

一种分层的和多分辨率的镜头边界检测方法^{*})

韩冰 高新波 姬红兵

(西安电子科技大学电子工程学院 西安 710071)

摘要 提出了一种分层的和多分辨率的镜头边界检测方法。该方法对各种不同的镜头间过渡类型给出了用不同方法进行联合检测的方案,该方案主要分为突变镜头检测(即视频切分)、淡化过渡检测、溶解过渡检测及划变过渡检测4个部分。检测方案并不是简单地将各种方法拼接在一起,而是通过小波变换的多分辨率分析将它们有机地结合起来,相互关联,达到有效检测结果。首先用FCM聚类算法进行视频切分,然后根据聚类结果分别在整数小波分解后的高频部分用 Gaussian 加权 Hausdorff 距离结合边界改变率算法检测淡化过渡;对分解后的低频部分用所提出的 SCD 算法(Similarity of color distribution based method)检测溶解过渡,并通过自适应调节权系数(系数自适应)使检测相异度函数更能适用于多种视频片段。最后根据切分以及前面两种过渡检测的结果,利用三维小波分解后高频成分中的运动部分所定义的运动矢量来检测划变过渡。用实际视频数据所做的仿真实验结果表明,该方法不但能同时检测突变过渡和渐变过渡,而且能准确地判断渐变过渡的类型及其位置。此外,还能有效地抑制闪光、运动等的影响,从而提高了检测精度。

关键词 镜头边界检测,整数小波变换,三维小波变换,FCM 聚类,自适应系数,Gaussian 加权 Hausdorff 距离,颜色分布,运动矢量

A Hierarchical and Multiresolution Detection Method of Shot Transition for Video Shots Segmentation

HAN Bing GAO Xin-Bo JI Hong-Bing

(School of Electronic Engineering, Xidian Univ., Xi'an 710071)

Abstract According to the drawbacks of available algorithms, a new hierarchical and multiresolution approach to the detection and classification of scene breaks in video sequences is presented in this paper. This method gives the detection scheme for different shots using different algorithms. This scheme includes four parts: cut detection(video cut), fade detection, dissolve detection and wipe detection. Firstly, the video clips are cut by FCM clustering method, then the fade and dissolve are detected using Hausdorff distance and SCD algorithm in the high components and the low components respectively by integer-to-integer wavelet transform. Finally, according to the cut and the former detection of gradual changes, we use a motion vector in high-components of 3D-WT to detect the wipe transition. Experimental results with real video clips demonstrate that our method can detect and classify a variety of scene breaks, including cuts, fades and dissolves, even in sequences involving significant motions and flash.

Keywords Shot boundary detection, Integer-to-integer wavelet transformation, 3D-WT, Fuzzy c-means clustering, Automatic coefficient, Gaussian weighted hausdorff distance, Distribution of color, Motion vector

1 引言

多媒体和视频压缩技术的发展及广泛应用导致大量视频信息的出现。如何高效、方便、快捷地对视频进行分类、管理和检索,现已成为一个研究的热门问题。在各种视频分析方法中,首要任务是视频分段,即镜头分割。镜头分割的正确与否将直接影响到后续分析步骤(如关键帧提取、场景分析、镜头运动分析等)的准确性。镜头是根据镜头间的过渡而划分的。

镜头间的过渡可分为突变过渡(Cut)和淡入(Fadein)、淡出(Fadeout)、溶解(Dissolve)及划变(Wipe)等渐变过渡。目前,针对切变检测已有很多有效的检测手段,如利用对应像素差异、直方图比较、边界变化以及统计运动信息等方法^[1]。这些方法大都需要预先设定检测门限,从而增加了人为干预造成的影响,使得这些方法不适用于有大量变化的视频。至于

渐变过渡,由于镜头边界不明显,使得渐变镜头的切分成为研究的难点,人们在这方面所做的研究工作相对较少。

H. J. Zhang 等人提出的双门限检测的方法^[2]是一种常用的方法,但阈值难以选择,且阈值本身严重影响了算法的性能。模板匹配法^[3]以两帧对应像素差的绝对值之和作为帧间差,其缺点是对噪声和镜头或物体运动非常敏感。Hampapur^[4]使用了相邻帧之间相对亮度的变化来检测淡入、淡出,但是这种方法不适用于检测溶解或划变过渡。Alattar^[5]利用视频帧统计特性的变化规律来检测划变。该方法首先假设在划变过程中,两个相邻镜头画面区域的均值始终不变。这个假设在大多数视频中不能成立,因为局部与整体的均值一般是不相等的。

另外,现有的方法只是针对以上几种镜头过渡中的某种较为有效,目前还没有一种通用有效的镜头检测方法。当然,有研究者也提出了将不同的方法进行简单的组合,以检测不

^{*} 国家自然科学基金(No. 60202004)、教育部重点项目(No. 104173)资助。

同镜头的方案^[6],但是这并没有形成一个完整的检测系统。再者,由于视频数据是海量的,计算机不可能一次处理很大的数据量。基于此,本文利用小波变换的多分辨特性,针对各种不同的镜头过渡模式提出了相应的检测算法,构建了分层的和多分辨的镜头过渡检测方案,使得各种检测方法能够有机地结合起来。

