

一种融合颜色和运动特征的镜头探测方法^{*}

陈剑贇 李云浩 老松扬 吴玲达

(国防科技大学多媒体研究开发中心 长沙410073)

A Shot-Detection Scheme Combining Color and Motion Features

CHEN Jian-Yun LI Yun-Hao LAO Song-Yang WU Ling-Da

(Multimedia R&D Center, National University of Defense and Technology, Changsha 410073)

Abstract Video Partitioning serves as the preliminary step to structure the content of videos and is the basis of Content-based Video Retrieval. This paper presents a shot detection scheme combining color and motion features of videos. It first extracts the color and motion information from the compressed domain videos, then detects cut transitions and gradual transitions separately incorporating color features with motion ones. The experiment demonstrates real time performance with high accuracy.

Keywords Shot detection, Combining color and motion features, Cut transitions, Gradual transitions

1 引言

随着多媒体技术的发展和应用需求的不断增长,多媒体数据库中视频数据越来越多,对视频数据的检索要求也相应增加。镜头探测是实现基于内容视频检索最基本的步骤,因为镜头是视频内容表现最基本的单位,一个镜头即是一段未被打断的视频,它描绘同一场景,表示的是一个摄像机动作,一个事件或连续的动作。在基于内容的检索中,对视频的处理如视频代表帧的抽取等操作都是基于镜头进行的,因此镜头的分割是视频内容分析的基础。

根据两个相邻镜头间视频内容的变化,可以把镜头边界分为突变和渐变两种。突变镜头边界由硬切造成,渐变镜头边界则由溶化(dissolve)、扫换(wipe)等编辑手法造成。对于突变探测,主要方法就是衡量相邻帧之间亮度或色度的差异。其中应用最为广泛的是计算帧间颜色直方图的差值,因为颜色的分布不受摄像机或者对象运动的影响。设视频帧 f_i 和 f_j , 两者直方图表示分别为 H^i 、 H^j , 直方图共包含 n 种颜色, 则其颜色直方图的差值定义为: $Dis(H^i, H^j) = \sum_{k=1}^n (H_k^i - H_k^j)^2$, 当 Dis 大于阈值, 就判定此处发生镜头突变。尽管通过比较帧间颜色差异可以较好地探测突变, 但这种方法对渐变并不很奏效, 因为相对突变而言, 渐变持续时间较长, 并且帧间颜色的差异没有突变那么大。渐变探测比较经典的方法是文[1]提出的双重比较法, 就是采用两个阈值来检测渐变。低阈值用于探测潜在的渐变边界, 然后累计候选区域相继帧之间的颜色差值, 当累计值超过高阈值时, 表明有渐变发生。文[2]提出了一种利用二维参数运动模型探测突变渐变的统一的框架。这种方法侧重于利用运动信息进行镜头探测, 但是这种方法计算比较复杂, 不能达到实时的效果。由于造成渐变的视频编辑手法多种多样, 目前渐变探测仍然是一个难点问题。

以上简要介绍了现有镜头探测的基本方法, 可见多数算法都是利用单一的颜色信息或者单一的运动信息分析视频帧的差异, 而很少考虑两者的结合。本文提出一种融合颜色和运动特征的镜头探测方法, 这种方法的主要特点是: 在压缩域内

