

一种压缩域中的体育视频慢镜头探测方法

侯绿林 白亮 老松杨

(国防科技大学多媒体研发中心 长沙 410073)

摘要 慢镜头探测是基于内容视频分析的一项重要技术。在介绍视频压缩的基本理论及 MPEG-1 视频压缩标准的基础上,对压缩域内的慢镜头探测进行了研究,提出了一种两阶段压缩域下的慢镜头探测方法。该探测方法的第一步是基于宏块类型信息的镜头探测,在镜头探测的基础上定义和计算镜头密度,探测可能的慢镜头出现位置;第二步是计算压缩域下连续帧间差,从候选慢镜头中探测慢镜头。实验结果显示,相对于已有的方法,该方法具有较快的探测速度和较高的准确率。

关键词 慢镜头探测,镜头切割,基于内容视频检索,MPEG 压缩域,宏块类型

Method for Slow-motion Replay Detection on Compressed Domain in Sports Video

HOU Lu-lin BAI Liang LAO Song-yang

(Multimedia Research and Development Center, National University of Defense Technology, Changsha 410073, China)

Abstract The Slow-motion replay (SMR) detection is an important technique for content-based video analysis. This paper introduced the basic theory of video compression and MPEG-1 video compression standard. A two-step method for SMR detection on MPEG uncompressed domain was proposed, in which firstly shot density was defined and calculated based on shot detection results using MPGE macroblock types information to detect the occurrences of possible SMR shots, and secondly the difference between consecutive frames was measured to detect true SMR in candidate SMR shots. Experiments results show that unlike previous approaches, our method has great improvement in both speed and accuracy.

Keywords Slow-motion replay detection, Shot cut, Content-based video analysis, MPEG compressed domain, Macroblock type

21 世纪是信息时代,随着因特网技术和多媒体业务的发展,人们对视频数据的存储、检索、加工等应用提出了更新更高的要求。视频实际上是集图像序列、图像、文字等为为一体的、被人们广泛使用的一种综合性媒体,视频语义内容分析已成为当前研究的热点。体育视频作为一种深受人们喜爱、具有广大群众基础的视频类型,其语义内容分析技术具有广阔的应用前景。体育视频中的慢镜头是体育视频的重要语义内容。进球、精彩射门镜头、犯规等都是观众感兴趣的语义内容,导演常常以慢镜头的方法重播以上镜头。利用低层特征来获取慢镜头,不仅可以标定比赛中精彩片断的位置,用来生成比赛的精彩集锦,还可以辅助体育视频的内容分析。

1 慢镜头探测技术研究现状

慢镜头探测是视频处理领域中的一个难点,这是由于在不同的视频中,慢镜头的产生方法可能不同,编辑手法也因人而异,并且慢镜头也没有一个很明显的边界,因此很难找出一种通用有效的慢镜头探测算法。

国内外对慢镜头探测技术的研究并不是很多,主要有以下方法用于对慢镜头的探测:美国夏普实验室 H. Pan 等人运

用隐马尔可夫模型(HMM)进行慢镜头探测,这种方法对于由帧重复产生的慢镜头具有很强的通用性,但对于不同类型的视频需要重新建立 HMM;日本 Osaka 大学的 Noboru^[3]等人针对美式足球比赛中扫描编辑效果进行探测,这种方法不仅可以探测扫描,还可以扩展到其他的编辑效果,但是依赖于体育视频的编辑模式,准确度不够理想;马里兰大学的 V. Kobla 等人利用 MPEG 流中的宏块信息进行探测,根据慢镜头由帧重复产生这一原理,对帧重复时宏块信息的特点做了分析,并根据这一特点来判断视频中是否存在慢镜头,这种方法计算简单,但是精确度不高,通用性和边界定位较差。

综上所述,慢镜头探测存在的主要问题是算法的通用性不强,依赖于视频的编辑模式,并且慢镜头边界定位的准确率有待进一步提高。本文根据这一现状,在实验分析研究的基础上,提出了基于压缩域下的宏块类型和帧间差的方法来实现慢镜头的探测。

