

基于 H.264 压缩域内宏块预测模式的镜头突变检测

郭俊超, 孙树森, 桂江生, 张华熊

(浙江理工大学信息学院, 杭州 310018)

摘要: 根据 H.264 的宏块编码特性,提出了一种直接在 H.264 压缩域内进行镜头突变检测的算法。对于 I 帧,通过对帧内所有宏块的预测模式建立一个矩阵,最后与当前 I 帧相对应位置宏块的帧内预测模式进行比较,检测是否发生突变;对于 P 帧和 B 帧,通过分析帧内宏块的预测模式及相邻帧之间的相关性来进行检测。实验结果表明,该算法具有较高的查全率和查准率。

关键词: H.264; 镜头突变; 预测模式; 矩阵

中图分类号: TP391 **文献标志码:** A

0 引言

进入 21 世纪,随着互联网技术和视频获取技术的飞速发展,互联网以及电子设备上的视频内容正以惊人的速度增长,面对如此海量的视频数据,如何对其进行有效的管理与利用,已日益成为一个非常重要的课题。以往的基于文本信息的检索以及用户定位搜索,已远远不能满足人们有效利用视频的需求,这就使基于内容的视频检索变得越来越迫切。视频边界检测是基于内容的视频检索的基础。视频边界检测算法可分为基于像素域和基于压缩域两大类,基于像素域的视频边界检测首先将压缩的视频流完全解压,然后在像素域内进行分析,其解压通常需要消耗大量的时间和计算资源^[1],而基于压缩域的视频边界检测则不需要完全解压,直接在压缩域内进行。现有的视频边界检测算法大多针对 H.261、H.263、MPEG-2 和 MPEG-4 等视频编码标准。作为当前技术水平最高的视频编码格式,H.264 标准引入了一系列先进的视频编码技术,并得到了越来越多的推广和使用,而基于 H.264 压缩域内的视频边界检测与特征提取也受到越来越多研究者的关注和研究^[2]。

本文在文献[3-4]的基础上,通过对 H.264 码流中的 I 帧、P 帧和 B 帧的不同宏块预测模式以及相邻帧之间所表现出来的相关性等特性进行分析,提出了一种新的基于压缩域的视频镜头突变检测算法。

1 H.264 镜头突变情况分析

在 H.264 码流中,有 I 帧、P 帧和 B 帧 3 种类型的帧,在基本框架内,支持 I 帧和 P 帧编码;在扩展框架、主要框架以及高级框架内,支持 I 帧、P 帧和 B 帧编码,其中 B 帧穿插在 I 帧和 P 帧之间。在理论上,码流中的任何一帧都有可能发生突变,但实际上,两个突变帧之间总会存在一定的时间跨度。图 1 通过最常见的 IBBPBBPBIBBP 编码结构给出码流中可能存在的 5 种突变情况。

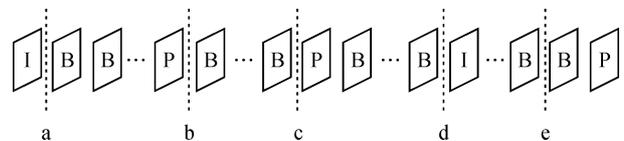


图 1 常见的 5 种镜头突变编码示意

在有关视频边界检测方面,相对于其它视频编码标准,H.264 最重要的改变就是引入了帧内预测

收稿日期: 2012-06-04

基金项目: 国家自然科学基金(61105035),浙江省自然科学基金(Y1100656)

作者简介: 郭俊超(1986-),河南商丘人,硕士研究生,主要从事视频编码、数据挖掘方面的研究。

通信作者: 孙树森,电子邮箱: shusensun@zstu.edu.cn

技术和多帧运动补偿预测技术。所谓帧内预测技术,是指帧内的宏块通过参考空间上相邻并且已经解码过的宏块上的像素来进行编码^[5]。帧内预测主要有3种模式,即 Intra_4×4 预测模式、Intra_16×16 预测模式和 I_PCM 模式。Intra_4×4 预测模式是对宏块中的每一个 4×4 亮度块分别进行帧内编码,该预测模式多用于纹理结构比较复杂的地方。Intra_16×16 模式是对一个 16×16 亮度块来进行预测编码,该预测模式则多用于纹理结构比较简单的地方。除此之外,H.264 还引入了 I_PCM 模式,即不参考其它宏块,直接对帧内宏块进行编码^[6]。所谓多帧运动补偿技术是针对 P 帧、B 帧而言,该技术可以让编码器在已经解码过并且保存在缓冲区的 N 个帧图像中,选择最优的参考图像。其中,P 帧最多可以前向参考 16 帧图像,B 帧不仅可以进行前向参考、后向参考,而且还可以同时进行双向参考,进行单向参考时,最多可以参考 16 帧图像,进行双向参考时,最多可以参考 32 帧图像^[7]。