2 分层的和多分辨的镜头过渡检测

不同的镜头过渡在特征和编辑检测的难易程度方面都不一样,因此很难用固定和单一的检测特征和算法去刻画所有的镜头过渡类型。一段视频往往包括很多的突变、溶解、淡化等过渡,这就相当于一个大问题由很多小问题组成一样,因而自然想到要分层的、分步骤地解决这些“小问题”。对于不同类型的镜头过渡,我们采用不同的检测手段,以达到更高的检

测效率。再者,从人类视觉角度分析,不同分辨率下的图像有不同的特征及视觉效果,因而将图像分解到不同分辨率下进行检测是合理的。而小波变换正是一个这样的工具,它能提供给我们所要提取的特征参数,而且能快速有效地计算。为了提高运算速度,本文采用整数小波。将原始视频序列经过整数小波变换后得到不同的子带。在第一层上用 FCM 聚类方法对低频子带上的图像序列进行突变镜头检测;在第二层上对两个突变帧之间的视频片断进行分析,对帧数大于 20 的视频片断,认为其可能存在渐变镜头过渡。在高频子带上,用 Gaussian 加权 Hausdorff 距离来检测淡化。同时,在低频部分用 SCD 算法对溶解过渡进行判断。最后,利用三维小波和矢量的运动特性在高频部分对划变过渡进行确定。分层检测方案如图 1 所示。

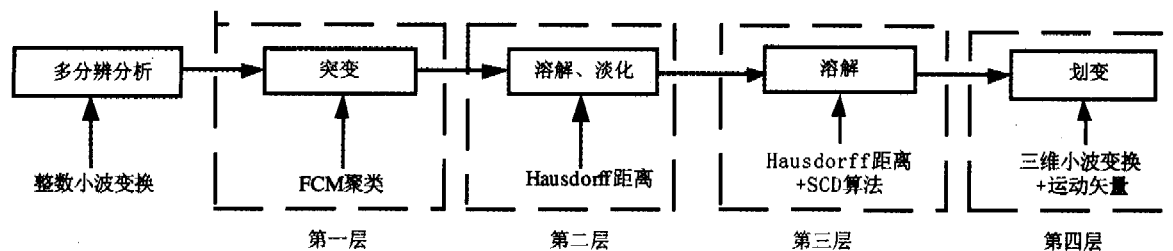


图1 分层的视频分割方案

2.1 整数小波变换

与第一代小波相比,整数小波是位内运算,且输入输出都是整数,因而运算复杂度较低^[7,8];且由于不需要对变换系数进行量化,因此可实现整数加和移位运算的快速算法。无疑,用它来实现镜头检测算法也具有不引入量化误差等优点。此外,低频图像的平滑性可在一定程度上补偿物体和摄像机的小运动带来的影响。对于高频部分主要包含图像的噪声与细节信息,我们正好可以利用细节部分进行边界提取,并结合 Gaussian 加权 Hausdorff 距离来衡量帧间变化。以上分析表明,利用小波变换将图像分解成高频和低频部分,再从各部分中提取不同的特征对不同类型的镜头过渡进行检测是行之有效的。

2.2 突变镜头检测

由于视频数据是海量的,文[9]中为了节省计算机内存,提出了对视频进行分段处理的思想,每段最多 400 帧。这种不根据视频内容任意地划分容易将一段渐变过渡划分开,而造成漏检和误检。而本文采用先切分后检测的思想,即先利用突变镜头将视频切分成若干片段。这种利用突变镜头进行切分的方法不会将可能的渐变区间划分开,从而在大大降低计算机的存储量的同时,提高了检测精度。本文采用 FCM 聚类算法检测视频序列中的突变镜头,从而对整个视频序列进行切分。选取的聚类特征为:空间差测度(SDM: Spatial difference metric)^[10]和量化直方图差测度(QHDM: Quantized histogram difference metric)。由于空间差测度(SDM)并不考虑摄像机的运动,因此它在检测镜头边界的时候会造或误检。此外,视频内容的变化是多样的,仅用一个特征无法准确表征视频内容。为此,我们引入了另一个检测特征:量化直方图差测度(QHDM)。

对于 QHDM,我们首先对 HSV(Hue, Saturation, Value)空间进行量化。因为人眼对色度(Hue)的敏感性比对饱

和度(Saturation)和亮度(Value)都要高,所以它的量化级数较后者多。经过实验比较,把 Hue、Saturation 和 Value 分别分成 32、10 和 16 三个等级可以取得令人满意的效果。因此所采用的直方图是 $32 \times 10 \times 16$ 的 3D-HSV 彩色直方图。令 $c \in \{H, S, V\}$ 和 $h_c(t)$ 表示在 HSV 模型中的第 t 帧的色度、饱和度和亮度彩色直方图,帧大小为 $M \times N$,则量化直方图差测度(QHDM)定义为:

$$D_h = \frac{1}{M \times N} \left(\sum_c \alpha_c \cdot \sum_{k=1}^n |h_c(t) - h_c(t+1)|^2 \right)^{1/2} \quad (1)$$