2 基于镜头密度和帧间差的慢镜头探测

考虑到数字视频的数据规模,绝大多数视频数据以压缩形式进行存储和传送。且随着基于内容需求的增加,解压缩

到稿日期:2008-10-31 返修日期:2009-02-25 本文受国家自然科学基金(60572137),国家高技术研究发展计划(863 计划)(2006AA01Z316),教育部博士点基金资助。

侯绿林(1987-),男,硕士生,主要研究方向为数字视频、音频处理和检索等,E-mail:houlvlin@126.com;白亮(1978-),男,博士生,主要研究方向为数字视频、音频处理和检索等;老松杨(1968-),男,教授,博士生导师,主要研究方向为数字视频、音频处理和检索等。

所需的额外费用不断增加。而在压缩域内进行检索,可以不需要解码或只需要部分解码,就可以避免解压缩的开销,提高处理的速度。实验分析了体育视频中慢镜头的一般特征,根据慢镜头出现时的特点检测慢镜头,有效地增强了算法的通用性。准确的镜头边界定位则解决了慢镜头边界探测问题。

该慢镜头探测算法包括3个过程。首先是将视频序列切分为镜头,然后利用镜头密度探测候选慢镜头区域,接着在候选慢镜头区域进行帧间差的比较,筛选出虚假的区域,得到慢镜头,如图1所示。

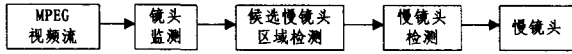


图1 慢镜头检测过程

2.1 镜头检测

由于一个镜头只能拍摄相邻地点连续发生的事情,描述能力有限,因此大多数视频都是由许多镜头通过编辑连接而成的。镜头的切换分为突变和渐变(abrupt change and gradual change)两类。突变是一个镜头直接转换为下一个镜头,中间没有时间上的延迟;渐变则是加入了一些空间或时间上的编辑效果,由前一个镜头慢慢地转换为下一个镜头。

2.1.1 突变镜头检测

将镜头突变分为3种类型:镜头突变发生在P帧或I帧处(SCPI);镜头突变发生在前一个B帧处(SCFB);镜头突变发生在后一个B帧处(SCRB)。这个分类包括了所有突变发生的情况,因为前一个SGOP的最后一帧是后一个SGOP的第一帧,因此只需要考虑后3帧。每种类型的突变都会产生一个特定的宏块类型模式。

通过分析发现,虽然不同类型突变中B帧前向和后向宏块的类型不同,但是都会导致B帧中双向预测类型宏块FB数量明显减少,故FB宏块数量是镜头切换的一个显著特征,如图3所示。

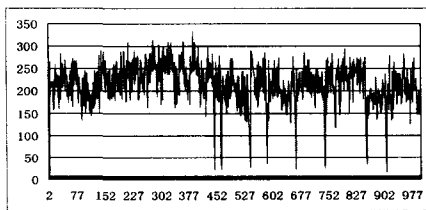


图2 B帧中FB宏块数

在检测镜头时,阈值选择是个很重要的问题。如果不能合理地确定阈值,算法也将失去价值。在不同类型的视频中,突变发生位置SGOP中B帧FB宏块虽都明显减少,但不同类型视频中宏块总数量不同,不可能存在一个通用的阈值。因此,选择连续的3个B帧宏块数比值进行比较,消除不同类型视频中宏块总数不同的影响。

定义当前B帧中FB宏块数为 n_{FB_1} ,前一个B帧中FB宏块数为 n_{FB_0} ,后一个B帧中FB宏块数为 n_{FB_2} ,阈值参数为T,若满足下面公式之一:

$$n_{FB_1} \leq n_{FB_0} / T \quad (1)$$

$$n_{FB_2} \leq n_{FB_0} / T \quad (2)$$

就认为在此SGOP中发生了镜头切换。

2.1.2 渐变镜头探测

渐变是许多镜头切换方式的总称。其特点是整个切换过

程是逐渐完成的,从一个镜头变化到另一个镜头常可能延续十几帧或几十帧。与切变只有一种不同,渐变有许多种,如淡入、淡出、擦出、滑动等。相对来说,突变检测算法比较成熟,可以达到较高的精度,而渐变检测方法还不完善。本文主要是基于宏块类型的方法来探测镜头的渐变。

