

文章编号:1004-4213(2010)06-1058-6

应用于视频编码的实时多测度联合突变场景切换检测算法*

马彦卓,常义林,杨海涛

(西安电子科技大学 ISN 重点实验室,西安 710071)

摘要:提出一种与视频编码器共用运动搜索模块的实时场景切换检测算法.通过分析原始图像与其运动补偿图像的一阶与二阶差值信号统计特性,分别构造对应于不同场景切换类型的两种测度函数,并联合使用该测度函数检测突变场景切换,同时自适应调节每个测度函数输出值的判决门限.分析和实验结果表明,该算法在视频编码位移估值基础上只增加少量的计算,而对视频序列中的突变场景切换检测的查全率和查准率均优于已有的几种实时检测算法.

关键词:图像信号处理;突变场景切换检测;多测度;实时;视频编码

中图分类号: TN911.73

文献标识码: A

doi: 10.3788/gzxb20103906.1058

0 引言

场景切换检测是视频信号处理研究领域的一个关键问题.它不仅可用于传统的基于内容的视频数据管理领域,也可用于视频编码领域用以提高视频编码效率和恢复视频质量.现有视频编码器,如 MPEG-4、H.264/AVC 等都包含率失真控制和码率分配模块,用于提高视频编码效率和得到平稳的输出码流.如果在视频编码的过程中,能够有效检测场景变化,尤其是突变场景变化(Scene Cuts, SC),并依据场景构造图像组(Group of Picture, GOP),再进行率失真控制,即对 GOP 中不同特性的帧分配比特,就能够提高码流的平稳性和解码恢复视频质量.

已有的突变场景检测算法使用的帧间非连续性的测度主要有帧间像素差域或变换域值^[1-4]、图像直方图^[5-7]、图像中的边缘等的变化特性^[8]或者视频压缩流数据特征^[9]等多种特征信息.在这些算法的门限值设置方法中,自适应设置门限值证明比固定门限值方法更为有效^[1,6].而在已有应用于视频编码速率控制的场景切换检测主要使用了帧间像素差值方法^[10-12],同时为了保证实时性而采用了较为简单的测度函数和预设门限值,因此影响了检测准确度.实际上,在已有各种场景检测方法中,与运动补偿相

结合的基于帧间像素差值的方法最为直观有效^[1,5].在将场景切换技术应用于视频编码速率控制时,可以将基于运动补偿帧间像素差值的检测方法运动搜索与视频编码中的位移估值结合起来,同时结合自适应门限值设定方法,这样就不会显著增加运算复杂度,同时可以得到较高的检测准确度.

本文提出了一种与视频编码器共用运动搜索模块的场景切换实现框架,并在框架下实现基于帧间像素差值信息的自适应门限值场景切换检测算法.因为场景变化会同时引起视频相邻帧多方面统计数据的多种变化特征,与已有使用单测度函数单门限值的各种场景检测方法不同,在分析视频相邻帧间数据多种变化特征的基础上,本算法联合使用对应于运动补偿后帧间绝对差值图像不同统计特性的多个测度函数来检测视频内容中的场景变化.

1 实时多测度场景切换检测算法

本文中场景切换检测使用与视频编码中的位移估值一样的运动搜索方法.首先以规定大小的图像块为基本单位进行块匹配搜索,在参考帧中找到当前帧每一个块的一个“最匹配”的块,然后,通过运动补偿可生成当前帧的运动补偿帧,当前帧和运动补偿帧的对应像素相减并取绝对值形成帧间绝对差值图像,文中简称为运动帧差图像(Compensated Absolute Difference Picture).