其中, k 是可能的颜色级 n 中的一个; $\alpha_c, c \in \{H, S, V\}$ 为权系数。由于大多数视频片段的基本特征都可以用其低级语义特征(如颜色、纹理、亮度等)来描述,而我们采用的 QHDM 特征同样能反映原始视频的这些特征。根据不同类型的视频(如新闻类、广告类、电影类等),可以通过调节权系数的取值使检测特征更能反映原始视频序列。对于新闻类的视频,由于它们的亮度不会发生太大的变化,因而可以设 $\alpha_h > \alpha_s$ 以及 $\alpha_h > \alpha_v$,从而弥补色度的不足。而对于广告类视频,通常这类视频为了吸引观众的注意力而会有较大的亮度和色度的变化,尤其是亮度。因而可以通过调节权系数 $\alpha_v > \alpha_h > \alpha_s$,从而降低视频中闪光灯、爆炸等造成的亮度突变的影响。而对于电影类视频,由于其视频内容的不确定性,使得系数的调节必须根据具体情况而定。如为战争题材的电影,其场景一般较暗,因而可以设定 $\alpha_v > \alpha_s, \alpha_v > \alpha_h$,使得镜头的变化主要以亮度为主;而戏剧题材的电影,一般其色彩和亮度均很丰富,且三个分量的比重差不多,因此一般会将三个参数的值设置为差不多大小。对于参数的调节,目前我们还仅是采用人工的方法,并且需要预先知道视频的类型,即系数非盲调节,今后可以在自动设置权系数上做进一步的工作。

由于本文采用 FCM 聚类算法不需要预先设定经验阈值,避免了由于无法找到一个普遍适用于各类视频的阈值而

造成的误检,从而提高了检测效率和精度。

2.3 淡化检测

2.3.1 Gaussian 加权 Hausdorff 距离

根据 FCM 聚类的结果,找出可能的渐变区间,再利用 Gaussian 加权 Hausdorff 距离的方法来检测是否为淡化过渡及其位置。

Hausdorff 距离是描述两组点集之间相似程度的度量方法^[11]。它与点集内所有点的分布情况相关,因而易受外界噪声等干扰的影响。基于此,有学者提出了修正的 Hausdorff 距离,其定义如下:

$$h(A, B) = \frac{1}{N_A} \sum_{a \in A} \min_{b \in B} \|a - b\| \quad (2)$$

其中, N_A 是点集 A 中点的个数。即对点集 A 中的每个点 a_i 到距离此点最近的 B 集中中点 b_j 之间的距离 $\|a_i - b_j\|$ 求平均,将这个平均值作为 $h(A, B)$ 的值。

由于图像的主要内容大都位于图像的中心区域,因而它的特征较为重要。在此,我们将修正的 Hausdorff 距离做进一步的改进,得到高斯加权的修正 Hausdorff 距离,其定义如下。

首先,我们将帧图像分成 M 块;然后,根据(2)式计算连续两帧对应块之间的 Hausdorff 距离 $h(A_m, B_m)$;最后,由这些块的 Hausdorff 距离经高斯加权和得到连续两帧之间的 Hausdorff 距离,即

$$H(A, B) = \sum_{m=1}^M \alpha_m \cdot h(A_m, B_m) \quad (3)$$

其中, $\alpha_m, m=1, 2, \dots, M$ 为高斯加权系数,即

$$\alpha_m = \frac{1}{2\pi\sigma_m^2} \exp\left\{-\frac{(i_m - i_0)^2 + (j_m - j_0)^2}{2\sigma_m^2}\right\}$$

(i_m, j_m) 和 (i_0, j_0) 分别为第 m 块的中心点像素和整幅图像中心点的像素。权系数的取值就加强了图像中心区域的重要性,从而使相异度函数更适合于原始视频。

2.3.2 淡化检测具体流程

在本方法中,用小波分解后的第一尺度上的三个高频分量来进行检测。在进行边缘提取的时候,边界点易受光照和运动的影响。经过大量的观察和比较发现,用 Sobel 算子提取的边界再用半径为 r (6 个像素) 的圆作为膨胀因子进行膨胀,能取得较好的检测效果。对于膨胀后的图像,用文[11]中的方法进行运动补偿。再利用(4)式计算相邻两帧的 Gaussian 加权 Hausdorff 距离。然后用 ECRin(出现的边界点数)和 ECRout(消失的边界点数)^[11] 以及 ECR(边界改变率)来表征镜头间的变化。为了更好地反映出现的边界点数与消失的边界点数的关系,我们定义第 n 帧和第 $n-1$ 帧之间的边界改变率 ECR_n 为:

$$ECR_n = \frac{ECRin_n / N_n}{ECRout_{n-1} / N_{n-1}} \quad (4)$$

其中 $ECRin_n$ 和 $ECRout_{n-1}$ 代表第 n 帧和第 $n-1$ 帧出现和消失的边界改变率; N_n 和 N_{n-1} 分别为第 n 帧和第 $n-1$ 帧的边界像素数。