抽取视频帧的颜色和运动特征, 融合视频颜色和运动特征进行突变和渐变探测, 快速有效, 鲁棒性强。

2 融合颜色和运动特征进行镜头探测的基本过程

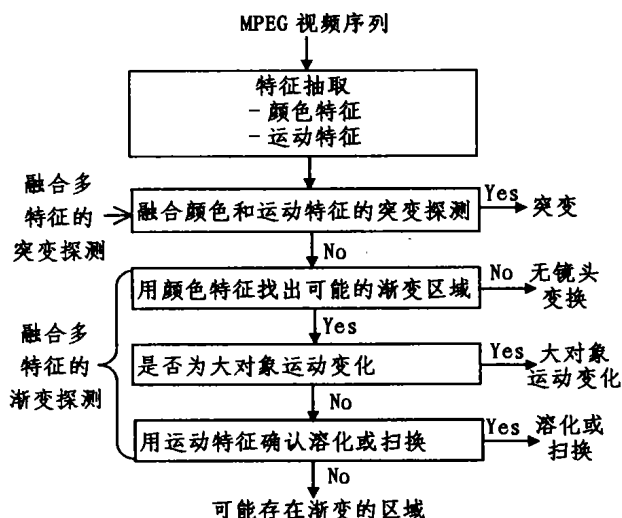


图1 融合颜色和运动特征进行镜头探测过程图

图1为融合颜色和运动特征的镜头探测过程图。对于要处理的视频序列, 首先抽取压缩域内的颜色和运动特征。由于这些特征的抽取在压缩域内进行, 所以运行速度快。第3节将详细介绍从 MPEG-1 视频序列中抽取颜色和运动特征的问题。由于突变探测的精度较高, 我们首先探测在当前帧是否有突变发生, 突变探测的准则融合了颜色和运动特征, 比单纯依靠直方图差值鲁棒性要强很多。如果没有突变发生, 则探测在一段区间内是否有渐变发生。首先利用颜色特征探测出可能存在渐变的区域, 但是大对象运动变化和渐变具有类似的颜色特征, 所以需要剔除出大对象运动变化。所谓大对象运动变化, 是指一个镜头内有一个占画面大部分的运动目标在运动, 如一辆汽车驶过画面, 或蓝色的大海频繁地翻卷着白色的浪

^{*} 本文受十五国防预研项目资助。

花等。这种视频帧间颜色变化较大,使得颜色差值连续出现高值,类似于渐变过程的颜色特征,但大对象运动过程实际上有着与渐变过程不同的特点。在此简单考虑了大对象运动变化的探测。接下来利用运动特征进一步确认溶化或者扫换的发生,如果该区域不具有溶化或扫换发生的运动特征,表示这

个区域可能存在渐变,但是仍然不确认。第4节将有详细的探测算法的介绍。

3 压缩域内视频帧颜色和运动特征的抽取

图2(a)为按显示顺序表示的典型的MPEG-1视频序列

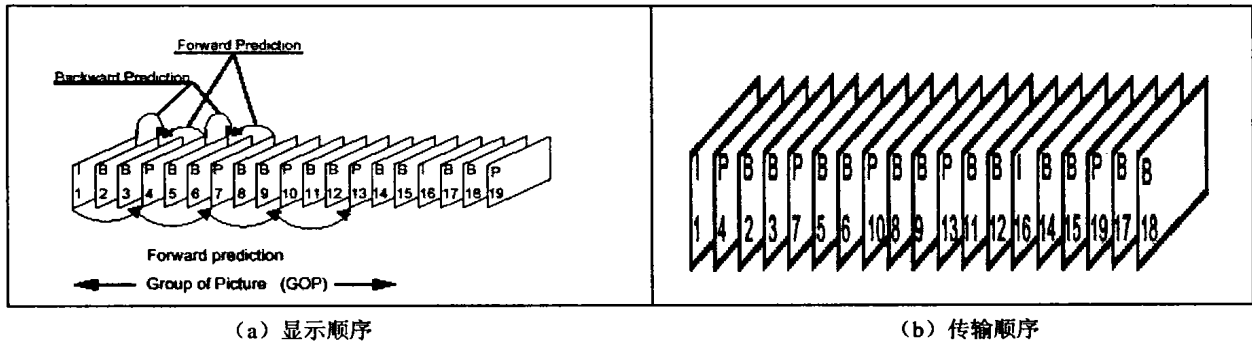


图2 典型的MPEG-1视频序列

图: IBBPBBPBBPBBPBBIBBP。可见,在MPEG-1压缩序列中,视频帧组织在图像组层(GOP)中。一个GOP中,第一帧为帧内编码帧——I帧,P帧为预测编码帧,I、P帧合称为锚帧(Anchor Frames),B帧是双向预测帧。图2(b)是(a)的视频序列的传输顺序表示。

在以上典型MPEG-1视频序列的基础之上,我们抽取以下的颜色和运动信息的统计量来衡量视频帧之间的差异:

对于颜色信息,考虑锚帧之间在YUV色彩空间上的颜色差异。设锚帧*i*和*j*,则颜色距离定义为:

$$D(i, j) = |Y_i - Y_j| + |\sigma_Y - \sigma_Y| + w * (|U_i - U_j| + |\sigma_U - \sigma_U| + |\sigma_V - \sigma_V|)$$

其中 Y, U, V 是DC图像 Y, U, V 值的平均值, $\sigma_Y, \sigma_U, \sigma_V$ 分别是 Y, U, V 的标准方差; w 是色度值 U 和 V 的权重。