当未发生场景切换时,多个运动帧差图像像素均值大致相等;而发生场景切换时,运动帧差图像中像素的特征会发生明显变化.本文通过分析运动补偿后的一阶与二阶帧差信号特征,针对不同突变场景切换

* 国家自然科学基金(60772134)、高等学校学科创新引智计划(B08038)、西电一华为多媒体通信联合实验室合作专项基金和西电研究生创新基金(创05018)资助

Tel: 029-88202527-828 Email: yzma@mail.xidian.edu.cn

收稿日期: 2008-10-21

修回日期: 2009-02-18

类型,把变化大的区域占整个图像的比例以及运动帧差图像像素均值的变化幅度作为非连续性测度,联合自适应检测场景变化,提高检测的正确性.

1.1 算法综述

图 1 给出了用于视频编码的突变场景切换检测算法的实现.圆角方框内是场景检测模块,虚线框内模块则完成视频编码的功能.场景切换检测模块对帧间位移估计和运动补偿后得到的运动帧差图像进行统计分析,自适应判决当前编码帧是否发生场景切换,从而为编码器速率分配提供参考,并为视频内容管理提供信息.

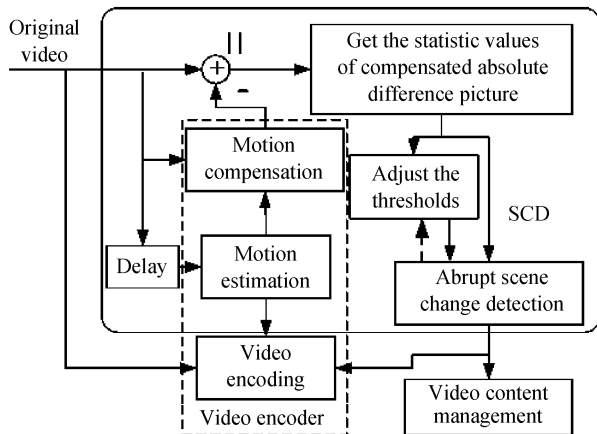


图 1 用于视频编码的突变场景切换检测

Fig. 1 Scene cut detection used for video coding

由于视频编码器在视频编码过程中也需要对视频进行帧间位移估计,所以在场景检测中帧间位移估计所得到的运动矢量仍然可以用于视频编码,减

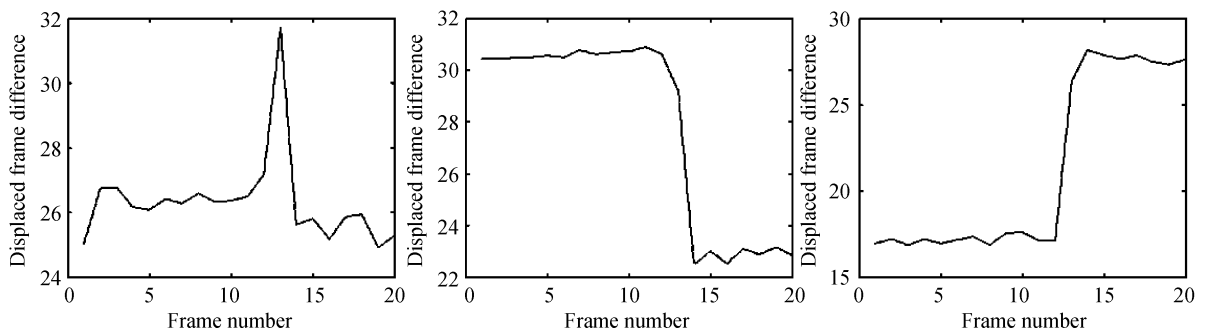
少了重复计算.这样由于共用了运动搜索(即视频编码器的位移估值)模块,对视频编码器而言引入的计算量不大,本算法可以为编码器提供准实时的场景切换信息.

1.2 非连续性测度函数与自适应判决门限

1.2.1 非连续性测度函数

通常认为,同一场景内部各个图像之间存在很强的相关性,因此运动帧差图像中的像素值都较小.但对于两个场景交界的相邻帧,图 2 为不同场景切换类型中的运动帧差图像像素均值,即运动补偿帧间差值(Displaced Frame Difference, DFD)均值的变化,该图为突变场景切换位于第 13 帧.由于内容的剧烈变化,运动帧差图像像素均值会陡然增加,呈现如图 2(a)所示的峰值特性.因此,已有算法多根据这种变化情况进行场景检测^[1-2,11-12].

