main color

1 引言

随着多媒体压缩技术的发展、计算机处理能力的提高以及存储介质价格的降低,人们能够接触到的视频信息也越来越多。为了在大量的视频信息中找到感兴趣的内容,需要建立视频检索和浏览系统。视频结构化组织是建立这种系统的基础,而镜头边界检测是视频结构化的第一步。

目前镜头边界检测一般都直接利用压缩视频数据的某些特征展开的,如 DC 系数^[1]、DCT 系数^[2]、运动矢量^[3]、宏块编码^[4]和比特数^[5]等。研究在半解码基础上直接对镜头边界进行检测的算法具有省去全解码时间和充分利用压缩数据的统计结果等优点。

镜头边界检测的主要根据是镜头内部的相邻帧之间具有很大的相似性,当发生镜头边界时,该相似性将被破坏。镜头突变和渐变对相似性的破坏程度是不同的,因此使用一个阈值,难以同时检测出突变和渐变。如果阈值过小,就会出现过检测;而如果阈值过大,就无法检测出渐变的镜头边界。文献

[1] 最早提出镜头渐变检测的双阈值比较的方法。该方法设置两个阈值 T_h 和 T_l , 其中 $T_h > T_l$ 。首先计算相邻两帧的帧差,当帧差大于 T_h 时存在镜头突变;当帧差小于 T_h 而大于 T_l 时可能存在镜头渐变,此时认为该帧差的后一帧为渐变边界的潜在起始帧;然后同时计算两种帧差:一种帧差是相邻两帧的帧差 $fd_{k,k+1}$;另一种帧差是相隔帧的帧差 fd_m ,即相隔 m 帧的帧差。当相隔帧的帧差 fd_m 超过 T_h 而相邻帧的帧差低于 T_l 时,这一帧便为镜头渐变的终止帧。在相隔帧差开始计算后,同时观察相邻帧的帧差 $fd_{k,k+1}$,如果 $fd_{k,k+1}$ 小于 T_l ,而相隔帧的帧差 fd_m 不超过 T_h 时,废除该潜在起始帧,接着重新寻找新的起始帧。该方法对渐变和突变检测都具有较好的效果,但是其阈值是经验值,因此该方法不适合长视频序列的镜头边界检测。

2 MPEG 编码中的比特数

对于 MPEG 压缩域中的镜头边界检测,最自然的解决方法是利用与像素域相关的 DC 系数,即用 MPEG 数据流中的离散余弦变换的直流分量(DC 系数)组成原图像的缩微图(又称 DC 图)检测镜头边界。I 帧的 DC 系数可以直接从 MPEG 视频流中获得,而 P 和 B 帧的 DC 系数不能直接从 MPEG 视频流中获得。因此本文中我们只考虑 I 帧。I 帧是内部编码帧,即编码时不需要任何参考帧,因此一个 I 帧编码所需的比特数反映了该 I 帧内容的复杂性。对于高活动度的帧,它所需要的比特数会比低活动度的帧多。但连续的两个 I 帧编码所需的比特数没有明显变化时,也并不说明一定没有内容变换。对不同内容的两个 I 帧,即使全局特征是相同的,局部特征却一定是不同的。因此只要镜头内容发生了变化,在局部特征上就会表现出来。因此我们定义了下面两个公式来判断相邻 I 帧是否发生了内容变换

$$d(R_n, R_{n+1}) = \sum_{i,j} |R_n(i, j) - R_{n+1}(i, j)| \quad (1)$$

其中 $R_n(i, j)$ 表示帧 R_n 的第 (i, j) 宏块的比特数, $d(R_n, R_{n+1})$ 是帧 R_n 和 R_{n+1} 的比特数差值。

$$RI_n = \frac{d(R_n, R_{n+1})}{T_n} \quad (2)$$

其中 T_n 是帧 R_n 编码所需的比特数。 RI_n 的大小可以反映相邻 I 帧的内容是否发生了变换。一个具有 31 个突变 420 个 I 帧的视频序列,其 RI_n 分布如图 1 所示(见下页顶部)。 RI_n 值大于 0.4 的点共有 165 个,其中包括 31 个真正的镜头边界点。即在检测镜头边界时,255 个 I 帧可以忽略不计。适当地选取 RI_n 的范围,就可以达到减少参与镜头边界检测的 I 帧个数,从而减少了计算 DC 图的时间。