根据本文的检测思想,淡化的检测过程如图 2 所示。

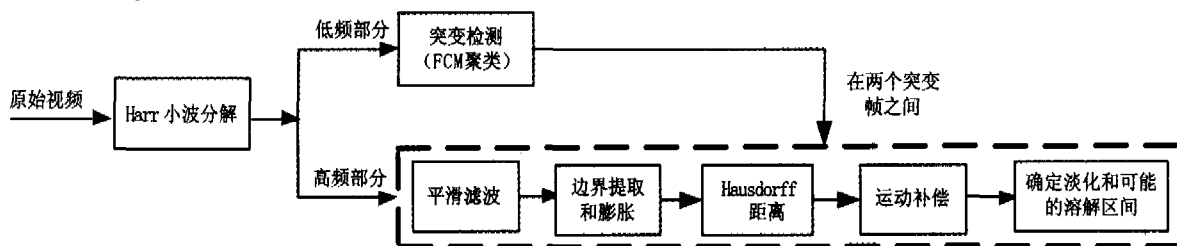


图 2 淡化的检测流程图

2.4 溶解检测

2.4.1 SCD 算法

根据 FCM 聚类的结果,我们对可能的渐变区间用加权 SCD 算法 (Similarity color distribution based method) 进行检测。SCD 算法的数学基础是任何概率分布都可以由它的矩来表征。因此,如果我们将颜色分布作为一种概率分布的话,颜色分布就可以由它的矩来表示^[13]。由于在镜头的渐变过程中,帧间的差异很小,我们不能仅用图像的某一种特征来检测,并且图像大部分内容信息主要集中在低阶矩上,因而我们将一阶矩(均值)、二阶矩(方差)和三阶矩(偏态)三种统计量加权求和后的特征值作为渐变过渡算法的检测特征。同时,为了增加帧间的差异,我们采用累积差的方法。因此,两帧图像之间的相异性函数可定义为:

$$dis(P, R) = \sum_{c \in \{H, S, V\}} (\omega_{1c} |E_c(P) - E_c(R)| + \omega_{2c} |\sigma_c(P) - \sigma_c(R)| + \omega_{3c} |S_c(P) - S_c(R)|) \quad (4)$$

其中 E_c, σ_c 和 S_c 分别代表图像帧的一、二和三阶矩; P 为当前帧, R 为代表帧。 $\omega_{ic}, i=1, 2, 3; c \in \{H, S, V\}$ 为权系数。可以通过调整权系数的取值使相异性函数更有效地适用于多种类型的视频序列。系数的调节可通过图 3 来进行。

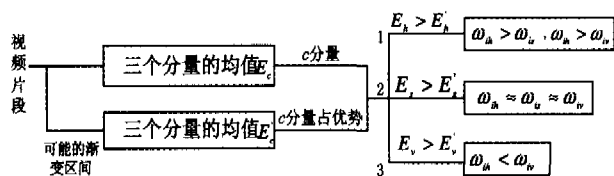


图 3 自动调节系数框图

所给的视频序列亮度变化不是很明显(如新闻视频),我们可以将权系数设置为 $\omega_h > \omega_s, \omega_h > \omega_v$, 让它能更有效地弥补平均颜色的不足。若所输入的视频序列为比较暗的视频片段(如海底世界),一般方法在进行检测的时候首先要进行亮度补偿。而在我们的算法中,只要将 ω_h 的值设置得小一点,而将 ω_v 的值设置大一点即可。通常,对于 HSV 颜色空间,为了能使色度比饱和度和亮度更能与原视频匹配,需设置色度通道中矩的权值比其它通道中的稍高一点。当然,在突变镜头检测的 QHDM 特征中,系数 α_c 的调节也可以参照图 2。但由于 FCM 算法中引用两个聚类特征,因而这种自动调节可能会影响 SDM 特征。所以在此我们并不主张对 QHDM 的权系数进行自动(盲)调节。

2.4.2 溶解检测具体流程

由溶解过渡的特性我们可以知道,当镜头消失时,统计量减小,而镜头出现时统计量增大。在一个理想的溶解过渡过程中,视频序列的SCD会呈现抛物线的形状。检测流程如图4所示。

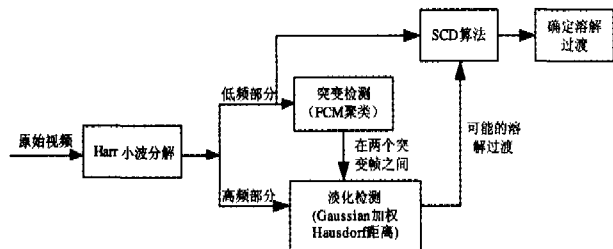


图4 溶解的检测流程图

2.5 划变检测

2.5.1 三维小波变换

视频序列实际上是由许多单帧图像构成的图像组,这些图像帧之间存在一定的关系,而这些关系是无法用图像内的信息来表示的。三维小波变换(3D-WT)则可实现对图像序列信号在空间和时间上的多分辨分析。因此,利用三维小波即可将帧内信息和帧间信息有机地结合起来。塔式分解算法的引入使得三维小波可以通过对视频序列的空间二维和时间一维分别做一维小波变换来实现,即在一幅图像内用二维小波变换描述帧内信息,再在时间序列上用一维小波变换描述帧间信息^[14]。图5所示为一级三维小波变换图。

对图像序列的三维小波而言,既可以利用图像多分辨的优点,同时在时间上又可以得到图像的低通帧和各级高通帧。低通帧包含了视频镜头中的静止成分,高通帧包含了视频镜头中的运动成分。本文将三维小波变换的思想应用到视频镜头的分割中,以期利用其中的高通帧来检测目前最难以检测

的划变过渡。

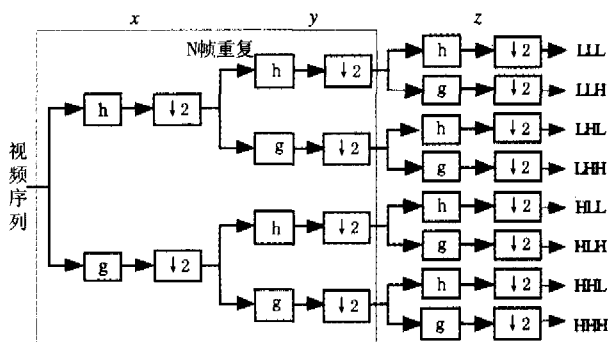


图5 一级三维小波变换框图

2.5.2 划变检测具体流程

划变的种类繁多,一般可分为线性划变过渡和非线性划变过渡。在此,仅考虑线性划变,即由边到边和角到角的划变。划变检测分为两个步骤:首先,确定划变的过渡区域,即划变带;其次,判别划变的类型,即通过运动矢量的方向来确定划变的种类。具体检测流程如图6所示。