镜头渐变是一个持续时间相对较长的过程,镜头渐变的探测主要是对起始帧和结束帧的探测。由于视频内容的变化,起始帧和结束帧中P帧I类型宏块数量显著增加。定义P帧I类型宏块数量为 n_I ,阈值为 T_I ,当满足式(3)时,

$$n_I \geq T_I \quad (3)$$

就判定镜头发生渐变。

该方法探测速度快,且可以避免其他一些方法中由于运动物体引起亮度变化而导致假算法,但是无法区分出快速的摇镜头(Fast Panning),因此需要进一步筛选。相对于摇镜头,渐变过程中B帧存在大量的FB宏块。定义B帧FB宏块数量为 n_{FB} ,阈值为 T_{FB} ,当满足式(4)时,

$$n_{FB} \geq T_{FB} \quad (4)$$

便可筛选掉多数的摇镜头,从而提高算法精度。

2.2 基于镜头密度的候选慢镜头区域探测

体育视频中的镜头主要分为3类:远景、中景、特写。远景多表现场上的形势,近景多表现对象特写,而中景是介于两者之间的体育视频视图,如图3所示。慢镜头是视频中的一类特殊镜头。慢镜头中的镜头持续时间短,一般在持续时间较短的中景或特写镜头后面出现。在慢镜头出现的地方镜头切换频繁,如图4所示。



图3 正常的镜头切换

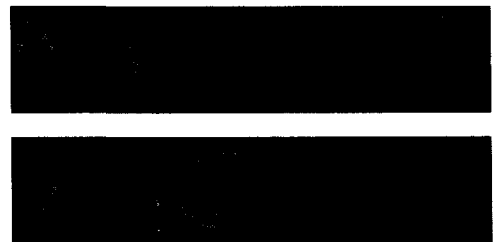


图4 慢镜头出现时的镜头切换

定义每300帧为一个时间段T,一个T内存在的镜头切换边界数为 T_i ,判定公式如下:

$$T_i \geq 4 \quad (5)$$

由于一般慢镜头中还存在镜头切换,且单个镜头持续时间不超过10s,故为了探测到整个慢镜头部分,将检测到的T向后扩展300帧,得到新的T,即为候选慢镜头区域。以2006年足球世界杯视频为例,在长度为10000帧的视频中进行分析,如图5所示。丢失是个别慢镜头为远景镜头回放,持续时间长,但这类慢镜头数量很少;造成误判则是因为镜头在远景和中景间切换频繁,例如摄像者既想拍摄全局又想清楚显示

足球的运动位置轨迹,就需要不停地切换镜头。误判的存在只会影响慢镜头算法的效率,但不会影响慢镜头探测的准确性。

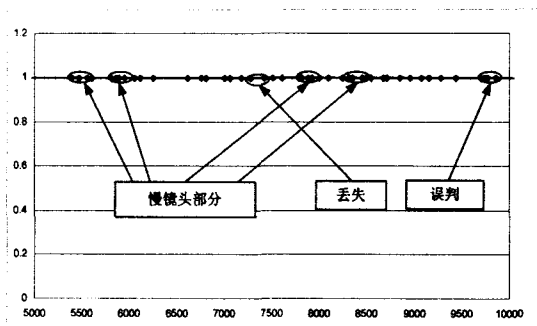


图5 镜头密度

2.3 基于帧间差的慢镜头探测

该方法主要是针对通过插入重复帧产生的慢镜头。这类慢镜头由于成本低廉,在体育视频中得到了广泛的应用。定义帧间差为 $D(t)$,有

$$D(t) = \frac{1}{M * N} \sum_{p=1}^M \sum_{q=1}^N (I_t(p, q) - I_{t-1}(p, q))^2 \quad (6)$$

