

一种在 H. 264/AVC 压缩域中检测镜头边界的方法

夏定元^{1),2)} 谢惠琳¹⁾

¹⁾(武汉理工大学信息工程学院, 武汉 430070) ²⁾(桂林电子科技大学信息与通信学院, 桂林 541004)

摘要 针对像素域和 MPEG 压缩域镜头边界检测方法的不足,充分利用 H. 264/AVC 视频编码标准的新特性,提出一种从 H. 264/AVC 压缩码流中直接检测镜头边界的方法。首先,对图像帧做 4×4 块划分,然后,统计相邻帧对应块的帧内预测模式差异或帧间预测模式差异,最后,通过自适应阈值判决,实现对不同类型的视频镜头边界检测。实验结果表明,该方法具有良好的查全率和查准率。

关键词 H. 264/AVC 标准 镜头边界检测 预测模式 自适应阈值

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2009)12-2595-04

Shot Boundary Detection Based on H. 264/AVC Compressed Domain

XIA Ding-yuan^{1),2)}, XIE Hui-lin¹⁾

¹⁾(School of Information Engineering, Wuhan University of Technology, Wuhan 430070)

²⁾(School of Information & Communication, Guilin University of Electronic Technology, Guilin 541004)

Abstract Aiming at some defect in pixel domain and the MPEG compressed domain for shot boundary detection, making full use of new characteristics of the H. 264/AVC video coding standard, in this paper, we present a shot boundary detection method based on the H. 264/AVC compressed domain. Firstly, the frame should be decomposed into 4×4 blocks, and then the difference information between the corresponding blocks in the intra-prediction mode or the inter-prediction mode is counted up. At last, the shot boundary detection is accomplished for different types of video by using adaptive threshold decision. Experimental results show that the proposed method achieves satisfactory performance of recall and precision.

Keywords H. 264/AVC, shot boundary detection, prediction mode, adaptive threshold

1 引言

2003 年,ITU-T/ISO 正式公布了 H. 264/AVC 数字视频编码标准^[1]。与以往的视频编码标准相比, H. 264/AVC 采用了许多新技术^[2],如不同尺寸宏块分割、帧内预测编码模式、多参考帧以及 CABAC 编码等,虽增加了算法复杂度,但性能上却得到了极大提升,显示出优异的压缩性能和良好的网络亲和性。同时,还兼顾了不同速率和清晰度要求,可以满足多种场合的应用需求。因此,越来越多的压缩视频采

用 H. 264/AVC 标准。

对数字视频进行镜头边界检测并提取代表镜头的关键帧,从而得到该视频的摘要,是基于内容的视频检索技术的前提。过去基于内容的视频检索主要在像素域或基于 MPEG 的压缩域实现。传统的像素域方法对压缩视频进行镜头边界检测,需要先解压再做处理,检测效率较低;现有的基于 MPEG 压缩域的检测方法因无法直接应用于 H. 264/AVC 码流中,不能充分利用 H. 264/AVC 的新特性。如何在 H. 264/AVC 压缩域直接进行视频镜头边界检测已成为视频检索研究领域的新热点。

基金项目:广西科学研究与技术开发计划项目(05112001-7A2)

收稿日期:2009-06-19;改回日期:2009-09-18

第一作者简介:夏定元(1963 ~),男,教授。2004 年于华中科技大学获通信与信息系统专业博士学位。主要研究方向为多媒体通信、多媒体信息处理、图像处理与机器视觉。E-mail:labxiady@263.net

2 现有算法分析

针对 MPEG 压缩视频,国内外研究人员提出了许多直接对其进行镜头边界检测的算法,主要包括基于 DCT(离散余弦变换)系数的方法^[3]、基于 DC 图(利用 DCT 直流系数构成的新图像)的方法^[4]、基于运动矢量的方法^[5]、基于宏块类型的方法^[6]等。而关于直接在 H. 264/AVC 压缩码流中检测视频镜头边界的研究工作才刚刚起步,提出了一些新的方法。文献[7]利用帧内预测模式直方图的差异找出可能存在镜头转换的帧,再利用帧间预测模式宏块数作为特征,使用隐马尔可夫模型来确定发生转换的帧,而训练该模型需要四分之一的数据,其运算量比较大。文献[8]通过检测两个连续 I 帧的帧内宏块分割模式来检测镜头边界,并对图像划分子块来消除运动干扰,但该方法只能判断两个 I 帧之间是否发生了镜头转换,无法得到精确的转换帧。文献[9]利用帧内编码宏块比率找出候选的镜头转换帧,再用帧内预测模式 8 个预测方向的边缘直方图来具体定位转换帧。该方法对消除闪光灯干扰效果较好,但对于发生在 I 帧的镜头转换检测能力较弱。