2 基于宏块类型的突变检测算法

在 H.264 编码过程中,如果相邻帧在同一个镜头内,则两帧图像将会表现出很强的相关性。如果发生镜头突变,即由一个镜头的最后一帧直接切换到下一个镜头的第一帧,由于突变边界的两帧分别属于两个镜头,两帧图像内容的相似性非常低,两帧之间的相关性就非常弱。

2.1 码流中 I 帧突变的检测

在 H.264 编码中,I 帧的宏块均采用帧内预测模式,据此,从码流中的第 1 个 I 帧——IDR 帧开始,用矩阵 \mathbf{M} 记录帧内所有宏块的帧内预测模式,令:

$$\mathbf{M} = \begin{pmatrix} MB_{11} & MB_{12} & \cdots & MB_{1m} \\ MB_{21} & MB_{22} & \cdots & MB_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ MB_{n1} & MB_{n2} & \cdots & MB_{nm} \end{pmatrix}$$

其中, MB_{ij} 代表 I 帧中第 i 行第 j 列宏块的帧内预测模式。然后对位于两个 I 帧之间的 P 帧和 B 帧的每一个宏块的预测模式进行检测,如果是帧内预测模式,则让该宏块的预测模式取代矩阵 \mathbf{M} 相应位置上的预测模式;当遇到下一个 I 帧时,我们将此 I 帧的所有宏块预测模式组成的矩阵 \mathbf{S} 与矩阵 \mathbf{M} 进行比较,并统计两矩阵相同位置上帧内预测模式不同的总数。设 MB_{ij}^M, MB_{ij}^S , 分别为矩阵 \mathbf{M} 和矩阵 \mathbf{S} 中第 i 行第 j 列的宏块的帧内预测模式,若 $MB_{ij}^M = MB_{ij}^S$, 记 $|MB_{ij}^M - MB_{ij}^S| = 0$; 若 $MB_{ij}^M \neq MB_{ij}^S$, 记

$|MB_{ij}^M - MB_{ij}^S| = 1$; 则当前 I 帧与矩阵 \mathbf{M} 之间差异为:

$$MB_{(M,D)} = \frac{\sum_{i=1}^n \sum_{j=1}^m |MB_{ij}^M - MB_{ij}^S|}{N} \quad (1)$$

其中 N 为一帧内总的编码宏块数。此外,如果当前 I 帧的前一帧是 B 帧且其后向参考宏块或者双向参考宏块所占总宏块的比例较高时,则说明当前 I 帧与前一帧在内容上存在很大的相似性,当前 I 帧一般不会发生突变。

设当前帧为 F_i , 其前一帧记为 F_{i-1} , 把一帧内所有的宏块数记为 N , 如果同时满足以下两式:

$$\frac{\varphi(i-1) + \beta(i-1)}{N} < T_1 \quad (2)$$

$$D_{(M,D)} = \frac{\sum_{i=1}^n \sum_{j=1}^m |MB_{ij}^M - MB_{ij}^S|}{N} > T_2 \quad (3)$$

则认为当前 I 帧发生了突变,其中 $\varphi(i-1)$ 为前一帧 F_{i-1} 采用双向参考编码的宏块数, $\beta(i-1)$ 为 I 帧前面的 B 帧采用后向参考编码的宏块数。通过本文第 4 节的实验统计, T_1 与 T_2 分别定为 0.08 和 0.35。

2.2 码流中 P 帧和 B 帧的突变检测

相比较 I 帧只能进行帧内编码, P 帧和 B 帧内的宏块不仅可以进行帧内编码,还可以充分利用相邻帧之间的相关性来进行帧间编码。如果当前帧是 P 帧并且发生了突变,则当前 P 帧与先前已编码帧之间的内容相关性非常低,所以很少采用前向参考来进行帧间编码,而是将大量采用帧内编码,且当前 P 帧的相邻帧也很少采用后向参考编码模式^[8]。

