

基于自适应阈值和傅里叶拟合的镜头边界检测*

刘 烽^{1,2}, 郭延文^{1,2+}, 殷昆燕^{1,2}

¹(计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210046)

²(南京大学 计算机科学与技术系, 江苏 南京 210046)

Shot Boundary Detection Using Adaptive Threshold and Fourier Fitting

LIU Feng^{1,2}, GUO Yan-Wen^{1,2+}, YIN Kun-Yan^{1,2}

¹(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210046, China)

²(Department of Computer Science and Technology, Nanjing University, Nanjing 210046, China)

+ Corresponding author: E-mail: ywguo@nju.edu.cn

Liu F, Guo YW, Yin KY. Shot boundary detection using adaptive threshold and Fourier fitting. *Journal of Software*, 2012, 23(Suppl. (2)): 105-114 (in Chinese). <http://www.jos.org.cn/1000-9825/12030.htm>

Abstract: As a key technology in content-based video indexing and management, the video shot boundary detection has attracted considerable research attention. However, the traditional detection algorithm can only handle a hard cut boundary. No satisfactory results have been achieved on gradual transition until now. A method based on the combination of adaptive threshold and Fourier fitting is presented in this paper to detect shot transition. A non-uniform histogram in HSV color space on each frame is first accumulated to calculate a similarity sequence of videos, depending on which thresholds are generated by the adaptive threshold to detect hard cut transition. Gradual transition has a much more complex transform in its span. After extensive experiments and observation, the study has found that fixed patterns exist in every gradual transition. Candidate transitions are detected by finding segments which show traits of a gradual transition pattern. Gradual transition boundaries of different types are collected to train a set of standard templates, which can be used to judge whether a candidate is a real gradual transition boundary and also, to further determine its transition type. The algorithm is accelerated by exploiting the parallel computation power of GPUs using CUDA. The effectiveness is verified through extensive experiments and is compared with other methods.

Key words: shot detection; hard cut detection; gradual transition detection; adaptive threshold; Fourier fitting

摘 要: 作为基于内容的视频检索与管理中的一项关键技术,视频镜头边界检测近年来得到了广泛的研究与应用.但是,传统的视频镜头边界检测算法只能处理切变镜头边界,渐变镜头边界的检测结果则远低于人们的期望.结合自适应阈值与傅里叶函数拟合技术,提出了一种用于检测切变与渐变镜头边界的方法.该方法首先计算视频中每一帧在HSV颜色空间上的非均匀分块直方图,再根据直方图距离得到视频数据的相似度序列,然后在其上使用自适应阈值法计算出相应的阈值,从而检测切变镜头边界.渐变镜头边界处较为复杂,通过大量的实验观察发现,渐变镜

* 基金项目: 国家自然科学基金(61073098, 61021062); 国家重点基础研究发展计划(973)(2010CB327903); 江苏省自然科学基金(BK2009081)

收稿时间: 2012-05-20; 定稿时间: 2012-09-29

头边界处存在着特定的变化模式,并在相似度序列上寻找具有这种变化模式的片段作为候选边界,最后用傅里叶函数拟合形成统一的表示.收集不同类型的渐变边界,训练了一组标准渐变模型,用于判断候选边界是否为渐变边界以及确定其渐变类型.为了提高计算效率,采用通用并行计算架构对算法进行加速.通过大量的实验测试以及与他人方法的对比,验证了算法的有效性.

关键词: 镜头检测;切变边界检测;渐变边界检测;自适应阈值;傅里叶函数拟合

随着多媒体信息技术的迅速发展,大量的视频数据开始涌入人们的日常生活,常见的如新闻、广告、电影等.海量视频数据的出现极大地推动了视频存档、编目以及索引技术的进步.镜头边界检测技术由于能够在视频中自动检测出镜头边界,近年来得到了广泛的研究与应用^[1,2].

镜头是指一组由同一摄像机不间断拍摄连续帧.一段完整的视频通常由多个镜头按照一定的顺序连接而成,镜头之间的连接称作镜头边界,镜头边界有两种最基本的类型,一种是切变边界,是指从一个镜头直接切换到另外一个镜头,中间不存在任何连续性的过渡,相邻两帧之间在内容、颜色上都存在非常明显的差异;另一种是渐变边界,镜头连接处加入了时间或空间上的变换,边界不再如切变类型清晰.渐变边界可以细分为多种类型,常见的有淡入、淡出和消融^[1].画面由亮变暗,直至完全变暗称为淡出;淡入则与此相反,画面由暗变亮,最终达到清晰,淡入与淡出通常会一起使用,使视频画面渐显渐隐.消融,表现为前一个镜头慢慢消失时,后一个镜头已经开始显露,两段画面在一定的时间内叠加在一起.消融常用于前后镜头组接不畅、镜头质量不佳时,避免切换镜头的跳跃.渐变边界这种与视频内容糅杂的特性使得传统的镜头边界检测方法失效,针对渐变镜头设计专有的检测方法因而显得十分必要.