但是,事实上,实验中还观测到图 2(b)、(c)的类似阶跃函数的变化情况.图 2(b)是从 Children 序列切变为 Mother&Daughter 时的数据,此时场景交界的相邻两帧帧间运动帧差图像的均值没有明显高于前面一个场景内的运动帧差图像.通过分析发现其原因在于前一场景内物体的复杂运动产生一些绝对差值较大的大变化块,而后一场景内纹理和运动都相对简单,帧间绝对差值远小于前一个场景.这时场景切换后运动帧差图像均值显著减小,近似呈现负阶跃函数特征.同理,在场景切换时也会产生如图 2(c)所示的近似正阶跃函数特征.



(a) DFD changing with a peak value (b) DFD changing in negative step function pattern (c) DFD changing in step function pattern

图 2 不同场景切换类型中的运动帧差图像像素均值

Fig. 2 Variation of average compensated absolute difference

然而,仅靠运动帧差图像像素均值的变化不能够准确判断突变场景切换,当视频图像局部区域发生剧烈变化时会发生误判.在这种情况下,如果能够综合考虑运动差值图像中具有较大能量区域的大小,就可以更加有效地判断场景切换.

定义测度函数 $CAR(n)$ 为当前运动帧差图像中具有较大像素值的图像区域占整幅图像面积的比重,以运动搜索块为单位统计. $CAR(n)$ 表示为

$$CAR(n) = \frac{N_{CB}(n)}{N} \cdot \frac{S_{CA}(n)}{S} = \frac{N_{CB}(n)}{N} \cdot \frac{H_{span}(n)gV_{span}(n)}{H \cdot V} \quad (1)$$

当对应第 n 帧有 $CAR(n) > Th_1(n)$ 时,判决发生场景变化.其中 N 为一帧图像内总的块的个数, $N_{CB}(n)$ 为当前帧变化大的块的个数.

$$N_{CB}(n) = \sum_{i=0}^{N-1} CB_i(n) \quad (2)$$

式中 $CB_i(n)$ 表示当前第 n 帧第 i 块是否为变化大的块

$$CB_i(n) = \begin{cases} 0, & DFD_{B,i}(n) < Th_3 \\ 1, & DFD_{B,i}(n) \geq Th_3 \end{cases}, i \in [0, N-1] \quad (3)$$

式中 $DFD_{B,i}$ 为运动帧差图像中第 i 块的均值, Th_3 为检测门限. 式(1)中 $S = H \cdot V$ 为图像的总像素数, $S_{CA} = H_{span} \cdot V_{span}$, 为变化大的区域的面积. H_{span} 和 V_{span} 分别代表了大变化块所占区域的水平和垂直跨度. 设帧差图像最左上角的像素坐标为 $(0, 0)$, 第 i 个块的左上角和右下角像素坐标分别为 $(h_{il}, v_{il}), (h_{ir}, v_{ir})$, 定义

$$\begin{cases} H_1 = \min(h_{il} | DFD_{B,i} \geq Th_3, i \in [0, N-1]) \\ H_2 = \max(h_{ir} | DFD_{B,i} \geq Th_3, i \in [0, N-1]) \end{cases} \quad (4)$$

则有 $H_{span} = H_2 - H_1$; 采用同样的方法定义可得到垂直方向跨度 V_{span} .