1) 确定划变带

大量的观察发现,线性划变过程中都会有一个划变带,且划变带的速度一般是线性的,而视频图像中物体的运动比起划变带的运动速度要小得多。根据以上的分析,我们可以利用划变带的运动特性来检测划变过渡。

划变过程中的图像可分为已覆盖和未被覆盖的区域,用前后两帧图像相减即得到可能存在的划变带的差图像,再利用水平和垂直投影将划变带从图像中分离出来,具体步骤如图7所示。对于找到的划变带,定义它的中心作为其运动矢量。根据它的运动轨迹来判定划变过渡的类型。为了获得较宽的划变带,选用相邻4帧图像作差。

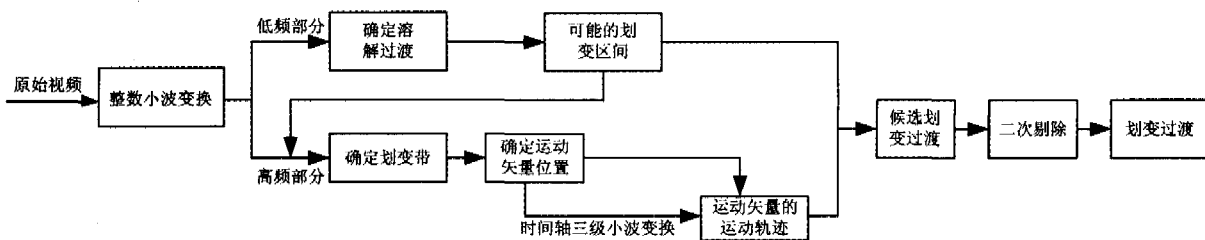


图6 划变检测流程图

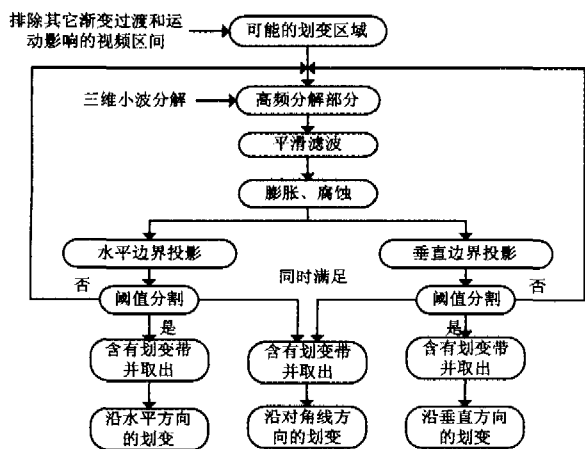


图7 划变带检测流程

2) 运动矢量的运动方向

为了确定运动矢量的运动轨迹,我们对原始序列图像沿时间轴做小波分解,选取窗口大小为8帧。当场景中存在快速运动时,变换后图像的高频子带中会存在很多较大的系数。我们通过提取高频子带中各帧对应像素最大值的方法来提取高频图像(高频信息)。设 $P_t(t=0,1,\dots,F-1)$ 表示原图像序列在小波变换以后的各帧, S_t 表示从变换后高频子带中提取出的高频图像,大小为 $M \times N$,则

$$S_t(i,j) = \max\{P_t(i,j) | 2^k - 1 < t < F\} \quad (5)$$

由此得到的高频图像包含了图像中最大的运动成分。由于划变带的运动速度远远大于视频图像中物体的运动速度,因而所提取出的高频图像中必定包含有划变带的运动成分。在确定了划变带之后,即可以通过划变带来提取运动矢量,从而根据运动矢量的运动方向来判断划变过渡的类型。

划变过渡对应于前一帧的场景图像逐块被后一帧取代的过程。图 8 所示为由左至右的划变过渡的划变带以及划变带

运动方向判断的示意图。

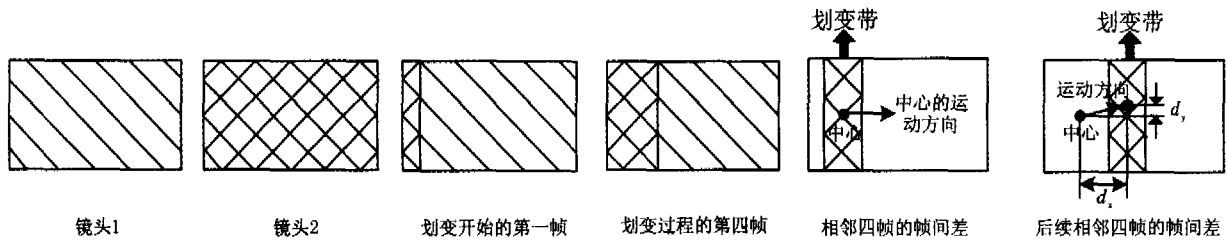


图 8 水平方向的划变的划变带及中心运动的示意图

确定了划变过渡之后,根据运动矢量的方向即可以确定出划变过渡的类型。设运动矢量的位移为 (d_x, d_y) ,则当 d_y/d_x 取不同值时,对应不同的过渡类型,具体判别由表 1 给出。