3 基于基色调的自适应阈值

将阈值作为重要参数的镜头边界检测方法中,怎样选取最佳阈值是一个重要问题。基本上,阈值的选取方法有两种。一是通过实验预先给定经验阈值;二是根据输入的视频数据自动获取阈值。由于视频序列的不同风格,在镜头边界检测中预先给定一个适合的固定阈值是很困难的,因此镜头检测中需要自适应地选取阈值。在本文中通过比较相邻两帧的颜色直方图差值,定义基色调来获取自适应的阈值。

3.1 基色调

在定义基色调之前,首先看一下在同一个电影中分别以黎明和太阳落山时分作为背景的两个视频片断的部分帧图

收稿日期 2005-9-14 收修改稿日期 2006-8-14
Received September 14, 2005; in revised form August 14, 2006
国家自然科学基金(60462001)资助
Supported by National Natural Science Foundation of P. R. China (60462001)
1. 内蒙古师范大学计算机与信息工程学院 呼和浩特 010020 2. 北京邮电大学电信工程学院 北京 100876
1. Department of Computer, Inner Mongolia Normal University, Hohhot 010020 2. School of Telecommunications Engineering, Beijing University of Posts and Telecommunications, Beijing 100876
DOI: 10.1360/aas-007-0655

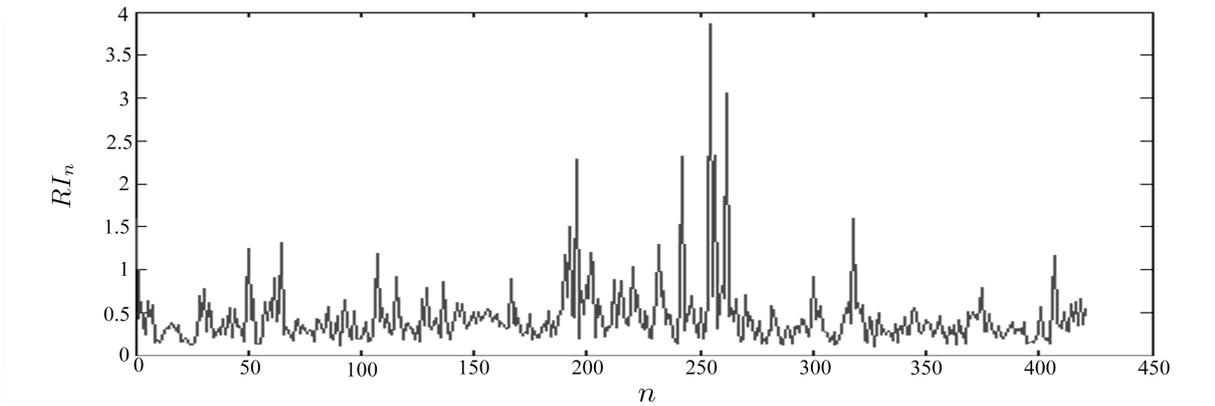


图 1 RI_n 分布图
Fig.1 Distribution of RI_n

像, 如图 2 和图 3. 在黎明的视频片断中, 每个帧图像的主演色是灰色; 在太阳落山视频片断中, 每个帧图像的主演色是黄色. 我们把视频片断的主演色定义为视频片断的基色调. 图 4 和图 5 给出了这两个视频片段的帧间颜色直方图差值的分布. 突变检测的阈值显然是不同的, 分别为 0.11 和 0.24. 由于视频片段的基色调不同, 对不同的电影甚至对同一个电影都不能只使用一个全局的阈值进行镜头切换检测.