若当前帧为I帧,则计算 $D_{NEAR}(i)$ 和 $D_{FAR}(i)$ 。其中 $D_{NEAR}(i)$ 表示当前帧与其前面最近锚帧之间的颜色差值,即 $D_{NEAR}(i) = D(i, i - M - 1)$ (M 表示相邻两锚帧间B帧的数目,在此取2)。 $D_{FAR}(i)$ 表示当前帧和相隔一段传输距离的锚帧之间的颜色差值,即 $D_{FAR}(i) = D(i, i - (M + 1) * k)$ ($k > 1$)。若当前

帧为P帧,同I帧一样,计算 $D_{NEAR}(i)$ 和 $D_{FAR}(i)$ 。此外,计算比率 R_p 以衡量运动特征。 R_p 定义为: $R_p = \frac{M_{INTRA}}{M_{PRE}}$,其中 M_{INTRA} 和 M_{PRE} 分别表示帧内编码和预测编码宏块的数目。

对于B帧,主要考虑其运动信息,计算比率 R_f 和 R_b ; $R_f = M_{fwd} / M_{bwd}$, $R_b = M_{bwd} / M_{fwd}$,其中 M_{fwd} , M_{bwd} 分别记录了B帧中向前预测编码宏块和向后预测编码宏块的数目。

4 融合颜色和运动特征的镜头探测方法

以下将在第3节颜色和运动特征抽取的基础之上,详细介绍融合颜色和运动特征的镜头探测方法。对于突变,重要的是探测出帧间差异的峰值,以往的算法^[8]只考虑了I帧之间的颜色差值,而忽略了突变发生在P帧和B帧上的情况,鲁棒性不强。4.1小节将结合颜色和运动特征讨论突变分别发生在3种帧上的情况。4.2小节讨论渐变的探测。

4.1 融合颜色和运动特征的突变探测

图3为传输顺序显示的一小段MPEG视频序列。假设突变发生在第*k*帧。

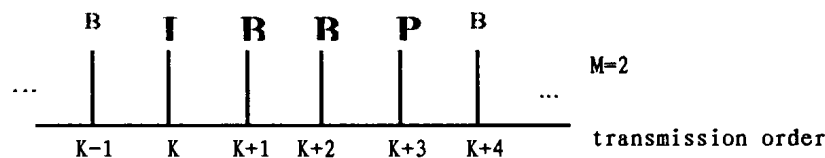


图3 传输顺序显示的一小段MPEG视频序列

如图3所示,第*k*帧为I帧,计算 $D_{NEAR}(k)$, $D_{NEAR}(k + M + 1)$ 和 $R_b(k + j)$, $j = 1, \dots, M$ 。 M 为当前锚帧与下一个锚帧之间B帧的数目。首先确认 $D_{NEAR}(k)$ 大于阈值,即在*k*帧处可能有突变发生;然后确认 $D_{NEAR}(k + M + 1)$ 小于阈值,即在下一个锚帧*k* + M + 1处没有突变发生,因为在此假设紧挨的几帧内不出现连续的突变;最后利用运动特征确认所有的 $R_b(k + j)$ ($j = 1, \dots, M$)都小于阈值,也就是 $R_b(k + j)$ 足够小,在*k* + j 帧处没有突变发生。过程如D1总结所示:

D1: if ($D_{NEAR}(k) < \tau_{COLOR}$) return false
Else if ($D_{NEAR}(k + M + 1) > \tau_{COLOR}$) return false
Else if (All $R_b(k + j) < \tau_{RB}$, $j = 1, \dots, M$) return TRUE
Else return false

如果第*k*帧为P帧,计算 $R_p(k)$, $D_{NEAR}(k)$, $D_{NEAR}(k + M$

+ 1)和 $R_b(k + j)$, $j = 1, \dots, M$ 。判断过程与I帧类似,唯一区别的是: $R_p(k)$ 大于某阈值时,表明可能在*k*帧处有突变发生。过程如D2总结所示:

D2: if NOT ($D_{NEAR}(k) > \tau_{COLOR}$ OR $R_p(k) > \tau_{RP}$) return false
Else if ($D_{NEAR}(k + M + 1) > \tau_{COLOR}$) return false
Else if (All $R_b(k + j) < \tau_{RB}$, $j = 1, \dots, M$) return TRUE
Else return false