Zhang 等人在文献[3]中提出了镜头边界检测领域中的一个开创性的工作.该方法利用一组相似性度量值描述连续帧之间的内容变化,然后通过双阈值法检测镜头边界.然而,该方法受限于阈值的选择,过大的阈值会产生漏检,而过小的阈值则会引入较多的虚假镜头.通常情况下,没有一组阈值可以适用于所有的视频数据.Ekin 等人^[4]扩展了文献[3]中的方法,但为了节省计算时间,直接在压缩数据上进行检测,利用帧间像素的变化、压缩因子和运动向量来检测镜头切换,正如该文献中所提到的,此方法不能有效识别出渐变的类型.为了突破参数对镜头边界检测产生的限制,Zabih 等人在文献[5]中提出了一种基于图像边缘特征的检测算法.他们认为,在镜头边界处,后一帧的图像边缘会远离其在前一帧中的位置.基于此,该算法利用相邻帧的边缘信息以及统计滑入边缘与当前帧的边缘来进行检测和识别切变类型的边界.渐变边界则通过分析滑入与当前帧的边缘像素的空间分布信息来检测和识别.该算法的缺点在于,边缘检测算法不能处理全局的亮度变化,如全黑或者全白,无法检测边缘信息.另外,该方法的运动补偿算法不能处理多物体快速运动的情况.近年来,Nam 等人^[6]提出通过分析镜头切换背后的规律,然后利用 B-样条插值来分析视频数据,从而检测并识别镜头.但是,由于不能准确地定位出镜头边界的开始位置与结束位置,且淡入、淡出检测内在地依赖于它之前的渐变镜头边界的检测,该方法的检测结果的准确性受到较大的影响.Efthymia 等人^[7]利用视频中帧间颜色的连续性计算视频帧在时间域上的多种特征,每种特征都计算出其特征向量用于 SVM 分类器分类,从而识别镜头边界,但是,SVM 分类器的训练需要耗费较多的计算资源,且由于淡入、淡出特效与消融特效的同质现象,它们的特征向量比较相似,使得 SVM 分类器不能准确地区分这两种渐变.

直方图法检测切变边界虽然已经取得较好的效果^[8,9],但是这类方法通常都会受到阈值选取问题的困扰,过大或过小的阈值都会影响检测的精确度和召回率,目前的切变边界检测希望能够摆脱阈值设置的烦恼,自动完成检测^[10];渐变镜头边界在经过多年的研究以后,其检测结果仍然不能达到人们的预期,至今仍面临两个主要问题,一是如何精确定位渐变边界,二是准确地识别渐变边界的类型.

为了解决直方图法的阈值问题以及渐变边界的定位与识别问题,本文在直方图法的基础上将其改进为 HSV 颜色空间中的非均匀分块直方图来计算视频的相似度序列,然后使用自适应阈值技术计算出跟相似度序列相匹配的阈值来检测切变边界,切变边界检测在本文中不需要设置任何参数且能取得很高的检测精确度.渐变边界相对于切变边界显得更为复杂,是在两个镜头的相邻段上附加特定的时-空变化效果而产生.渐变边界

处,相邻帧之间内容上的变化由原本视频内容的变化和编辑特效共同作用而成,不同的特效会使帧之间的内容发生不同形式的变化.基于此观察,本文发现渐变边界处帧内容变化有一定的规律可循,这种特殊的规律即是渐变特征.以此特征为基础,我们预先训练提取出渐变边界的特征模式,存在模板库中,将检测得到的候选边界与模板库中的标准渐变模型进行比较,从而确定渐变类型.我们为视频中每一帧计算 HSV 颜色空间中的非均匀分块直方图,相邻帧间的相似度可以通过计算对应块直方图距离的加权和得到.通过计算视频中连续帧之间的相似度可以得到视频的一个中间表示,即相似度序列.相似度序列在很大程度上保持了视频原有的变化特性,且提高了计算效率.渐变边界检测的第 1 步是确定候选边界,我们利用滑动窗口在相似度序列上寻找具有渐变特征的片段.渐变边界检测的第 2 步是采用傅里叶函数拟合对候选边界生成统一的表示,然后与模板库中的标准渐变模型进行对比,以判断是否为真正的渐变边界和确定准确的类型.最后,本文利用图形处理芯片的通用并行架构进行算法加速,在 1 080p 的高清视频上可以达到 20fps 的处理速度.

本文工作的主要贡献是,首先在切变边界检测方面提出非均匀分块直方图结合自适应阈值以自动检测切变镜头边界,不需要设置任何参数;其次,提出一种新的描述渐变边界特征的方法,并利用预先训练的标准渐变模型进行渐变边界检测,在精确度与召回率上,本文算法相对于现有算法具有较大的优势;最后利用图形芯片组的并行计算能力进行算法加速,取得了较高的计算效率.