表 1 划变过渡类型判别

d_y/d_x	过渡类型
0	由左至右
π	由右至左
$\pi/2$	由上至下
$-\pi/2$	由下至上
$\pm\pi/4 \pm 3\pi/4$	角对角

3 实验结果及分析

为了验证本文中的各种检测算法的有效性,我们选取了三类真实视频片段作为实验数据,其中视频 1 是电影《云中漫步》、《西雅图未眠夜》以及《赤裸特工》的片段,视频 2 为 CCTV-1 的一段宣传片,视频 3 和视频 4 分别是 CCTV-1 的广告和新闻片段。另外,在这些视频中还含有各种如摄像机和物体的高速、低速的运动以及各种如闪光、爆炸等的影响。首先用人工方法找到镜头的边界,作为我们检测结果的评价标准。由于视频序列的数据量很大,我们认为误差在 10 帧以内都是允许的。然后将本文方法和文[11]中方法的检测结果做对比,如表 2 所示。另外,我们还将本文的方法与一些传统的镜头边界检测方法,如直方图方法等,做了对比实验,结果在后续图中有所比较。表 2 中定义:

$$\text{Recall} = \frac{n_c}{n_c + n_m}, \text{Precision} = \frac{n_c}{n_c + n_f} \quad (6)$$

n_c, n_m, n_f 分别表示正确、漏检和误检的镜头个数。

表 2 本文方法与文[11]方法的对比结果

测试视频	视频 1	视频 2	视频 3	视频 4	总和	
视频时间	34'	4'12"	12'36"	66'28"	117'16"	
镜头数	268	52	193	713	1226	
本文方法	检出数	264	58	208	725	1255
	漏检数	14	2	2	13	31
	误检数	10	8	17	25	60
	Recall	94.7%	96.1%	98.9%	98.2%	97.4%
	Precision	96.2%	86.2%	91.8%	96.5%	95.2%
文[11]方法	检出数	282	55	210	733	1280
	漏检数	20	7	13	28	68
	误检数	34	10	30	48	122
	Recall	92.5%	86.5%	93.2%	96.0%	93.8%
	Precision	87.9%	81.8%	85.7%	93.4%	89.7%

由表 2 中实验结果我们可以看出,文[11]中的方法对于

电影的误检很多,这是因为在电影中存在大量的物体运动,从而很容易误检为溶解;而广告中的溶解过渡的漏检较多是因为广告中的溶解持续的时间较短且较难分辨,而运动的影响远远大于溶解的影响,从而使得检测率不高。

检测过程中的漏检主要是由于在一些镜头边界中帧对间小的内容变化,而大多数错检是由于在渐变过程中不规则的摄像机运动造成的。

在视频 2 中存在误检主要是由于片段中存在物体的快速运动以及在制作过程中编辑人员对影片做了“透明”处理(将两个镜头进行了叠加但并不是溶解过渡)。

图 9 和图 10 所示为一段新闻视频的聚类检测结果及其 ECR 的检测结果。由图中可以看到:相邻的两个突变帧为第 280 帧和第 450 帧,而实际的渐变过渡为 380~420 帧,它的前后几十帧为没有任何编辑特性的过渡。

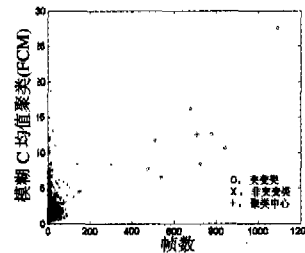


图 9 一段新闻视频的聚类结果

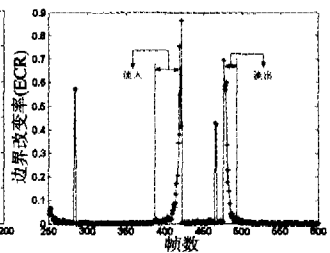


图 10 淡化检测的结果

由于淡化的一个重要特点就是无论是淡入还是淡出,它必定有一个单色帧,因而对于淡化过渡必定会有一个单侧峰,而且单色帧一般受光照和运动的影响较小,因而可以同时根据 ECR、ECRin 和 ECRout 来共同确定淡化。在部分实验中,由于 Fade 的坡过于陡且它的峰又不如它两侧 Cuts 的峰高,因此有时候会将 Fade 与 Cuts 混淆,造成漏检和误检,如图 11 所示。在这种情况下,我们可以同时观察 ECRin 和 ECRout 来进行判断,而非只利用单一的 ECR,如图 12 所示。

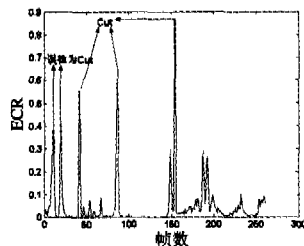


图 11 ECR 的检测结果

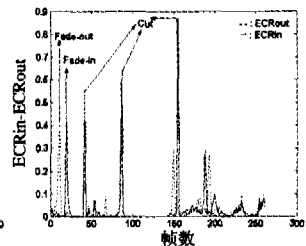


图 12 ECRin-ECRout 的检测结果

文[7]中的方法直接应用 Hausdorff 距离对溶解过渡进行检测,虽然能够达到一定的检测效率,但是易受运动的影响。虽然进行了运动补偿,但它只能对小物体的小运动有一定的抑制作用,对于运动速度较快的大物体,它的影响在某些