$CAR(n)$ 可以准确检测对应于图 2(a) 的场景切换检测, 并避免局部区域变化引起的误检. 针对图 2(b) 和 (c) 中的场景切换类型, 定义第二个场景切换检测测度函数为帧间像素值的二阶差值

$$V_{DFD}(n) = DFD(n) - \max(DFD(n \pm j) | j \in [1, W]) \quad (5)$$

当对应第 n 帧有 $V_{DFD}(n) > Th_2(n)$ 时, 同样判决发生场景切换. 其中 $DFD(n)$ 为第 n 帧运动帧差图像信号的均值, 减去的部分是参考窗口中 $DFD(n)$ 的最大值, “+”号对应于后向时间窗口, “-”号对应前向时间窗口. $DFD(n)$ 表示为

$$DFD(n) = \frac{1}{N} \sum_{i=0}^{N-1} DFD_{B,i}(n) \quad (6)$$

1.2.2 自适应判决门限

由于各个序列和场景中场景复杂度、场景中物体运动剧烈程度不同, 采用固定门限不能检测出不同情况下的场景切换, 需要自适应调节判决门限. 判决门限是根据参考时间窗口内连续几帧的运动帧差图像数据的统计值调节. 对于第 n 帧 $Th_1(n)$ 调节如式(7), 其中 W 为窗口宽度. λ 为可变参数, T_1 的引入是防止场景内部部分原静止物体产生运动而发生虚假场景切换误判. $Th_2(n)$ 的调节算法如式(8), 其中 γ 为可变的参数, T_2 的引入同样是为了防止在场景内部分原运动物体变为静止而发生虚假场景切换误判. 式中“+”号对应于后向, “-”号对应于前向窗口, 即对 $V_{DFD}(n)$ 在两个方向的窗口内分别对应有一个自适应门限值.

$$Th_1(n) = \max(\lambda \times \max(CAR(n-1), \dots, CAR(n-W)), T_1) \quad (7)$$

$$Th_2(n) = \max((\max(DFD(n \pm j) | j \in [1, W]) - \frac{1}{W} \sum_{j=1}^W DFD(n \pm j)) \times \gamma, T_2) \quad (8)$$

式(3)中使用的门限值调节算法为

$$Th_3(n) = \frac{1}{W} \sum_{j=1}^W (\max(DFD_{B,i}(n-j))), i \in [1, N] \quad (9)$$

1.3 场景切换检测

本文提出的算法在两种情况下会判决当前发生场景切换, 其中一种是当前图像中不只是有相当数量的大变化块, 而且这些块大面积覆盖图像; 另一种情况是当一帧运动帧差图像的均值与检测窗口中帧差图像的最大均值之差超过一定门限. 检测算法的执行过程如图 3.

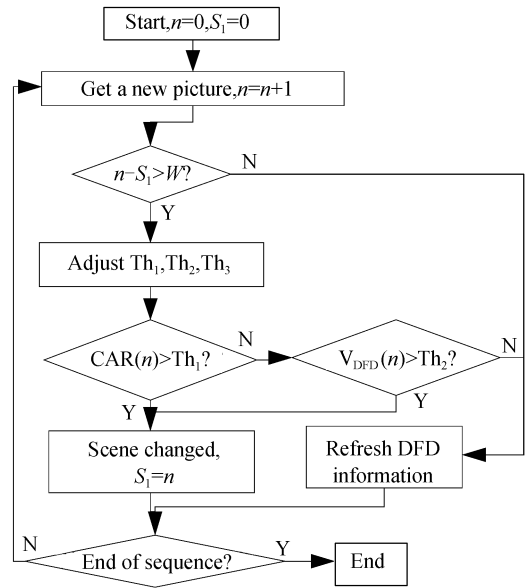


图 3 场景切换检测流程

Fig. 3 Flowchart of scene cut detection

为了减少计算量, 算法执行前可先对原始序列进行下采样.

2 场景切换检测实验及其结果分析

为检测突变场景变化检测方法的准确度, 实验中使用了多种不同特性的序列作为场景检测的对象, 有标准视频序列的组合序列、新闻、电影、歌曲 MV 等. 不失一般性, 所有序列下采样为 QCIF (176×144 个像素大小) 格式后进行场景变化检测.