基于以上研究,提出一种从 H. 264/AVC 压缩码流中直接提取编码信息来检测镜头边界的方法,利用宏块的帧内预测模式和帧间预测模式衡量相邻两帧间的差异,采用自适应阈值使算法适用于不同类型的视频。

3 本文检测算法

在 H. 264/AVC 中^[2],一个编码图像通常是由多个宏块(MB)构成,并以片的形式排列。I 宏块在当前片范围内进行帧内预测,P 宏块利用前面已编码的图像做前向预测,而 B 宏块既可以利用前面已编码的图像作为参考帧,也可以将未来的已编码图像作为参考做预测。

对于亮度像素,帧内预测可采用 4×4 子块或 16×16 块。 4×4 子块适用于图像细节较多的块,包括有 9 种预测模式,如图 1(a)所示的 8 个预测方向以及 DC 预测(利用当前块左边和上面的像素均值来推出当前像素值)(模式 2)。 16×16 块适用于图像较为平坦的区域,包括 4 种预测模式,如图 1(b)

所示。帧间预测按照参考方向可分为前向预测、后向预测以及双向预测,如图 2 所示。其子块分割方式较为复杂,包括 16×16 、 16×8 、 8×16 、 8×8 、 4×8 、 8×4 以及 4×4 ,每个分割的子块分别被编码传输。

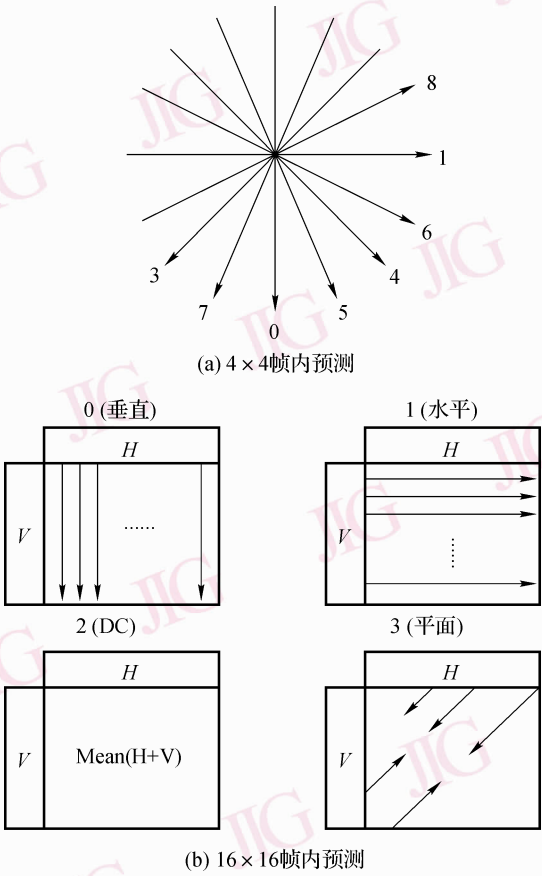


图 1 帧内预测模式

Fig. 1 Intra-prediction mode

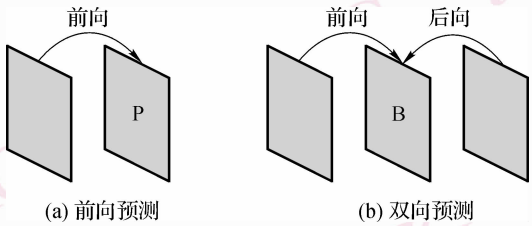


图 2 预测方向

Fig. 2 Prediction direction

对于一段视频,处于镜头内部的连续帧具有高度的时间及空间相关性,此时帧间预测编码方式被大量采用。而在镜头切换发生处,帧的内容发生了跳变,时空上的不连续性使得编码器选择帧内预测编码方式或后向预测编码方式。对一段 MTV 视频