式中, $I_t(p, q)$ 表示在时刻 t 帧中坐标为 (p, q) 的宏块的平均亮度值(DC), M 和 N 分别代表帧的宽度和高度。

慢镜头是由重复帧产生的,因此每过若干帧, $D(t)$ 就会出现一个很小的值,相邻帧之间的平均亮度差值就会很小。但实验显示, $D(t)$ 值的大幅度振荡不仅与慢镜头的帧重复有关,还受摄像机运动的影响。摄像机运动越大, $D(t)$ 的振荡幅度也越大,为了补偿摄像机运动对 $D(t)$ 的影响,将窗口中每个 $D(t)$ 的值减去该窗口所有 $D(t)$ 值的均值 $\bar{D}(t)$; 另一方面是为以下多门限过零率的计算提供便利。

首先计算在 S 长的滑窗内的多门限过零率。 $Z_c(t, \theta_k)$ 是指 $D(t)$ 在长度为 S 的窗口内超过门限值穿过零点的次数,可由式(7)计算,其中 θ_k 是门限值。

$$Z_c(t, \theta_k) = \sum_{i=1}^{S-1} f(D(t-i) - \bar{D}(t), D(t-i-1) - \bar{D}(t), \theta_k) \quad (7)$$

$$f(x, y, \theta) = \begin{cases} 1, & \text{if } (x \geq \theta \text{ and } y \leq -\theta) \text{ or } (x \leq -\theta \text{ and } y \geq \theta) \\ 0, & \text{else} \end{cases} \quad (8)$$

门限过零率反映了信号的频率和幅度。易见,当门限值 θ_k 很大时,慢镜头的门限过零率会远远大于正常播放速度的镜头。为了综合度量 $D(t)$ 振荡的频率和幅度,定义了 $P_x(t)$,其含义指零点穿越次数大于一定阈值后量化等级 θ_k 的最大脚标,由式(9)计算:

$$P_x(t) = \begin{cases} \arg \max_k \{ \theta_k | Z_c(t, \theta_k) > \beta \}, & \text{if } (Z_c(t, \theta_k) > \beta) \\ 0, & \text{else} \end{cases} \quad (9)$$

当 $P_x(t)$ 大于一定阈值时,就可将慢镜头与正常播放速度的镜头区别开来。首先 $D(t)$ 值减去均值 $\bar{D}(t)$,只是在某种程度上补偿了镜头运动的影响。例如,远景镜头中摄像机的运动和近景镜头中运动员的运动都可以造成 $D(t)$ 大幅度振荡,故根据平均运动大小,自适应地调整量化级 θ_k 大小,即 θ_k 的大小与 $D(t)$ 在整个镜头中的平均值成正比。量化步长的范围从 1 到 4。运动量低的镜头,量化步长取 1,而运动量大

的镜头,量化步长则取 4。当最大脚标 k 大于 100 时,则判定此镜头为慢镜头。慢镜头过程中 $D(t)$ 值可能不是零值,而是一些很小的值。由于选择的窗口通常不应包含太多只有普通运动的帧,因此 S 取较大的值通常效果不好,所以选择 L 为 7。阈值 β 的选取较为容易,可以选取 $\beta=1$,只要证明存在零点穿越即可。

3 试验与分析

选取 2006 年足球世界杯 3 个片段、NBA 篮球比赛两个片段、2007 年橄榄球世界杯两个片段作为实验数据进行试验,结果如表 1 所列。试验结果表明,该算法能够较好地检测到视频中的慢镜头,取得了令人满意的结果。影响查全率的原因一方面是在候选慢镜头区域探测时个别慢镜头区域未检测到,影响后面的检测效果,可进行人工修正;另一方面一些慢镜头零相交特征不明显。影响准确率最主要的原因是在远景镜头中相机突然或快速的运动和近景镜头中物体不规则的运动。

表1 慢镜头探测试验结果

视频类型	视频帧数	实际个数	正确个数	错误个数	查全率	准确率
足球 1	30021	12	10	0	83.3%	100%
足球 2	26703	11	10	1	90.9%	90.9%
足球 3	47066	17	15	1	88.2%	94.1%
篮球 1	34533	13	11	1	84.6%	91.7%
篮球 2	52004	19	16	2	84.2%	88.9%
橄榄球 1	14791	10	9	0	90%	100%
橄榄球 2	29546	17	15	1	88.2%	94.1%