表 1 本文实验所使用的序列

Table 1 Sequences used in the experiments

Sequence	Content
A	A merged sequence by Akiyo etc., including the first 17 frames of each sequence, CIF(352×288)
B	Clips of the film King of Kong fu, HDTV(1280×720)
C	Clips of news, VGA(640×480)
D	Clips of MV of the song iShop, VGA(640×480)

实验中将提出的算法和两个性能公认较好其他场景切换检测算法, 以及与视频编码器结合的几种算法

作了对比分析,分别是自适应门限值的 χ^2 直方图方法[14]、和文献[13]中提出的帧间像素差方法,以及文献[10-12]中的实时检测算法。

场景切换检测算法的性能一般使用查全率 (Recall) 和查准率 (Precision) 两个指标来衡量。查全率是检测到的正确场景切换 N_c 与所有实际场景切换数之比,式(10)中 N_m 为漏检数;查准率是正确检测到的场景切换数与检测报告的所有场景切换数之比,式(11)中 N_f 为误检数。

$$\text{Recall} = \frac{N_c}{N_c + N_m} \quad (10)$$

$$\text{Precision} = \frac{N_c}{N_c + N_f} \quad (11)$$

算法中的五个参量, λ 、 T_1 、 γ 、 T_2 以及窗口宽度 W , 通过对上述四个序列之外的大量视频的统计获得,如表 2。特别地,对于窗口宽度的取值,我们发现其对不同类型的视频序列应有不同取值,对大部分视频序列,如新闻、电影、电视剧等,该取值可以不少于半秒的持续时间对应的帧数,如 12;而对于广告、歌曲 MV 等充满大量编辑特效的视频序列,该取值应相对较小,如 7。另外,实验中运动搜索块大小为 4×4 。

表 2 本算法实验中选取的配置参量

Table 2 Selected parameters for the experiments using the proposed algorithm

Parameter	Value	Parameter	Value	Parameter	Value
λ	1.5	γ	0.5	Forward W	12/7
T_1	0.063	T_2	1	Reverse W	3

本方法是面向编码应用的场景检测,主要目的是为有效的速率控制等提供重要信息和依据,因此突变场景是它的主要检测目标。表 3 给出了进行对比的六种方法分别对测试序列中突变场景切换的检测结果。

从表 3 中给出的结果可以看到,对于序列 A~D 中突发的场景切换,提出的方法得到的查全率保持在 90% 以上,有些达到 100%;查准率也均在 80% 以上。总体看来,本文提出方法查全率和查准率优于其他几种方法。然而对于序列 D,本文方法略差于改进的 χ^2 直方图方法。这是由于歌曲 MV 中的人工编辑多而复杂,场景内出现大范围的突然的运动变化,而本文算法对这种情况较敏感而造成了误判。 χ^2 直方图方法提供了与提出的方法近似的性能;文献[13]提出的方法对序列 A 和 C 有较好的检测结果,接近于 χ^2 直方图方法,而对序列 B 和 D 检测时漏检率很高,这是由于该方法只能采用固定门限值,缺乏对多种序列的自适应性造成的。而文献[10-12]中的多种实时检测算法检测准确度都比较

低,并体现出对不同序列性能差异较大的特性。文献

表 3 对各个序列的突变场景切换检测结果

Table 3 Results of scene cut detection

Sequence	Frame/SCs	Method	False detections	Missed	Precision ratio/%	Recall ratio/%
A	884/51	χ^2 Histogram	2	2	96.08	96.08
		Reference[10]	0	35	100	31.37
		Reference[11]	7	2	87.50	96.08
		Reference[12]	0	11	100	78.43
		Reference[13]	8	1	86.21	98.04
		Proposed method	0	0	100	100
B	2532/46	χ^2 Histogram	9	3	82.69	93.48
		Reference[10]	4	29	80.95	36.96
		Reference[11]	159	4	20.90	91.30
		Reference[12]	0	30	100	34.78
		Reference[13]	3	36	76.92	21.74
		Proposed method	4	2	91.67	95.65
C	1468/24	χ^2 Histogram	3	3	87.50	87.50
		Reference[10]	4	9	78.95	62.50
		Reference[11]	570	0	4.04	100
		Reference[12]	12	4	62.50	83.33
		Reference[13]	8	3	72.41	87.50
		Proposed method	3	1	87.50	95.45
D	2866/53	χ^2 Histogram	8	4	85.96	92.45
		Reference[10]	3	41	80.00	22.64
		Reference[11]	79	0	40.15	100
		Reference[12]	9	25	75.68	52.83
		Reference[13]	5	17	78.26	67.92
		Proposed method	10	4	83.05	92.45