序列相邻帧之间预测模式差异分析可知,在发生镜头切变的位置,差值明显增大,如图 3 所示。因此,可以通过考察相邻帧的预测模式差异来检测镜头的边界。

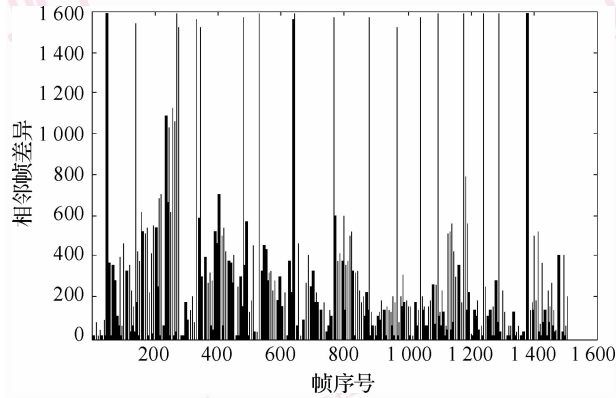


图 3 MTV 视频序列相邻帧差异

Fig. 3 The difference between frames in sequence MTV

在实际应用中,考虑到镜头切换通常不会在短时间内连续发生,可以通过检测两个相邻 I 帧之间的预测模式差异来判断该串帧序列是否发生了镜头转换。若差值小于所给的阈值,则认为这两个 I 帧同属于一个镜头,继续检测下一对 I 帧的差异。若差值大于所给的阈值,则认为该帧序列中存在转换过程,进一步检测其中 P/B 帧之间的预测模式差异,若在该 P/B 帧序列中未检出镜头转换帧,则认为当前 I 帧即为转换帧。

由于在 H. 264/AVC 亮度预测中,帧内预测采用 4×4 块或 16×16 块,而帧间预测则采用了 7 种不同的子块划分模式,因此需要对其进行归一化处理。规定以 4×4 子块作为基本块,所有 16×16 块均被看作是 16 个 4×4 子块,其预测模式保持不变,其他不同尺寸块划分做同样处理。

首先,计算当前 I 帧与前一个 I 帧之间的帧内预测模式差异:

$$D_{\text{Intra}} = \sum_{j=1}^N d_j \quad j = 1, 2, \dots, N \quad (1)$$

其中, N 为图像帧可划分的 4×4 子块的数目。设当前帧为 i 帧, $Mode_i^k$ 表示第 i 帧第 k 块的预测模式,则 d_j 表示两帧间对应子块差异:

$$d_j = \begin{cases} 0 & Mode_i^k = Mode_{i-1}^k \\ 1 & Mode_i^k \neq Mode_{i-1}^k \end{cases} \quad (2)$$

然后,设定阈值 T_{Intra} 进行检测。若 D_{Intra} 小于给定的阈值 T_{Intra} ,则不存在镜头转换,继续计算 $i + 1$

帧与 i 帧之间的帧内预测模式差异。若 D_{Intra} 大于给定的阈值 T_{Intra} ,则检测这两个 I 帧之间的 P/B 帧序列,计算相邻两帧间的预测模式差异 D_{Inter} ,其计算方法与 D_{Intra} 类似。

需要注意的是,此时的 d_j 统计的是帧内预测模式以及后向预测块差异,即若当前帧当前块为前向预测块,则跳过该块不处理。这是由于若当前块采用了前向预测,则表明该块与前一帧具有一定的镜头相关性。得到的 D_{Inter} 若大于阈值 T_{Inter} ,则判断当前帧为镜头转换帧,若 D_{Inter} 小于阈值 T_{Inter} ,则认为在 I 帧处发生了镜头转换。

在镜头边界检测中,阈值设定的好坏直接影响检测的准确性,如何确定合适的阈值是一个值得重视的问题。显然,对于一段内容复杂的长视频,固定阈值往往达不到满意的检测效果,鉴于此,引入自适应阈值来解决这一问题。文献[10]中直接采用差异值的均值与方差作为阈值。经实验检测,该阈值对于不同的视频类型其检测效果波动较大。因此,在这里对其加以改进,取得了较好效果。

$$\begin{cases} m_l = \frac{1}{L} \sum_{l=1}^L D \\ \sigma_l = \sqrt{\frac{1}{L-1} \sum_{l=1}^L (D - m_l)^2} \end{cases} \quad (3)$$

其中, L 为窗口长度, m_l, σ_l 分别为窗口内两帧间模式差异 D 的均值与标准差,二者的值表明了窗口内帧模式差异的均值与波动程度。利用二者来定义自适应阈值

$$T = a \cdot m_l + b \cdot \sigma_l \quad (4)$$

其中, a, b 为常数。实验结果表明, a, b 二值的大小对检测效果影响极大。对于不同类型的序列,通过反复实验,最终将它们确定为 0.98 和 -1.20。

4 实验结果及分析

为了验证算法的检测效果,选取 3 组不同类型的视频序列进行实验,分别为影片《赤壁》片段 Movie(1 500 帧),足球比赛片段 Football(1 800 帧)以及 Backstreet Boy 的 MTV 片段(1 600 帧)。采用 JM9.5 模型进行编码测试,帧大小为 176×144 ,帧率为 30 fps,支持 B 帧帧间编码,每隔 60 帧插入一个 I 帧, B, P 帧交替出现,帧序列结构为 IPBPB……。