如果第*k*帧为B帧,计算当前B帧和它相继B帧的 R_f 和 R_b 的值。如果这些B帧的 R_b 都大于阈值,表明在*k*帧处存在突变。否则,如果当前B帧的 R_b 小于阈值,而相继B帧的 R_f 大于阈值,则表明传输顺序中向前与*k*帧最邻近的锚帧存在有突变。过程如D3总结所示:

D3: if (All $R_b(k + j) > \tau_{RB}$, $j = 0, \dots, L$) (L 为当前B帧与下一个

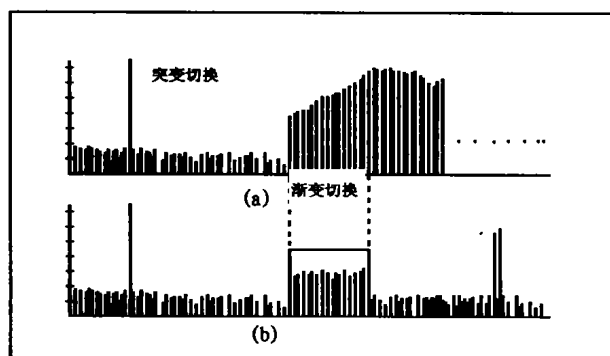
```

    锚帧间 B 帧的数目)
    return TRUE
Else
    if ( $R_b(k) < \tau_{RB}$  AND All  $R_f(j) > \tau_{RF}, j = k+1 \dots L$ )
        return TRUE

```

4.2 融合颜色和运动特征的渐变探测

渐变探测中,溶化和扫换是两种主要的编辑模式。融合颜色和运动特征的渐变探测过程分两步进行:一是用双颜色差值比较法探测出可能发生渐变的区域,并且根据颜色特征区别渐变与大对象运动变化;然后根据运动特征进一步确定溶化或者扫换的发生。



(a)与潜在起始帧比较的颜色差值; (b)帧间颜色差值
图4 双颜色差值比较法

双颜色差值比较法指检测过程中采用两个颜色差值,一个用于记录帧间颜色的差值,一个比较渐变区域内的锚帧与潜在起始帧的颜色差值。首先用第一个颜色差值,通过与较低的阈值比较来确定渐变切换序列的潜在起始帧。一旦确定了这个帧,后续所有锚帧均与这一帧进行比较,得到第二个颜色差值。如果这确是一个渐变切换,后续帧与潜在起始帧的颜色特征差值将随时间增加越来越大(因为起始帧将不断地被另一镜头的画面所代替,因此起始帧的特征将越来越小,而切换结束后的镜头的图像帧的特征将越来越大)(如图4(a)),且在切换结束时的图像帧与起始帧的颜色差值将具有突变的特征,即将超过第二个阈值。这以后的所有锚帧和起始帧的差值

可能都具有突变的差值甚至大于此差值,也可能又小于突变差值,但都较大。因此在进行这个序列比较的同时,仍然进行帧间颜色差值的比较,当序列为渐变序列时,帧间的颜色差值将大于一个镜头内的颜色差值,一旦帧间差值又减小了,说明切换结束。如图4(b)所示。

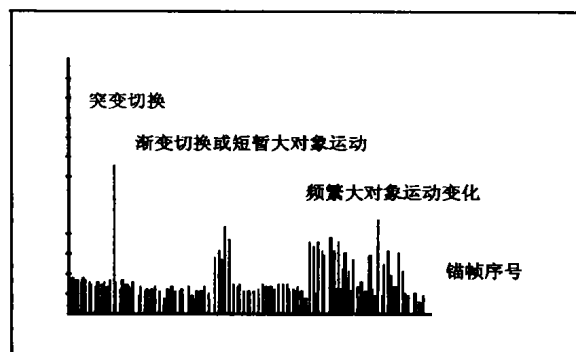


图5 大对象运行变化与镜头切换的差别

找出可能存在的渐变区域,但是大对象运动变化与渐变存在类似的颜色变化特征,如图5所示。大对象运动变化主要包括二种情况:(1)短暂的大对象运动变化:这种变化表现为在一个镜头内,突然有另外的目标进入镜头,并很快又离开镜头,如上面所提到的一辆汽车突然驶过整个画面的情况。这种情况不会由于短暂的目标闯入而影响整个镜头的内容,所以决不能将这种变化当作渐变过程把前后两段视频分割为两个镜头。另外一种情况是突然有一大目标进入画面并停留在画面上,这种情况虽然摄像机没有变化,但由于大目标的介入,使得整段视频内容发生相应的变化,因此也可以认为这是两个镜头,把大对象运动目标的突然进入作为一个突变过程处理,在此不作处理。(2)频繁的大对象运动变化:这种变化过程表现为在一个镜头内部有一个大目标在频繁运动,并在色彩、亮度上也表现为相应的变化,如上述提到的浪花的情况。这种情况在颜色差值上表现为很多高值或高值中间偶尔有一些低值。