如果当前帧是 B 帧且发生了突变,因该帧图像与其前一帧之间的内容相关性很低,当前帧很少采用前向参考。此时,对于当前 B 帧而言,如果其后面的帧图像已经编码并可以参考,则当前 B 帧内的宏块将会大量采用后向参考来进行编码,否则,当前 B 帧内的宏块会大量采用帧内编码^[9]。令当前帧记为 F_i , 其前一帧记为 F_{i-1} , N 为一帧内所有的宏块数,如果满足以下两式:

$$\frac{\delta(i-1) + \beta(i-1)}{N} < T_1 \quad (4)$$

$$\frac{\varphi(i) + \beta(i)}{N} > T_3 \quad (5)$$

即认为,当前帧发生了突变,其中 $\varphi(i)$ 为当前帧采用帧内编码的宏块数, $\beta(i)$ 为当前帧采用后向参考编码的宏块数, $\beta(i-1)$ 为帧 F_{i-1} 采用后向参考编码的宏块数, $\delta(i-1)$ 为帧 F_{i-1} 采用双向参考编码的宏块数。通过不同视频类型的实验统计, T_1, T_3 分别

定为 0.08 和 0.94 时,查全率 R 和查准率 P 均较为理想。

3 实验验证与分析

为了验证本文算法的检测效果,从视频库中选取三段不同类型的视频序列来进行编码,视频序列 MV 是李贞贤《阿里阿里》里的一个片段(5 000 帧),视频序列 Football 是《天下足球》里的一个片段(5 000 帧),视频序列 Movie 是电影《Big Bunny》里的

一个片段(5 000 帧)。用 JM8.6main profile 规格对上述 3 个视频序列进行编码,视频图像格式为 CIF(352 × 288),main profile 支持 I 帧、P 帧和 B 帧编码,采用常见的 IBBPBBPBBIBBP...模式进行编码,图 2 给出了从《Big Bunny》的视频流中连续提取的 500 帧,每一个 I 帧的 T_2 值以及其相邻的前一帧的 T_1 值, T_1 位于 T_2 的左边。