在镜头边界检测中,通用的评价方法是采用查全率和查准率,定义如下:

$$R = \frac{N_c}{N_c + N_m} \tag{5}$$

$$P = \frac{N_c}{N_c + N_f} \tag{6}$$

其中, N_c 为正确检出数, N_m 为漏检数, N_f 为错检数。检测结果如表 1 所示,其中 Cut 表示视频序列中镜头切变数。

表 1 镜头边界检测结果
Tab. 1 The result of shot boundary detection

序列	Cut	N_c	N_m	N_f	$R(\%)$	$P(\%)$
Movie	13	13	0	0	100	100
Football	16	15	1	1	94	94
MTV	16	16	0	2	100	89

在 Movie 序列中,展现了战争前敌对双方战斗前准备的场面,叙事风格相对和缓,镜头切换缓慢。因此,算法表现出色,取得了 100% 的查全率和查准率。在 Football 序列中,包括了远景与近景切换、运动员高速奔跑等画面,算法也取得了不错的效果。发生错检是由于快速回放导致的镜头高速移动。在 MTV 序列中,运用了许多特效,闪光的出现导致了错检的发生。总体而言,算法对于不同的视频类型均取得了令人满意的检测效果。

为了进一步验证算法的检测效果,针对不同类型的视频序列,与现有的几种算法进行了比较实验。表 2 给出本文算法与文献[10]算法的性能比较。

表 2 与现有算法比较
Tab. 2 Compared with the existing algorithm

视频类型	文献[10]算法		本文算法	
	$R(\%)$	$P(\%)$	$R(\%)$	$P(\%)$
Sport	90	96	94	94
Novel	93	99	100	100

由表 2 可以看出,对于不同类型的视频序列,本文算法在总体性能上明显优于文献[10]算法。

5 结 论

在 H. 264/AVC 压缩域直接进行视频镜头边界检测的方法,利用宏块的帧内预测模式差异和帧间预测方向,首先对相邻两个 I 帧做检测以考察是否存在镜头转换,再进一步精确定位镜头转换帧,简便高效。此外,自适应阈值的应用大大提高了算法的通用性。

参考文献 (References)

1 JVT-G050. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification[S].

2 Wiegand T, Sullivan G J, Bjontegaard G, *et al.* Overview of the H. 264/AVC video coding standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 560-576.

3 Arman F, Hsu A, Chiu M Y. Image processing on compressed data for large video databases [A]. In: Proceedings of the 1st ACM International Conference on Multimedia[C], Anaheim, California, USA, 1993: 267-272.

4 Yeo B L, Liu B. Rapid scene analysis on compressed video[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1995, 5(6): 533-544.

5 Zhang H J, Chien Y L, Smoliar S W. Video parsing and browsing using compressed data[J]. Multimedia Tools and Applications, 1995, 1(1): 89-111.

6 Pei S C, Chou Y Z. Efficient MPEG compressed video analysis using macroblock type information[J]. IEEE Transactions on Multimedia, 1999, 1(4): 321-333.

7 Liu Y, Wang W Q, Gao W, *et al.* A novel compressed domain shot segmentation algorithm on H. 264/AVC [A]. In: Proceedings International Conference on Image Processing[C], Singapore, 2004, 4: 2235-2238.

8 Kim S M, Byun J W, Won C S. A scene change detection in H. 264/AVC compression domain[A]. In: Proceedings of the 6th Pacific Rim Conference on Multimedia Advances in Multimedia Information Processing-PCM 2005, LNCS[C], Jeju Island, Korea, 2005, 3768: 1072-1082.

9 Hong B, Eom M Y, Choe Y S. Scene change detection using edge direction based on intra prediction mode in H. 264/AVC compression domain [A]. In: Proceedings of the IEEE Region 10 Annual International Conference[C], Hong Kong, China, 2006: 1-4.

10 Sarah D B, Davy V D, Jan D C, *et al.* A compressed-domain approach for shot boundary detection on H. 264/AVC bit streams[J]. Signal Processing: Image Communication, 2008, 23(7): 473-489.