时候其 ECR 特征类似于溶解过渡,如图 13 所示,此时易造成误检。而在我们的第三层检测中,可将其剔除,如图 14 所示。

另外,文[9]中的方法在检测淡化和溶解之前,首先要剔除突变的影响,这就需要设定门限,从而受人因素为影响。在本文方法中,突变的剔除已经在第一层的检测中过滤掉了,且不需要参数,因而在利用 Hausdorff 距离对淡化进行检测的时候,突变的影响就可以不予考虑。另外,直接用 Hausdorff 距离有一个缺陷,就是对于剧烈运动时 Cut 有误检,而我们的方法有效地抑制了这一点。再者,文[9]中的算法受光照的影响较大,尤其是对于溶解的检测。而本文的方法只是利用 Hausdorff 距离对溶解过渡进行初定位,再利用第三层的 SCD 算法进行检测。

溶解的检测就较为复杂,因为溶解前后的镜头具有任意性。若直接用 Hausdorff 距离进行检测,则受参数的干扰更大。因此本文在第二层的检测过程中,确定可能的溶解过渡,当然这些可能的溶解过渡中可能包含由于运动造成的干扰。在第三层的检测算法中可以一并剔除,如图 13 和 14。

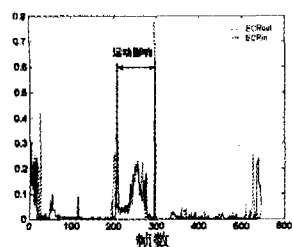


图 13 ECR 的检测结果

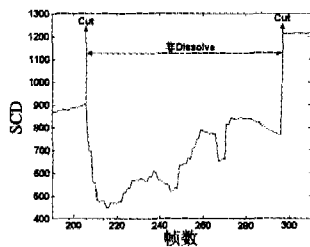


图 14 SCD 的复查结果

在 SCD 方法中,系数 $\omega_h, \omega_s, \omega_v$ 的设置对检测结果起着关键的作用。在视频 1 中有一处溶解过渡。当我们取 $\omega_h = 1, \omega_s = 1, \omega_v = 1$ 时,由于场景较暗,检测结果中不能很清晰地看出溶解过渡的边界,如图 15 所示。而将亮度系数 ω_h 调大,即令 $\omega_h = 0.05, \omega_s = 0.05, \omega_v = 0.9$ 时,我们可以得到较好的检测结果,如图 16 所示。

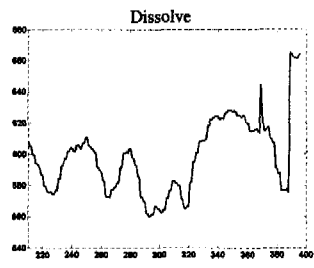


图 15 未经系数调节的溶解检测结果

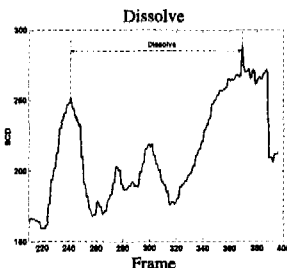


图 16 经系数调节的溶解检测结果

在直方图检测方法中,对视频 1 片段中的淡入过渡有一个漏检,这是因为在该片段中渐变过渡是在昏暗的背景下,人物亮度也不高的情况下由黑色的视频帧淡入到场景,如图 17 所示。而在我们的算法中,由于调整了权系数 $\omega_h, \omega_s, \omega_v$ 的值(增大亮度系数 ω_h 的值,降低其它两个系数 ω_s, ω_v 的值),得到了淡出的结果,提高了检测结果。本实验中取 $\omega_h = 0.10, \omega_s = 0.05, \omega_v = 0.75$,原视频中在 488~657 帧处为淡入,结果如图 18 所示。

另外,经过大量的实验观察,发现直方图方法容易将相连的淡入和淡出过渡或淡出和淡入过渡误检为溶解过渡。图 19 所示即为用直方图方法对视频 2 中的一个相连的淡出和淡入过渡的检测结果,它将 640~765 帧处的一段过渡误检为

溶解过渡。而由于本文采用的是先切分后检测的方法,因此在检测之前,已经将相连的淡出过渡和淡入过渡切分成了两个镜头,然后分别对这两个镜头进行检测,因而不会将其误检为一个溶解过渡,从而避免了直方图方法的误判。图 20 所示为用本方法对淡出为白色帧的过渡检测结果。

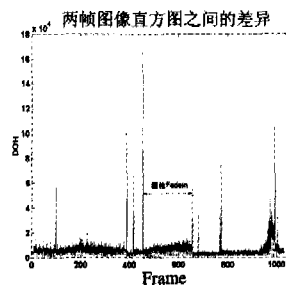


图 17 直方图方法的检测结果

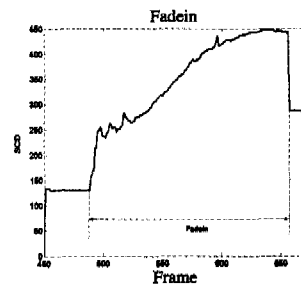


图 18 本文方法的检测结果

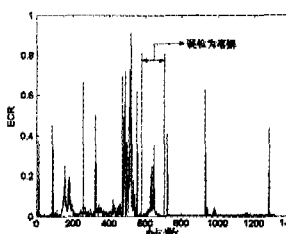


图 19 直方图差(DOH)的检测结果

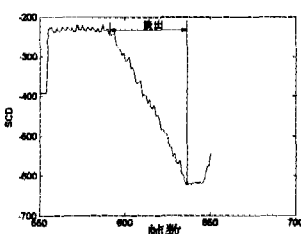


图 20 本方法检测出的淡出过渡

相对于淡化和溶解的检测,划变的检测更为不易。文[15]中利用差分图像在水平和垂直两个方向的投影,得到有关的结构和统计信息,来检测沿水平和垂直方向的划变。但是该方法仅能将划变与其它如突变、溶解等镜头过渡区分开,并不能判断出划变过渡的类型,而且对对角线划变产生漏检。另外,当帧内的物体运动速度较快时,该算法失效。而本文的算法在确定划变带时就已经将物体的快速运动点的影响排除在外,从而检测效率有所提高。