图 2 黎明时分的视频片断
Fig.2 Video sequence at daybreak



图 3 太阳落山时分的视频片断
Fig.3 Video sequence at sundown

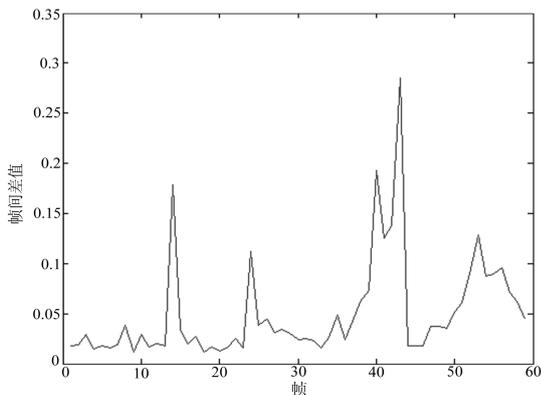


图 4 帧间差分布图
Fig.4 Distribution of frame differences

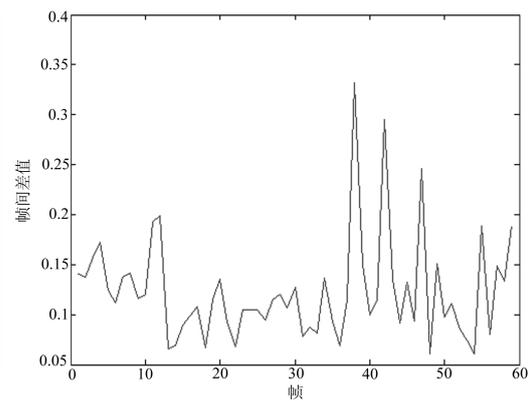


图 5 帧间差分布图
Fig.5 Distribution of frame differences

3.2 自适应的阈值

为了计算自适应的局部阈值, 我们设置了一个滑动窗口, 该滑动窗口的大小 W 是这样定义的: 窗口的最大值 W_{max} 是固定的, 最小值为 1, 其变化大小由视频序列基色调的变化来决定. 如果当前视频片段没有基色调的变换, 则窗口的大小为最大值 W_{max} ; 如果在第 n 帧和第 $n+1$ 帧 ($n < W_{max}$) 发生了基色调变换, 则窗口的大小为 n . 两个阈值 T_h 和 T_l ($T_h > T_l$) 分别定义为

$$T_h = \lambda_1(\mu - \sigma) \tag{3}$$

$$T_l = \lambda_2\mu \tag{4}$$

其中 $\mu = \frac{1}{W} \sum D(f_i, f_j)$, $\sigma = \frac{1}{W} \sum (\mu - D(f_i, f_j))^2$, $D(f_i, f_j)$ 为两个帧的颜色直方图差值, λ_1 和 λ_2 为比例参数, 一般选择为 1.5~2.

4 镜头边界检测算法

我们利用双阈值方法检测镜头边界, 与文献 [1] 相比除了选取阈值的方法不同外, 还有其它的不同. 在镜头边界检测时由于只考虑了 I 帧, 因此在实验中发现: 1) 渐变的起始帧与它前一个帧的帧间差有可能大于 T_h ; 2) 当一个镜头内大物体一直在运动或快速摇镜头时, 这段视频的帧间差会持

续大于 T_h 或 T_l , 但是相隔帧的帧差却不是单调的. 为了解决这两个问题, 对文献 [1] 的方法进行了如下改进:

1) 当前相邻帧的帧差大于 T_h 且下一个相邻帧的帧差小于 T_l , 发生突变, 否则转 2);

2) 当前相邻帧的帧差大于 T_h 或 T_l , 则把当前镜头的起始帧作为潜在起始帧; 然后同时计算相邻帧的帧差和相隔帧的帧差, 当相隔帧的帧差大于 T_h , 而相邻帧的帧差小于 T_l 时, 这一帧便为镜头渐变的终止帧. 在相隔帧差 fd_m 开始计算后, 同时观察相邻帧的帧差 $fd_{k,k+1}$ 和 fd_m , 如果 fd_m 不是单调增加, 或 $fd_{k,k+1}$ 小于 T_l 且相隔帧的帧差 fd_m 不大于 T_h 时, 废除该潜在起始帧, 即在此位置没有发生渐变. 接着重新寻找新的起始帧.