本文第 1 节介绍我们提出的视频序列中间表示方法.第 2 节详细论述切变镜头边界与渐变镜头边界的检测方法.第 3 节进行实验并与相关工作进行对比.第 4 节对本文工作进行总结并展望未来工作.

1 视频序列的中间表示

视频数据,尤其是正日益普及的高清视频的一个重要特点是内容丰富,数据量大.为了直观且快速地进行镜头边界检测,本文首先生成视频数据的中间表示.传统直方图表示具有准确、高效的特点,但其丢失了像素点的位置信息,基于此,本文采用改进的 HSV 颜色空间中非均匀分块直方图.本节将详细介绍如何生成视频数据的中间表示以及颜色空间选取、直方图计算、相似度衡量等工作.

颜色空间选取.目前,电视电影视频、体育娱乐视频都已摆脱传统的黑白画面,取而代之的是彩色制式.在镜头边界处,人眼可以感受到视频帧之间在颜色空间上强烈的不连续性.在镜头边界检测中如果能够引入一种与人眼感知类似的描述量来表示视频帧,则能在较大程度上提高检测的成功率与正确性.本文通过实验验证了 HSV 颜色空间较为符合人类色彩感知^[1],可以用作检测的颜色空间.在 HSV 颜色空间中,色调 H 、饱和度 S 和亮度 V 这 3 个分量相互独立,这种特性可使连续变化的镜头内的帧间相似度更大,而镜头间的帧间相似度更小,有利于发现镜头边界.

非均匀分块直方图计算.在视频边界检测中,用直方图表示原始视频数据进行计算比像素比较法和基于特征的算法高效,另外,直方图法可以减弱摄像机运动和物体移动对检测结果的影响.但是,传统的全局直方图不能记录像素点位置,难以反映视频的空间信息,因此本文采用改进的分块直方图.另外,考虑到人类的视觉关注点集中在视频的中间部分,而边缘处在一定程度上则被忽略^[4].为了模拟人类的这种视觉特性,本文以 3:5:3 的黄金分割比率非均匀地对视频帧划分区域(如图 1 所示),在每个区域上分别计算局部颜色直方图特征,整个视频的颜色直方图特征则可以由局部颜色直方图组合而成,表示为 $\{h_k(f)\}_{k=1}^9$.

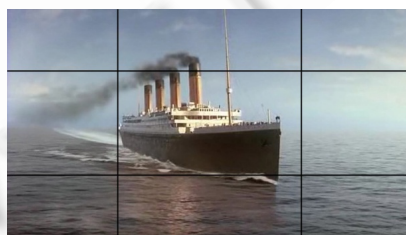


图 1 非均匀视频帧分块

相似度度量.本文定义了一个相似度值用于衡量视频帧之间的内容变化剧烈程度.对于帧 f_i 和帧 f_{i+1} ,其相似度可以通过计算对应块直方图距离的加权和得到,

$$d_i = \sum_{k=1}^9 w_k \cdot \text{dis}(h_k(f_i), h_k(f_{i+1})) \quad (1)$$

其中, $h_k(f_i)$ 表示帧 f_i 中第 k 块上的HSV空间颜色直方图, $\text{dis}(\cdot, \cdot)$ 表示相邻视频帧中对应块之间的距离. w_k 表示赋予第 k 块的权值,用于标定视频帧中各个区域的重要性.在图1中,最上方3个区域以及最下方3个区域的权值设定为1/14,两侧的 w_4 和 w_6 则设为1/7,最后中间区域的权值 w_5 设置为2/7.

相似度序列.通过计算视频中所有连续帧之间的相似度,我们得到一组相似度的序列,即视频数据的中间表示.这组相似度序列保持原有视频内容变化的特性,并且使得检测不必直接在原始数据上进行,极大地提高了检测算法的效率.对于一段长度为 n 的视频,其相似度序列如下:

$$\Omega = \{d_1, d_2, \dots, d_{n-1}\} \quad (2)$$

视频内容的剧烈变化会使相似度序列产生强烈的局部波动,即 Ω 上对应段会有略微高出周围区域的波峰出现,但是这些波峰的能量通常又小于镜头边界的能量.视频相似度序列剧烈波动时如果仍然选用较小的自适应阈值则会引入较多的虚假镜头.为了解决这一问题,本文在 Ω 上采用长度为 $2 \cdot \sigma$ 的一元高斯函数进行滤波得到平滑序列 Ω' .

$$\Omega' = \Omega \cdot \exp(-x^2/2 \cdot \sigma^2), x \in (-\sigma, \sigma) \quad (3)$$