图 3 和图 4 分别给出了视频流中 B 帧和 P 帧的 T_3 值和 T_1 值。

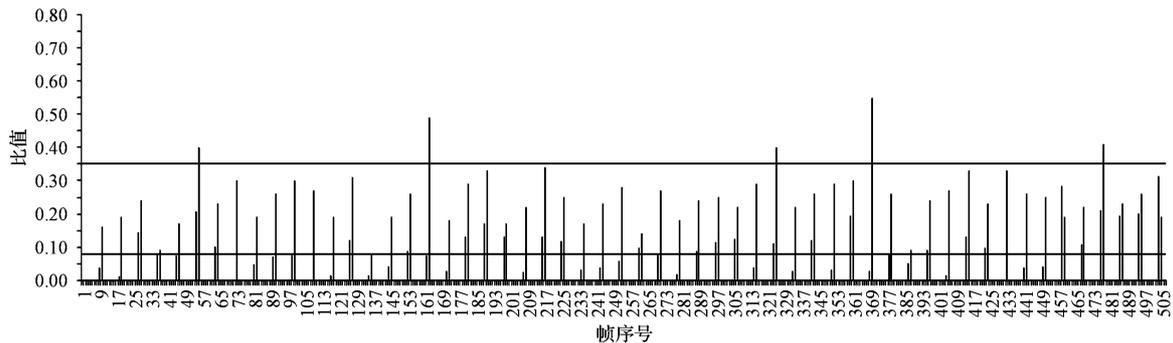


图 2 I 帧的 T_1 、 T_2 统计信息

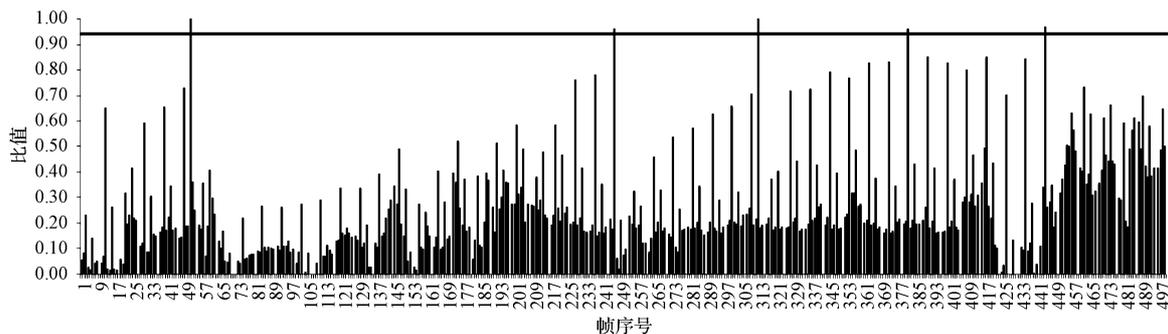


图 3 P 帧、B 帧的 T_3 统计信息

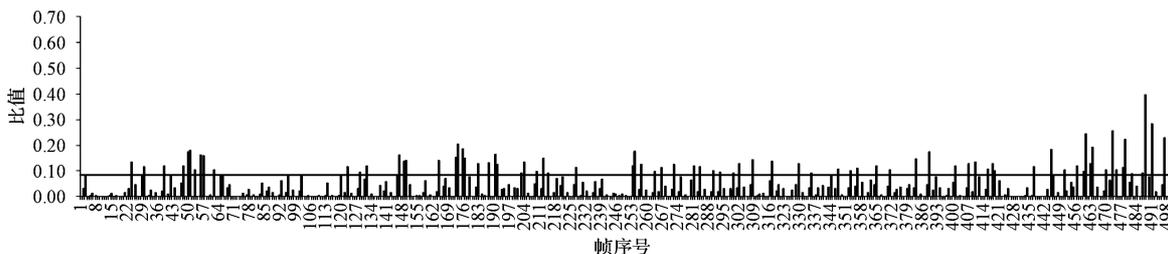


图 4 P 帧、B 帧的 T_1 统计信息

4 实验结果分析

在镜头边界检测算法评估中,通常采用查全率 R (Recall)和查准率 P (Precision)来进行评估和分析。其定义如下:

$$R = N_c / (N_c + N_m)$$

$$P = N_c / (N_c + N_f)$$

其中 N_c 为正确检出数, N_m 为漏检数, N_f 为误检数。图 5 和图 6 用三维图形分别展示了视频流《Big Bunny》中 T_1 、 T_2 和 T_3 的变化对查全率 R 和查准率 P 的影响。通过对不同视频类型的实验结果统计, T_1 、 T_2 、 T_3 分别定为 0.08、0.35、0.94 时,查全率 R 和查准率 P 均较为理想。

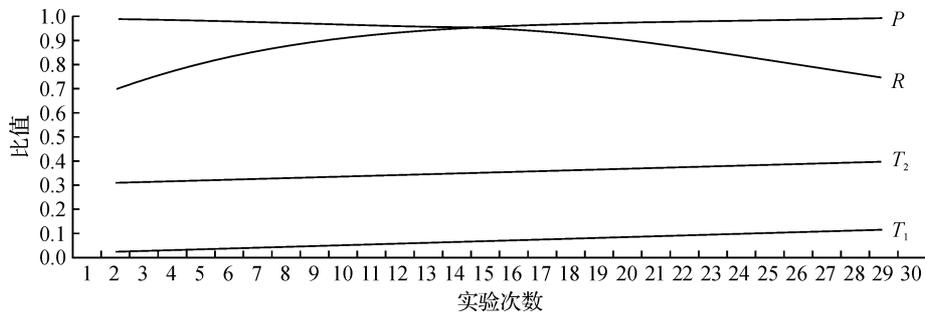
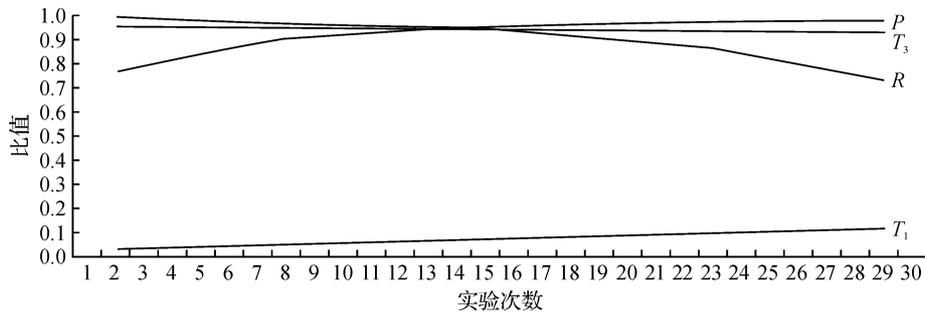
图5 T_1, T_2 对查全率 R 和查准率 P 的影响图6 T_1, T_3 对查全率 R 和查准率 P 的影响