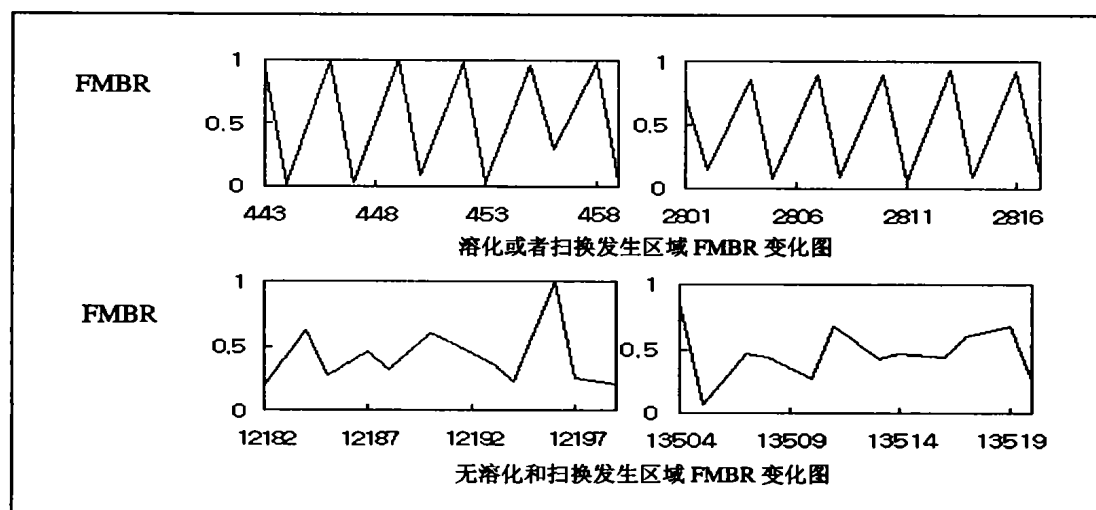


图6 FMBR 变化规律图

根据以上的特征,我们利用以下规则从可能的渐变区域探测出大对象运动变化:(a)如在第 i 个锚帧和第 $i+n$ (n 大于

1)个锚帧间的差值一直为高值,且 n 不太大,同时后面为连续的较低差值,说明这是一个渐变过程,渐变以后(高差值后)又

是另一镜头(连续多个低差值);(b)如后面只有较少个的低差值,则前面的连续高差值必然为一个镜头内的大对象运动,因为后面的低差值太少,没有表现为一个镜头的持续过程;(c)如果 n 较大,说明这是一个连续的镜头内的大对象运动变化过程,因为渐变不可能延续太长的时间。这种变化也适合于判断摄像机的连续运动情况。(d)对于短暂的大对象运动目标的运动情况,第 i 帧到第 $i+n$ 帧的前后的颜色差值满足渐变过程,则计算第 $(i+n+1)$ 帧和第 i 帧的颜色差值,如差值很小,说明是大对象运动目标过后又回到了原来的镜头,因此判断为大对象运动目标,仍为一个镜头;如差值高于阈值,说明第 $(i+n+1)$ 帧和第 i 帧间已换了一个镜头,因此是一个渐变过程。

剔除出大对象运动变化,再利用运动特征进一步确认溶化或者扫换的发生。 M_{fwd} , M_{bwd} 分别表示 B 帧中向前预测宏块和向后预测宏块的数目,定义比率值 $FMBR = M_{fwd} / (M_{fwd} + M_{bwd})$,文[4]研究表明,溶化或扫换发生时 $FMBR$ 的特征如图6所示,可见当溶化或扫换发生时 $FMBR$ 的变化呈锯齿