2 镜头边界检测

镜头边界检测按照镜头边界类型的不同可以划分为切变镜头边界检测和渐变镜头边界检测.传统的基于颜色直方图表示并采用固定全局阈值的切变边界检测方法,其检测的精度受限于阈值的设定,无法对所有的视频数据提供统一的阈值,为了解决此问题,本文设计了自动阈值的选取方法,使得切变边界检测实现自动化,无须用户干预.渐变镜头在经过多年的研究之后,仍然没有一个完善的解决方案,研究者们面对的两个主要难题是准确定位渐变的边界以及识别出渐变的类型.本文通过观察大量的渐变边界,发现在渐变边界处存在特定的变化规律,即渐变特征.我们收集了大量渐变边界,并以此为蓝本为淡入、淡出特效训练出3组标准特征模板,为消融特效训练出一组标准特征模板.而后再从视频的相似度序列上寻找具有渐变特征的片段,将其选作候选边界.所有的候选边界都会用傅里叶函数进行拟合,生成统一表示.然后和模板库中的标准渐变模型进行对比从而判定其是否为渐变边界以及确定边界类型.

2.1 切变镜头边界检测

切变边界的特点是镜头与镜头之间没有任何过渡帧.检测切变边界的主要理论依据是镜头间的帧间相似度远低于镜头内的帧间相似度.通常,检测切变边界的方法是设定阈值 T ,然后选取相似度序列中达到阈值的位置作为切变边界的位置.

$$hc(\Omega') = \bigcup_{i=1}^l (\max(0, \text{sig}(d_i - T)) \cdot i) \quad (4)$$

其中, l 是滤波后相似度序列的长度, hc 表示切变边界的集合, $\text{sig}()$ 是用于返回0或1的信号函数.公式(4)中,阈值 T 的选择直接关系到检测结果的好坏,不同分辨率和不同类型的视频所应设定的阈值相差较大,如果不加区分地使用固定阈值进行检测,则会产生远低于人们预期的结果.本文采用自适应阈值技术,基于相似度序列计算出相应的检测阈值.图2(a)表示了一段相似度序列,为了将其中的切变边界分离出来,本文对此相似度序列进行了频率统计,如图2(b)所示,可以明显看到两个波峰,分别对应非边界与边界区域的相似度分布,至此,寻找切变边界阈值的问题则转化为如何准确地分割两个波峰.本文利用Gonzalez在文献[11]中提出的基本门限法寻找合适的分割位置,即利用初始阈值做初步分割,对分割出的两组数据分别计算均值,两处平均值再做平均得到新的阈值,之后再利用新的阈值进行迭代,直到阈值开始呈现收敛趋势.基于自适应阈值的切变边界检测过程如算法1所示.

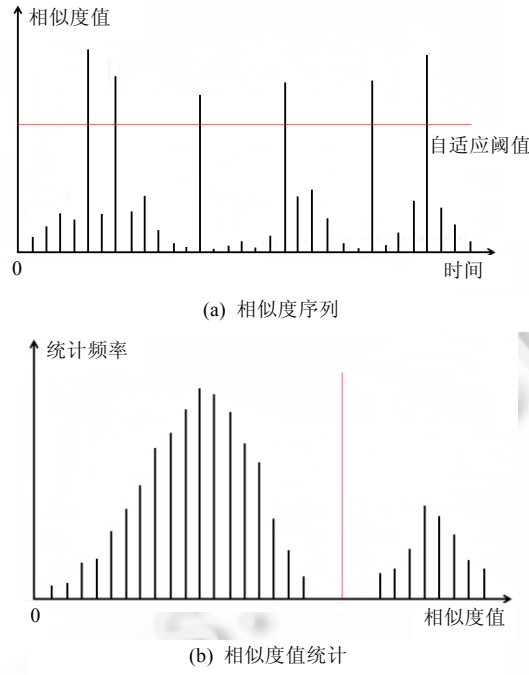


图2 自适应阈值选取

算法 1. 基于自适应阈值的切变边界检测.

Initialization:

Set $\varepsilon=0.1$; $T=0$; $T_r=\varepsilon$; $su=0$; $sb=0$; $hc=\{\emptyset\}$; $\Omega'=\{d_1, d_2, \dots, d_i\}$;

Iteration:

```

1: while  $|T-T_r|>\varepsilon$  do
2:    $i\leftarrow 0$ ;  $ncu\leftarrow 0$ ;  $ncb\leftarrow 0$ ;  $su\leftarrow 0$ ;  $sb\leftarrow 0$ ;
3:   while  $i<\text{length of similarity sequence}$  do
4:     if  $\Omega'(i)>T$  then
5:        $su\leftarrow su+\Omega'(i)$ ;  $ncu\leftarrow ncu+1$ ;
6:     else
7:        $sb\leftarrow sb+\Omega'(i)$ ;  $ncb\leftarrow ncb+1$ ;
8:     end if
9:   end while
10:   $T_r\leftarrow T$ ;
11:   $T\leftarrow (su/ncu+sb/ncb)/2$ ;
12: end while
13:  $j\leftarrow 0$ ;
14: While  $j<\text{length of similarity sequence}$  do
15:   if  $\Omega'(j)>T$  then
16:      $hc=hc\cup\{j\}$ 
17:   end if
18: end while