表1给出了上述三段视频突变检测结果。

表1 三个视频序列的实验结果

视频	总帧数/帧	突变帧数/帧	N_c	N_f	N_m	查准率 R /%	查准率 P /%
Movie	5 000	23	22	1	1	95.6	95.6
Football	5 000	39	38	3	2	92.7	95.0
MV	5 000	27	26	2	1	96.3	92.9

由表1实验结果可知:在Movie序列中,由于叙事相对缓慢,镜头变换较为缓慢,算法表现较为出色;在Football序列中,由于近景远景切换频繁,再加上运动员奔跑速度变化很快,算法效果略微有所下降;在MV序列中,由于闪光灯的频繁使用,导致了误检率的上升。总体而言,算法对于不同类型的视频序列均能取得较好的效果。

为进一步验证本文算法的检测效果,表2给出了本文算法与文献[4]算法对于不同视频类型的边界检测性能比较。

表2 本文算法与文献[4]算法的比结果

视频类型	文献[4]算法		本文算法	
	查准率 R /%	查准率 P /%	查准率 R /%	查准率 P /%
Movie	94.3	97	95.6	95.6
Football	93.5	93.9	92.7	95.0
MV	95.2	92.1	96.3	92.9

从表2可以看出,对于不同类型的视频序列,本文算法在总性能上明显优于文献[4]的算法。

5 结 语

在基于内容的视频检索中,视频镜头边界检测是最重要的基础,也是评估一个视频检索系统性能的重要指标。本文主要利用了H.264在编码过程中,宏块的编码模式以及当前帧与相邻帧之间的相关性来进行镜头突变检测,并为视频的渐变检测打下了坚实的基础。实验结果表明,本文所提出的算法具有很高的查全率和查准率。

参考文献:

- [1] 冯杰. 基于H.264压缩域的视频分割与特征提取方法研究[D]. 杭州: 浙江大学, 2009.
- [2] 百度百科. H264[EB/OL]. [2012-08-18]. <http://baike.baidu.com/view/403562.html>.
- [3] Sarah D B, Davy V D, Jan D C. A compressed-domain approach for shot boundary detection on H.264/AVC bit streams[J]. Signal Processing: Image Communication, 2008, 23(7): 473-489.
- [4] Bruyne S D, Neve W D. Temporal video segmentation on H.264/AVC compressed bitstreams[J]. Advance in Multimedia Modeling, 2005, 4: 351-359.
- [5] Wiegand T, Sullivan G J, Bjntegaard G, et al. Overview of the H.264/AVC video coding standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 3(7): 559-566.
- [6] Wei Zeng, Wen Gao. Shot change detection on H.264/AVC compressed video[J]. IEEE Int Symp, 2005, 4:

- 3459-3462.
- [7] Iain E R. The H. 264 Advanced Video Compression Standard[M]. 2nd Ed. Hoboken: John Wiley & Sons, 2010.
- [8] Kim S M, Byune J, Won C. A scene change detection in H. 264/AVC compression domain[C]//Proceedings of Pacific-Rim Conference on Multimedia, 2005: 1072-1082.
- [9] Schwarz H, Marpe D, Wiegand T. Overview of the scalable video coding extension of the H. 264/AVC standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(9): 1104-1110.

Abrupt Shot Transition Detection Based on Prediction Mode of Macro Block within H. 264 Compressed Domain

GUO Jun-chao, SUN Shu-sen, GUI Jiang-sheng, ZHANG Hua-xiong

(School of Informatics, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: This paper puts forward an algorithm of abrupt shot transition detection directly within H. 264 compressed domain according to macro blocking coding features of H. 264; for frame I, establishes a matrix for the prediction mode of all intra-frame macro blocks, finally makes a comparison with the intra-frame prediction mode of macro blocks in the corresponding position of the current frame I and detects whether abrupt transition occurs; and, for frames P and B, conducts a test on the prediction mode of intra-frame macro blocks and relativity of adjacent frames. The experimental result shows that this algorithm has a high recall ratio and precision ratio.

Key words: H. 264; abrupt shot transition; prediction mode; matrix

(责任编辑: 陈和榜)