对于表 1 中试验的视频,本文用单纯的基于帧间差的慢镜头探测算法进行了试验对比。现以测试视频足球 1、足球 2 和足球 3 为例,采用基于帧间差探测算法获得的平均查全率为 86%、平均准确率为 92%。可见本文提出的方法与基于帧间差的方法在探测性能上相当。但是在探测速度上,本文提出的基于镜头密度和帧间差的探测算法优势显著。同样地,以测试视频足球 1、足球 2 和足球 3 为例,采用本文提出的算法所需时间仅为基于帧间差的慢镜头探测算法的 1/10。分析其主要原因是一方面本文的算法避免了逆 DCT 变换,减少了很多的计算量;另一方面只需在候选区域进行多门限过零率计算,避免了逐帧比较,很大程度地提高了算法效率。

结束语 本文提出了一种新的慢镜头探测算法,定义了视频的镜头密度,提出了利用镜头密度进行候选慢镜头区域探测的具体方法。并在此基础上对如何通过帧间差实现慢镜头的探测进行了研究,取得了良好的效果。

在计算机、互联网和媒体技术迅速发展的信息时代,视频的生成方式和编辑手段越来越复杂和多样化。如何对大量种类繁多的视频进行处理、组织、管理和检索,成为迫切需要解决的问题,这需要更有效地对视频高层内容进行分析。分析视频高层内容涉及多个学科和技术领域。慢镜头探测技术也需要进一步深入研究,以下几个方面将是今后研究的主要方向:提高慢镜头探测算法的通用性、实现慢镜头探测中阈值的自动选择、研究对慢镜头内容进行分析的方法。

参考文献

- [1] Huang Q, Hu JM, Hu W, et al. A reliable logo and replay detector for sports video[C] // IEEE International Conference on

[2] Riley P F, Decker J C. Analysis architecture of a mobile sports replay system[C]//20th International Conference on Advanced Information Networking and Applications. Apr 2006

[3] Farn En-Jung, Hen Ling-Hweic, Iou J-H. A new Slow-motion Replay Extractor for Soccer Game Videos [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2003, 17(8):467-1481

[4] Pan H, Li B, Sezan M I. Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions[C]//ICASSP. 2002

[5] Pan Hao, van Beek P, Sezan M I. Detection of Slow-motion Re-

play Segments in Sports Video for Highlights Generation[C]//ICASSP'01. Salt Lake City, UT, USA, May 2001

[6] Babaguchi N, Kawai Y, Yasugi Y, et al. Linking Live and Replay Scenes in Broadcasted Sports Video[C]//Proceedings of the 8th ACM International Conference on Multimedia. CA, USA, October 30-November 4, 2000; 205-208

[7] Kobla V, DeMenthon D, Doermann D. Identification of sports videos using replay, text, and camera motion features [C] // Proc. of the SPIE Conference on Storage and Retrieval for Media Databases. Vol. 3972, Jan. 2000; 332-343

[8] 张福炎, 余崇志. 运动图像及其伴音的压缩与编码——MPEG1 [M]. 南京: 南京大学出版社, 1995

(上接第 282 页)

式(14)中, σ_0^2 是 t_0 上 R 的方差。

在分析窗 R_L 中, $P(s|R)$ 是均值为 $\mu_m(s|R)$ 、协方差为 $C(s|R)$ 的高斯分布。

2.2.2 计算回归参数 α, β

设在 5×5 的窗口 R_L 中, \bar{a}_n 是第 n 个 ($1 \sim 25$) 源像素点亮度值组成的矢量。 $\bar{a}_n = [a_{1n}, a_{2n}, \dots, a_{qn}]^T$, 其中 $a_{1n}, a_{2n}, \dots, a_{qn}$ 是 q 幅源图像同一个分析窗中的第 n 个像素点的亮度值。其矢量均值和协方差分别为

$$\bar{m}_a = \frac{1}{25} \sum_{n=1}^{25} \bar{a}_n$$

$$\Sigma_a = \frac{1}{25} \sum_{n=1}^{25} (\bar{a}_n - \bar{m}_a)(\bar{a}_n - \bar{m}_a)^T \quad (15)$$