```

2.2 渐变镜头边界检测

在本文的检测算法中,渐变镜头边界检测分两阶段完成.第1阶段通过在相似度序列上寻找具有渐变特征的片段以确定候选边界.通过实验观察发现,渐变边界处视频帧之间的相似度值会发生剧烈波动,表现出波峰形状,候选边界检测则可通过在相似度序列上寻找特定的波峰来完成.第2阶段是渐变边界认证过程,通过与模板库中的标准渐变模型进行对比可以准确地筛选出渐变边界并确定其变化类型,本文渐变镜头边界检测主要集中在淡入、淡出与消融特效.

图3展示了淡出特效,视频帧的亮度由暗转亮;淡入特效,视频帧的亮度由亮转暗;消融特效,前后视频镜头中的帧产生重叠,并且其中一帧在强度上减弱,另一帧加强.



图3 淡出特效、淡入特效以及消融特效

2.2.1 候选边界检测

渐变效果通常是视频在后期编辑时由视频编辑工具生成,不同的工具对渐变效果的实现方法并不相同,Lu等人在文献[12]中利用淡入、淡出特效中间区域相似度值为0来检测这种特效.但是,有些编辑工具并不会使淡入、淡出特效中间区域的帧亮度完全降为黑色,因而,这种特征并不稳定.鉴于此,我们需要找到一种不受编辑工具影响的渐变特征.我们首先对 Hampapur 等人在文献[13]中提出的渐变模型做分析.渐变边界模型如公式(5)所示.

$$f(t) = \alpha(t) \cdot f_1(t) + \beta(t) \cdot f_2(t), 0 < t < T \quad (5)$$

其中, $f(t)$ 为视频片段 $f_1(t)$ 和 $f_2(t)$ 通过控制函数混合而成的视频渐变区域, $\alpha(t)$ 与 $\beta(t)$ 均为控制函数, T 是时间域上渐变镜头边界的长度.在此模型中,如果 $f_1(t)$ 或 $f_2(t)$ 有一个是纯色的片段,那么淡入、淡出特效可被看作消融特效的一个特例,消融特效与淡入、淡出特效同质的这种特性使得对它们的检测可以在统一的框架下进行.

渐变特征.图4展示了控制函数 $\alpha(t)$ 与 $\beta(t)$.可以看到 $\alpha(t)$ 在时间域上递减,它的作用是将片段 $f_1(t)$ 的亮度从正常转换为黑色,而 $\beta(t)$ 则是增函数,它的作用是将 $f_2(t)$ 的亮度从黑色转换成正常.控制函数导数的模先从最小增到极值,然后再降到最小.这个过程在相似度序列上体现为控制函数刚作用到视频片段时,相邻帧的相似度值的差异较小,而当控制函数达到最快变化时差异也相应地最大,最后,控制函数的变化趋于平缓,差异也随着而变小.图5展示了对淡入、淡出边界的拟合结果,其中,原始数据是指帧间相似度,连续的相似度值则形成相似度序列,对该序列使用傅里叶函数进行拟合可以得到一条光滑的拟合曲线,即傅里叶拟合曲线.可以看到淡入、淡出边界的傅里叶拟合曲线呈现双峰波形,消融作为与淡入、淡出同质的特效,呈现出单峰波形,这则是渐变边界的渐变特征.在相似度序列上寻找具有渐变特征的片段,即可找到候选边界,当控制函数为线性时,渐变边界的这种特性仍然成立,所不同的是控制函数开始作用和结束作用区域的明暗变化将会更加明显.值得注意的是,视频内物体快速移动的时间区域也会具有这种渐变特征,但这些区域很容易在认证过程中排除.

候选边界.本文提出了一种基于逆序对计数的检测候选边界的算法,逆序对和顺序对的定义如下:

定义 1(逆序对与顺序对). 对于一个包含 N 个非负整数的数组 $A[1,2,\dots,n]$,如果有 $i < j$,且 $A[i] > A[j]$,则称 $(A[i],A[j])$ 为数组 A 中的一个逆序对;反之,如果有 $i < j$,且 $A[i] < A[j]$,则称 $(A[i],A[j])$ 为数组 A 中的一个顺序对.