状,我们利用这个特性进一步确认溶化或扫换的发生。

5 实验结果与分析

我们运用以上算法在多个视频源上实验。实验视频主要集中在两类:一是体育实况视频,二是新闻视频,因为这是应用前景广阔的两类视频。其中采用的体育实况视频包括排球、网球、篮球、足球等多种视频数据。这些视频数据通过视频采集卡从电视信号数字化压缩为 MPEG-1 视频流,采用 CIF 分辨率(352×240),视频流的比特率为 1.2Mbps 到 1.5Mbps。整个镜头探测系统运行在 Pentium III 750 的个人电脑上。

在我们的实验中,采用查全率 (Recall) 和查准率 (Precision) 来衡量实验结果。设定 N 为一段视频数据中实际包含的镜头数目, M 为漏判的镜头数, F 为误判的镜头数,则查全率和查准率定义如下:

$$\text{查全率} = \frac{N-M}{N}, \text{查准率} = \frac{N-M}{N-M+F}$$

具体的实验数据如表1所示。

表1 实验结果

视频类型	总帧数 T	N(ground num)	F(false num)	M(missed num)	Recall	Precision
Volleyball1	3782	51	4	6	88.24%	91.84%
Volleyball2	918	13	1	2	84.62%	91.67%
Tennis	4111	78	7	11	85.90%	90.54%
Basketball	2158	20	2	3	85%	89.47%
Football	2912	27	3	4	85.19%	88.46%
News1	3250	42	7	5	88.10%	84.09%
News2	1135	14	3	2	85.71%	80%

从表1可见,对于体育实况视频,查全率和查准率的结果都比较令人满意。这是因为体育实况视频中包含有丰富的运动信息,而算法融合了视频的颜色和运动信息,所以得到了较好的实验结果。但是体育实况视频中摄像机和对象的运动变化都比较复杂,故部分漏判和误判还是来源于摄像机和运动员的运动变化,例如摄像机的缩放(zooming)操作等。此外,部分漏判还发生在比赛实况与重播的渐变之间。

对于新闻视频,实验结果也比较好;但是与体育实况视频相比,查准率要稍低一些。这是因为:(1)尽管新闻视频是专业人员摄制,但是因为摄制环境的原因,可能存在一些不平稳的摄像机运动;(2)新闻视频中通常存在闪光等因素的影响。

总的来说,对于误判和漏判的镜头,原因主要源于两类:运动相关和亮度相关。运动相关主要包括摄像机和对象的运动,或者两者的同步发生。而亮度相关指实际环境的亮度的变化,例如闪光的影响等等。

小结 本文提出了一种融合视频颜色和运动特征的镜头探测算法。该方法在视频压缩域内提取视频帧的颜色和运动特征,并且融合两者进行突变和渐变的探测。实验结果表明,该算法快速准确。从实验结果分析来看,运动特征在镜头探测中占据重要地位,需要采用更好的方法获取更精确的运动信息。所以,采用更好的运动估计方法和更加复杂的特性(例如对象区域特征)分析镜头边界是今后研究发展的方向,但是运算势必更加复杂。

参考文献

- 1 Zhang H J, Smoliar S W. Developing Power Tools for Video Indexing and Retrieval, Proc. IS&T/SPIE. On Storage and Retrieval for Image and Video Database II, San Jose, CA, 1994. 140~149
- 2 Bouthemy P, Gelgon M, Ganancia F. A unified approach to shot change detection and camera motion characterization. IEEE Trans. On Circuits and Systems for Video Technology, 1999, 9 (7): 1030~1044
- 3 Meng J, Juan Y, Chang S F. Scene Change Detection in a MPEG Compressed Video Sequence. SPIE Symposium on Electronic Imaging: Science & Technology-Digital Video Compression: Algorithms and Technologies, San Jose, Feb. 1995
- 4 Jun S B, Yoon K, Lee H Y. Dissolve transition detection algorithm using spatio-temporal distribution of mpeg macro-block types. In ACM Multimedia, 2000
- 5 Zhong D. Segmentation, Index and Summarization of Digital Video Content: [PhD thesis]. Columbia University, 2001
- 6 Ngo C W, Zhang H J, Pong T C. Recent Advances in Content Based Video Analysis. IJIG2001
- 7 熊华. 视频内容结构化技术的研究与实现: [博士学位论文]. 长沙: 国防科学技术大学, 2001
- 8 曹莉华. 视频媒体的基于内容处理和检索的研究与实现: [博士学位论文]. 长沙: 国防科学技术大学, 1998