[10]中的检测方法除对序列 C 检测结果达到了查全查准率都在 60% 以上外,对其他序列检测漏检率都很高;文献[11]中的检测算法除对序列 A 提供了高于 80% 的查全查准率,对其他序列的检测中虽然漏检很少,但是误检都很多,导致非常低的查准率;文献[12]中的检测算法略差于文献[13]中算法的检测结果,多数情况下提供了优于文献[10]和文献[11]中的算法的性能,但是对序列 B 检测时漏检非常多。

以上结果表明提出算法的有效性和鲁棒性都较好。其原因在于,全面考虑了由于场景切换引起的帧差图像数据突变的情况,即不只是检测图 2(a) 的冲击函数型突变情况,而且要检测图 2(b) 和 (c) 的阶跃函数型突变情况,并相应提出两个非连续性测度函数,采用了自适应调节的判决门限。

实验中采用了较小的 4×4 块作为位移估值单元,当与 H.264/AVC 视频编码器结合时,它为视频编码种的帧内帧间不同编码模式提供了有效的判决依据。

对于场景切换帧可以确定为视频编码的 Intra 帧,从而不需要再进行编码模式判决。在没有对序列

下采样的前提下,在一个 16×16 宏块中,如果有多个 4×4 块有相同位移矢量,可将它们合并,从而可确定帧间预测编码的不同模式 $16 \times 16, 8 \times 16, \dots, 4 \times 8$ 等,无需再次计算帧间编码模式. 即使对于 4×4 的小的帧间编码模式,在场景检测中得到的位移估值完全可用作为视频编码的位移估值,无需再次运算. 在序列下采样的情况下,得到的位移估值也可在编码中作为搜索起始点等方式重用.

由此可见,场景检测过程中的运动搜索结果可以用于视频编码,提高视频编码器帧内帧间编码模式选择和位移估值的运算速度,从而整体上没有增加大的计算量.

3 结 论

本文分析了场景切换引起的运动补偿帧间绝对差值图像突变的有关情况,提出了一个完整的突变场景切换检测算法,包括构造了两个对应于帧间差值的统计特性的非连续性测度函数,判决门限的自适应调节方法,以及一些参量的适当选取. 一般算法只检测了运动帧间绝对差值图像均值的单峰型突变情况,而本文增加了阶跃函数型突变情况的检测,提高了场景切换检测的正确率.

本文提出的场景切换检测算法可用于视频编码,视频编码器根据场景变化的检测结果自适应生成 GOP,进一步结合 GOP 内关键帧检测^[15],编码器可以按照 GOP 中帧的重要性分配比特,实现率失真控制,提高视频编码效率和恢复视频质量. 我们将另文深入讨论这一问题. 同样,使用该算法得到的场景检测结果也可用于编码后的视频数据的内容管理. 由于该算法和视频编码紧密结合,联合实现视频编码中的位移估值和场景检测中的运动搜索,降低算法的运算复杂度.