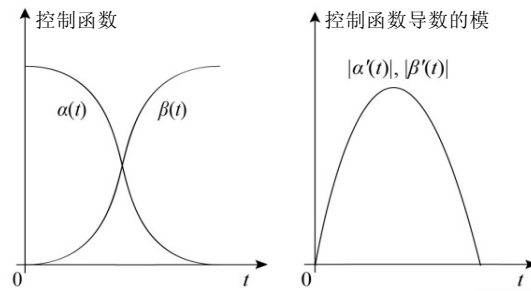


图4 渐变控制函数

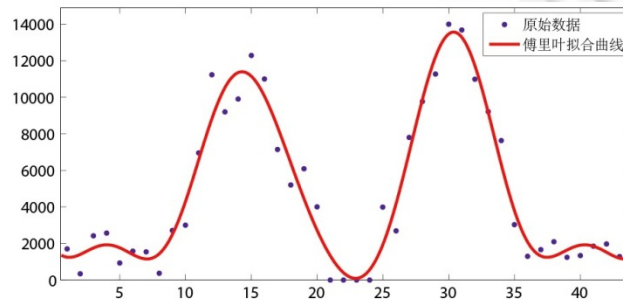


图5 傅里叶函数拟合淡入、淡出特效

确定一个候选边界需要在相似度序列 Ω' 上找到相邻的一对上行段和下行段,其中间的区域即为候选边界.我们使用长度为 W 滑动窗口在相似度序列上滑动,然后可以得到一组局部的相似度序列,

$$U_i = \{d_i, d_{i+1}, \dots, d_{i+W-1}\} \quad (6)$$

让 n_i^r 和 n_i^o 分别表示 U_i 中逆序对和顺序对的个数.如果 $\mu \cdot n_i^o < n_i^r$,则 U_i 为下行段,否则,如果 $\mu \cdot n_i^r < n_i^o$,则 U_i 为上行段.本文通过大量的实验测试发现,当 μ 设置为5时,可获得最佳检测结果.为了排除不可能是候选边界的区域,我们需要剔除变化较为缓慢的上行段和下行段,计算每组数据的方差,如果方差较小则直接作无效处理.一个区域如果由连续的上行段和下行段包围则可被认为是候选边界.

2.2.2 候选边界识别

淡入与淡出通常结合在一起使用,相似度序列上表示为连续出现的两个波峰,消融特效则表现为单个波峰.我们可以根据这个观察来实现鉴定渐变边界的类型.直接使用候选边界的原始数据无法进行比较,本文采用傅里叶函数对 U_i 进行拟合,得到一条光滑曲线,不同的渐变边界则有不同类型的曲线.曲线在进行标准化后与模板库中的标准渐变模型进行比较,从而确定渐变类型.

傅里叶函数拟合.可以用来拟合曲线的函数有很多,选取合适的拟合曲线对渐变边界检测起着关键的作用.我们首先尝试了多项式拟合,但是,要想获得比较满意的拟合结果则需要对每个候选边界拟合时调整多项式的阶,即无法使用同一个多项式拟合所有的候选边界且能取得令人满意的拟合效果.B-样条曲线拟合时需要对原始数据进行切割分段,然后将拟合结果综合起来,由于其对拟合结果无法提供一个统一的表达式,B-样条曲线拟合方案也被排除在外.傅里叶函数保留了多项式和B-样条的优点,能够精确拟合并提供统一的拟合结果表达式,最终,本文选定傅里叶函数作为候选边界的拟合函数.傅里叶函数是一系列正弦函数与余弦函数的组合,其表达式如下:

$$f(t) = a_0 + \sum_{i=1}^4 (a_i \cos(i \cdot w \cdot t) + b_i \sin(i \cdot w \cdot t)) \quad (7)$$

其中, a_0, a_i, b_i 是傅里叶函数的参数,拟合不同的渐变边界会得到不同的参数, w 是数据的频率信息.我们通过实验发现,傅里叶函数可以很准确地拟合出数据的走势与形状,拟合得到的结果会进一步做标准化以用于比较.

标准渐变模型.本文从新闻广告、电视电影、纪录片以及音乐会等视频中采集了大量的渐变边界.不同视频数据上的渐变边界千差万别,为了剔除不相关因素的影响,首先用傅里叶函数拟合渐变镜头边界,得到统一的光滑曲线表示,然后对同一类别渐变边界的曲线模型进行叠加平均.根据波形的形状,淡入、淡出特效的渐变模型有3种,如图6(a)~图6(c)所示.消融特效的渐变模型在标准化之后只有1种,如图6(d)所示.