5 实验结果

为了检验算法, 我们从不同故事片里选取了部分视频序列. 首先利用比特数信息减少了需检测的 I 帧个数, 表 1 给出了部分实验的结果. 从结果可知, I 帧减少的个数可以在 20% 左右, 即减少了计算 I 帧 DC 图的时间. 其次在计算阈值的同时检测镜头边界. 本文中采用了查全率和查准率两个参数来评价镜头边界检测结果.

表 2 给出了文献 [1] 和本文算法比较的实验结果, 从表中可以看出, 对短视频片断检测方法基本相同, 但是对长视频序列时文献 [1] 的检测率只有 50% 左右. 对具有 3980 个 I 帧, 282 个突变和 14 个渐变视频片段, 本文算法的实验结果如表 3 所示. 突变检测的查全率和查准率都超过了 94%; 而且渐变检测的查全率达到了 92.8%.

表 1 利用比特数信息减少 I 帧个数的实验结果
Table 1 I frame decreasing using bit rate

视频序列代号	I 帧个数	减少的个数	减少的比例
1	484	151	31%
2	557	90	16%
3	531	113	21%
4	502	77	15%
总计	2074	437	21%

表 2 文献 [1] 和本文方法的比较结果

Table 2 Comparison results between [1] and our method

	文献 [1] 的方法		本文方法	
	查准率 (%)	查全率 (%)	查准率 (%)	查全率 (%)
短视频 (20 个突变)	94	98	94.9	98.8
长视频 (84 个突变)	52.3	55.4	95	98.7

表 3 基于基色调的镜头边界检测算法实验结果

Table 3 Result of main color based shot boundary detection

实际镜头个数		漏检数		误检数		查全率 (%)		查准率 (%)	
突变	渐变	突变	渐变	突变	渐变	突变	渐变	突变	渐变
282	14	15	1	4	7	94.8	92.8	98.5	65

6 结论

镜头边界检测是实现基于内容的视频检索的重要步骤之一, 本文提出在部分解码的基础上, 基于基色调的镜头边界检测方法. 该方法可以对长视频序列进行较好的镜头突变和渐变检测, 并且减少了计算量, 同时能够排除大物体运动和摇镜头对镜头检测的影响.

References

- Zhang H J, Kankanhalli A, Smoliar S W. Automatic partitioning of full motion video. *Multimedia System*, 1993, 1(1): 10~28
- Zhang H J, Low C Y, Gong Y H. Video parsing using compressed data. In: *Proceedings of SPIE Conference on Image and Video Processing II*. New York, USA, 1994. 142~149
- Yeo B, Liu B. Rapid scene analysis on compressed video. *IEEE Transactions on Circuits and Systems for Video Technology*, 1995, 5(6): 533~544
- Zhang H J, Low C Y, Smoliar S W. Video parsing and browsing using compressed data. *Multimedia Tools and Applications*, 1995, 1(1): 89~111
- Koprinska I, Carrato S. Hybrid rule based neural approach for segmentation of MPEG compressed video. *Multimedia Tools and Applications*, 2002, 18(5): 187~212

智敏 内蒙古师范大学副教授. 主要研究方向为基于内容的视频检索, 视频检索和多媒体通信. 本文通信作者. E-mail: zhiminc@sohu.com (ZHI Min Associate professor at Inner Mongolia Normal University. Her research interest covers content based video retrieval, video abstraction, and multimedia communication. Corresponding author of this paper.)

蔡安妮 北京邮电大学教授. 主要研究方向为多媒体通信, 流媒体技术, 图像处理和图像识别. (CAI An-Ni Professor at Beijing University of Posts and Telecommunications Her research interest covers multimedia communication, image processing, and image recognition.)