参考文献

[1] KOPRINSKA I, CARRATO S. Temporal video segmentation: A survey[J]. *Eurasip Sig Proc: Image Communication*, 2001,

- 16(5):77-500.
- [2] HANJALIC A. Shot-boundary detection: unraveled and resolved[J]. *IEEE Trans on CSVT*, 2002, 12(2):90-105.
- [3] HOU Yu-qing, GE Bao, PENG Jin-ye. A novel approach to the detection of shot boundary [J]. *Scope on Acta Photonica Sinica*, 2005, 34(10):1597-1600.
侯榆青, 葛宝, 彭进业. 视频分镜头边界检测的一种新方法[J]. *光子学报*, 2005, 34(10):1597-1600.
- [4] DAI Xiao-wen, CAI Zhi-ping, ZHONG Gui-ying. An algorithm of video shot detection based on partitioning image wavelet entropy[J]. *Scope on Acta Photonica Sinica*, 2008, 37(7):1492-1496.
戴小文, 蔡志平, 钟桂英. 基于分块图像小波相对能量和熵的视频镜头边界检测算法[J]. *光子学报*, 2008, 37(7):1492-1496.
- [5] BORECZKY J B, ROWE L A. Comparison of video shot boundary detection techniques[J]. *J Electron Imaging*, 1996, 5(2):122-128.
- [6] GARGI U, KASTURI R, STRAYER S H. Performance characterization of video-shot-change detection methods[J]. *IEEE Trans CSVT*, 2000, 10(1):1-13.
- [7] TIAN Yu-min, WANG Hao. Dynamic window technique in shot boundary detection[J]. *Acta Photonica Sinica*, 2007, 36(10):1949-1953.
田玉敏, 王昊. 镜头边界检测的动态窗口技术[J]. *光子学报*, 2007, 36(10):1949-1953.
- [8] LIENHART R. Comparison of automatic shot boundary detection algorithms[C]. *SPIE*, 1999, 3656:290-301.
- [9] FOUAD M A, BAYOUMI F M, ONSI H M, et al. Real-time shot transition detection in compressed MPEG video streams [J]. *Journal of Electronic Imaging*, 2008, 17(2):1-16.
- [10] LIN Shu. Scene change detection for video [P]. World, WO2007142646A1, Dec. 2007.
- [11] DING Jun-ren, YANG Jar-ferr. Adaptive group-of-pictures and scene change detection methods based on existing H. 264 advanced video coding information [J]. *IET Image Processing*, 2008, 2(2):85-94.
- [12] LEE C, JUNG Y, LEE S, et al. Cost-effective scene change detection algorithm for real-time H. 264 rate control [J]. *Optical Engineering*, 2008, 47(3):1-3.
- [13] YI X, LING N. Fast pixel-based video scene change detection [C]. *IEEE Intl Sym Circuits Sys*, 2005, 4:3443-3446.
- [14] NAGASAKA A, TANAKA Y. Automatic video indexing and full-video search for object appearances, in *Visual Database Systems II*[M]. Elsevier, 1995:113-127.
- [15] MA Yan-zhuo, CHANG Yi-lin, YUAN Hui. Key-frame extraction based on motion acceleration [J]. *Optical Engineering*, 2008, 47(9):1-3.

Multi-Metric Combined Real-Time Scene Cut Detection Algorithm for Video Coding

MA Yan-zhuo, CHANG Yi-lin, YANG Hai-tao

(State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China)

Abstract: A real-time scene cut detection algorithm is proposed which shares the results of motion estimation with the video encoder. To deal with different kinds of scene cuts, two metric functions are introduced based on the statistics of the first and second order difference between the original pictures and their motion compensated pictures. The detection thresholds for each metric are adaptively adjusted according to the scenes. The experimental results and analysis show that the recall and precision of the proposed algorithm are generally higher than other tested ones at the cost of little computation time increment.

Key words: Image signal processing; Scene cut detection; Multi-metric; Real-time; Video coding



MA Yan-zhuo was born in 1980. She received the Ph. D. degree of engineering from Xidian University in 2009. Her research interests focus on video signal processing, compressing and communication.



CHANG Yi-lin was born in 1944. He is currently a professor at School of Telecommunications Engineering, Xidian University. His major interests focus on multimedia in network, wireless video transmission and network management.