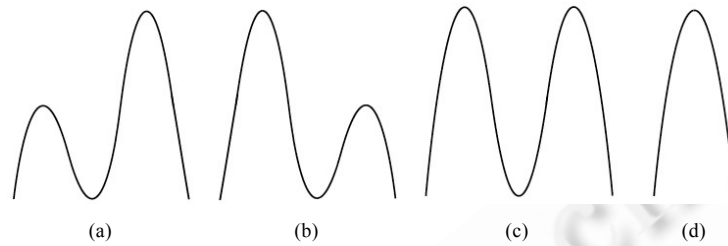


图6 淡入、淡出与消融的标准渐变模型

由于渐变边界的长度并不一致,对其进行拟合得到拟合曲线的长度也不一样.为了训练出统一的模板,本文首先对拟合曲线使用最近邻插值法进行插值,使所有的渐变模型的长度均达到 L ,然后对齐拟合曲线.用 P_1, P_2 和 P_3 分别表示渐变模型的两个波峰和一个波谷.

$$R(T_1, T_2) = \sum_{i=0}^2 |P_i - P'_i| \quad (8)$$

在坐标轴上从左往右移动新的拟合曲线,直到 $R(T_1, T_2)$ 达到最小,则认为曲线对齐.重复此过程,直到所有训练样本都对齐,然后将训练样本的幅值取平均,即可得到标准渐变模型.消融边界的标准模型亦可通过同样的方法训练得到.

类型识别.在相似度序列上通过寻找渐变特征可以检测到候选边界,但是仍无法确认其是否是真正的镜头边界及其准确的边界类型.本文通过对比较候选边界与标准渐变模型来识别渐变边界,我们对候选边界进行与标准渐变模型一样的前期处理,利用傅里叶函数拟合及标准化.至此,我们可以得到候选边界标准化之后的傅里叶函数表示,在与标准渐变模型校准之后,对其差值绝对值进行积分则可以得到候选边界与标准渐变模型的差异程度.

$$Diff(f(t), f_s(t)) = \int_0^T |f(t) - f_s(t)| dt \quad (9)$$

其中 $f(t)$ 和 $f_s(t)$ 分别是候选边界拟合得到的傅里叶函数与标准模板库中的边界模型, T 为候选边界时间域上的长度.如果差异程度不超过0.17,则认为匹配成功,当多个标准渐变模型匹配成功时,选择差异程度最小的作为匹配边界.

3 实验结果

本节选取了包括新闻、纪录片、音乐会、电影以及体育视频等各种不同类型的数据来进行测试,并且与文献[6]的实验结果进行对比以验证本文算法的有效性.

数据集.本文在不同类型以及不同长度的数据集上测试算法的性能.视频数据集包括电视新闻、文献纪录片、音乐会录像(《Tears in Heaven》)、2部完整电影(《Titanic》与《Star Wars: Episode 1 The Phantom Menace》)、1段羽毛球比赛视频,以及6段从电影数据中截取的切变片段,电影数据为1080p高清格式.

电视新闻中存在滚动字幕、电视台标和直播重播提示,由于本文采用了非均匀分块直方图,字幕所在区域分配的权值较小,使得其对检测结果影响甚微,但新闻中各种不同的特效和丰富的报道内容对于镜头边界检测则是相当大的一个挑战.文献纪录片与音乐会片段由于其内容变化较为缓慢,用编辑工具生成变化特效时,控制函数起主导作用,渐变边界的渐变特征较为明显,易于检测.本文选取的羽毛球比赛视频中,比赛中视频相对静止,运动员虽有剧烈动作,但对帧间相似度影响不大;而在比赛休息阶段的视频,则包含很多对运动员的特写,运动员的动作会对帧间相似度产生较大的影响.电影视频中以切变为主,辅以少量的淡入、淡出特效.表1展示了

渐变镜头与切变镜头的统计对照数据.其中切变边界有专门的数据集,故未在其他类型的数据中统计切变边界的数目.

表 1 视频镜头边界统计

		视频					总计	
		电视新闻	文献纪录片	音乐会	体育视频	电影(完整)		电影(片段)
边界	消融	134	93	43	27	38	0	335
	淡入、淡出	33	34	0	18	5	0	90
	切变	0	0	0	0	0	258	258
总计		167	127	43	45	43	258	-

评价标准. 镜头边界检测的通用评价标准是召回率与精确度,两者定义如下:

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

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

其中, N_c 表示正确检测出的镜头, N_m 表示漏检的镜头数, N_f 表示被错误识别成镜头边界的假边界数.

结果分析与对比. 表 2 中展示了本文切变镜头边界检测的实验结果.本文在 4 部电影中分别选取 4 个片段,共 258 个镜头,对于每个视频片段按照本文的自适应阈值计算阈值进行检测,召回率达到 97.7%,精确度达到 98.4%,高于当前切变检测的平均水平.表 3 展示了渐变边界的实验结果,由于音乐会视频的内容变化较为缓慢,镜头边界处很清晰,因而本文算法在此数据集上表现最为出色,而电影视频中,由于其部分内容较为剧烈,如汽车追逐、爆炸等场景等,本文算法在电影视频上引入了部分虚假镜头边界.为了进一步评价本文的工作,表 4 中给出了本文方法与文献[6]中方法的结果比较.在测试使用的数据集上,本文方法的召回率均超出文献[6]中的方法,对于视频分析和管理工作,丢失的镜头边界需要浏览整段视频重新寻找,所以引入虚假镜头所产生的代价要远小于丢失正确的镜头,本文方法在此数据集上的召回率达 88.9%,超过其他文章中的方法.