以最大似然法 (MLE), 设似然比为 Z , 则

$$Z = \sum_{n=1}^N \ln[p(a_n|R)] = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln|C(a|R)| - \frac{1}{2} \sum_{n=1}^N \text{tr}\{(C(a|R))^{-1}(a_n - at_0 - \beta)(a_n - at_0 - \beta)^T\} \quad (16)$$

先估计参数 β 。通过似然比 Z 求关于 β 的偏导数。令偏导为零, 得

$$\beta_{ML} = m_a - at_0 \quad (17)$$

用式(17)的结果替换式(16)中的 β , 似然比 Z 变为

$$Z = \sum_{n=1}^N \ln[p(a_n|R)] = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln|C(a|R)| - \frac{1}{2} \sum_{n=1}^N \text{tr}\{(C(a|R))^{-1} \Sigma_a\} \quad (18)$$

计算参数 α 。以似然比 Z 求关于 α 的偏导数。令偏导为零, 得

$$\alpha = \frac{(\lambda_a - \sigma_a^2)^{\frac{1}{2}}}{\sigma_a^2 + \sigma_0^2} X_a \quad (19)$$

λ_a 和 X_a 分别为传感器噪声协方差矩阵 $\Sigma_a - \Sigma_e$ 的特征值及特征向量。

3 图像配准准则及实现过程

图像配准过程中, 配准准则及算子的选择, 对于配准后图像质量的影响至关重要。本文所涉及的医学图像配准处理基本步骤如下:

(1) 对各待配准图像(源图像)进行小波分解, 建立源图像的小波金字塔;

(2) 对各分解层分别进行概率估计处理, 得到配准后的小

波金字塔;

(3) 对配准后所得小波金字塔进行小波逆变换(进行图像重构), 最终所得到的重构图像即为配准图像。

对两幅待配准图像分别进行小波分解时, 一般分解的层数越多, 配准结果中的细节就越丰富。但分解层数多, 运算量也大, 且顶层配准损失的信息也越大, 故小波变换分解的层数不宜过多。

进行两幅图像配准时, 当第二幅源图像包含相反极性时, α_2 为负值, 配准图就是两幅源图像相减的结果。当极性相同时, 配准图就是两幅源图像相加的结果, 因此配准就完成了相加/减。

就互补特性而言, 源图像 1 中没有的特征, 在式(9)中对应 $\alpha_1 = 0$, 可通过源图像 2 得到补偿, 因而克服了简单平均法中减少对比度会丢失图像细节信息、增大信噪比的缺点。

当图像有噪声时, 如源图像 1 噪声大, 则对于配准图的贡献就减弱, 同时, 源图像 2 的贡献就加大。当两幅源图像的噪声相同时, 对配准图的干扰相同。因而, 通过配准减弱了源图像的噪声影响, 克服了特征配准法中会同时选择更多噪声的缺点。

结束语 MATLAB 验证表明, 当采用 3 层小波分解时, 本算法配准图像细节清晰, 且运算量小, 有较满意的配准精度。

参考文献

[1] 王娟, 师军, 吴宪祥. 图像拼接技术综述[J]. 计算机应用研究, 2008, 25(7):1940-1947

[2] 王海晖, 彭嘉雄. 基于多小波变换的图像融合研究[J]. 中国图象图形学报, 2004, 9(8):1002-1007

[3] 郭雷, 李晖晖, 鲍永生. 图像融合[M]. 北京: 电子工业出版社, 2008

[4] Sharma P K, Leen T K, Parvel M. Probabilistic Image Sensor Fusion[J]. Advances in Neural Information Processing Systems, 1999, 11:824-830

[5] 刘卫光. 图像信息融合与识别[M]. 北京: 电子工业出版社, 2008

[6] Bilmes J A. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models[R]. TR-97-021. International Computer Science Institute and Computer Science Division, University of California at Berkeley, 1998; 335-357