表 2 切变边界检测结果

视频类型	实际边界	正确检测	漏检	错误检测	召回率(%)	精确度(%)
电影片段	258	252	6	4	97.7	98.4

表 3 本文算法的检测结果

视频类型	实际边界		正确检测	漏检	错误检测
	消融	淡入、淡出			
电视新闻	134	33	149	18	42
文献纪录片	93	34	112	15	13
音乐会	43	0	40	3	5
电影	38	5	37	6	20
体育视频	27	18	39	6	17
总计	335	90	377	48	97

表 4 与文献[6]中方法的对比

视频类型	召回率(%)		精确度(%)	
	本文方法	文献[6]中方法	本文方法	文献[6]中方法
电视新闻	89.2	87.4	78.0	79.8
文献纪录片	88.2	86.6	89.6	90.9
音乐会	93.0	90.7	88.9	86.7
电影	86.0	81.4	64.9	56.5
总计	88.9	86.8	80.8	80.3

算法加速. 本文算法中最耗时的部分是生成视频数据的相似度序列,尤其对于高清视频,如 1 080p 的电影.幸运的是,我们计算相似度序列时采用了通用并行计算架构来加速此计算过程,在 GPUs 的流处理器组上生成 12 288 个线程用于统计直方图,每帧统计时间只需要 27ms.本算法在配备 Intel Pentium Dual-Core 2.7GHz 处理器和 NVIDIA GTX580 显卡的计算机上,能够以 20fps 的速度处理 1 080p 高清电影视频,而处理较低分辨率如

720p 视频的速度则可以超过 25fps.

4 总结与展望

本文提出了一种新的用于切变与渐变镜头边界检测的方法——自适应阈值方法,解决了直方图方法中选定阈值的困难,可以直接根据相似度序列自身计算出与其最符合的阈值.对于渐变边界,我们的方法通过分析其变化特征,应用傅里叶函数拟合得到统一表示,然后与模板库中的标准渐变模型做对比,从而完成检测和识别的任务.本文中的算法目前只能检测切变边界和淡入、淡出、消融这两种渐变边界,在未来,我们还将尝试检测并识别 WIPE 类型及其他不规则的渐变边界.

References:

- [1] Mas J, Fernandez G. Video shot boundary detection based on color histogram. In: Proc. of the TRECVID Workshop 2003. 2003.
- [2] Smeaton AF, Over P, Doherty AR. Video shot boundary detection: Seven years of TRECVID activity. Computer Vision and Image Understanding, 2010,411-418.
- [3] Zhang H, Kankanalli A, Smolia SW. Automatic partitioning of full-motion video. Multimedia Systems, 1993,1(1):10-28.
- [4] Ekin A, Tekalp AM, Mehrotra R. Automatic soccer video analysis and summarization. IEEE Trans. on Image Processing, 2003, 12(7):796-807.
- [5] Zabih R, Miller J, Mai K. A feature-based algorithm for detecting and classification production effects. Multimedia Systems, 1999, 7(1):119-128.
- [6] Nam J, Tewfik AH. Detection of gradual transitions in video sequences using B-spline interpolation. IEEE Trans. on Multimedia, 2005,7(4):667-679.
- [7] Tsamoura E, Mezaris V, Kompatsiaris I. Gradual transition detection using color coherence and other criteria in a video shot meta-segmentation framework. In: Proc. of the ICIP_MIR. 2008. 45-48.
- [8] Seidl M, Zeppelzauer M, Breiteneder C. A study of gradual transition detection in historic film material. Journal on Computing and Cultural Heritage, 2010,10-20.
- [9] Zeppelzauer M, Mitrovic D, Breiteneder C. Analysis of historical artistic documentaries. In: Proc. of the Int'l Workshop on Image Analysis for Multimedia Interactive Services. 2008. 201-206.
- [10] Chan C. Shot boundary detection using genetic algorithm optimization. In: Proc. of the Int'l Symp. on Multimedia. 2011. 327-332.
- [11] Gonzalez RC, Woods RE, Eddins SL. Digital Image Processing. Electronic Industry Press, 2008. 482-495.
- [12] Lu HB, Zhang YJ, Yao YR. Robust gradual scene change detection. In: Proc. of the Int'l Conf. on Image Processing. 1997. 304-308.
- [13] Hampapur A, Jain R, Weymouth T. Digital video segmentation. In: Proc. of the 2nd ACM Int'l Conf. on Multimedia (MULTIMEDIA'94). 1994. 357-364.



刘烽(1988—),男,江苏南京人,硕士生,主要研究领域为图像与视频分析处理.



殷昆燕(1987—),女,硕士生,主要研究领域为图像处理与恢复.



郭延文(1980—),男,博士,副教授,主要研究领域为计算机图形学,图像与视频处理.