同样地,文[11]中的方法,也只能将划变与其它镜头过渡区分开,并不能将某些运动影响从中剔除,从而造成误检,如图 21 所示。而本文的方法有的运动影响已经在做淡化和溶解的检测时就已经剔除,剩下的也很容易从我们的检测算法中淘汰。图 22 给出的是划变过渡的检测结果。由图中可以看出,运动矢量沿 Y 轴的增量几乎为零(沿水平方向运动)而沿 X 轴单调递增,因而 $d_y/d_x \approx 0$,即运动矢量夹角为 0 度,从而可以判断该划变过渡是一由左向右的水平划变过渡。

此外,在我们的检测过程中,运动矢量的运动轨迹所确定的划变,仅作为候选划变过渡,再经二次剔除,即长度小于 3 帧的划变被认为是误检,将其去除。

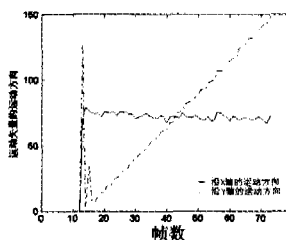


图 21 文[11]方法对划变的检测

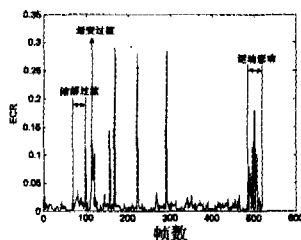


图 22 本文方法对划变过渡的检测

针对其它算法在判断具体位置上有所偏差,本文由于通过测量矢量的运动方向来定义过渡的类型,从而使得检测率有所提高。针对视频中常见的几种过渡类型进行检测,结果如表 3 所示。

表 3 不同类型的 Wipe 转换结果

实际的 wipe 转换模式	实际的位置	检测出的位置	矢量夹角(≈)	检测出的 wipe 转换模式
由左至右	64~73	66~73	0°	由左至右
由右至左	33~42	34~43	180°	由右至左
由上至下	49~58	48~57	90°	由上至下
由下至上	52~62	53~62	-90°	由下至上
由左上至右下 (page peel)	58~73	59~73	135°	由左上至右下 (page peel)
由右上至左下	55~68	55~68	45°	由右上至左下
由左下至右上	12~22	12~22	-135°	由左下至右上
由右下至左上	21~33	21~33	-45°	由右下至左上
Barn-door	16~75	16~75	0°,180°	Barn-door
Checker-wipe	8~74	9~73	多个矢量 0°	Checker-wipe

由于在视频中,各种 Wipe 过渡的速度有所不同,针对新闻视频中常见的由左至右的过渡和体育视频中由左上至右下的过渡,我们也对用 Adobe premier 6.5 生成的不同持续时间(速度)的划变进行测试,结果见表 4,表明该算法能够适用于不同的划变速度。

表 4 不同持续时间的 Wipe 转换检测结果

实际的 wipe 转换模式	实际的 Wipe 转换模式	检测到的 Wipe 转换模式
由左至右 1	12~18	11~18
由左至右 2	12~25	12~26
由左至右 3	12~50	12~50
由左至右 4	12~96	12~96
由左上至右下 1	15~23	16~23
由左上至右下 2	15~30	15~30
由左上至右下 3	15~63	15~63
由左上至右下 4	15~132	15~132

结论 本文根据分层的和多分辨的检测思想,综合运用小波、聚类等方法以及提出的高斯加权 Hausdauff 距离、加权 SCD 算法以及三维小波运动矢量分别对视频序列中的镜头过渡进行分类检测。该方法不但能同时检测突变镜头和渐变镜头,还可以判断出渐变过渡的类型以及首末帧的位置。并且,我们的检测方案并不是单一地将各种方法拼接在一起,而是将它们有机地结合起来,互相关联,达到有效检测。并且在检测方案的一些算法中,我们可以通过调节权系数而使检测相异度函数更能适用于多种视频片断。另外,本文的方法能有效地抑制如闪光及各种运动造成的影响,从而为镜头检测方面提供一个新思路,并为后续的镜头内容的分析以及进一步进行基于内容的视频检索奠定良好的基础。通过实验以及和其它算法的检测结果比较,本文给出的方法检测效果还是相当理想的。

参 考 文 献

- Lienhart R. Comparison of Automatic Shot Boundary Detection Algorithms. Proc. In SPIE Conf. on storage and retrieval for image & video debase VII,1999(3656): 290~301
- Zhang H J, Smoliar S W. Developing power tools for video indexing and retrieval. Proceedings of the SPIE: Storage and Retrie-

- al for Image and Video databases II, 1994, 2185: 140~149
- Jiang Haitao, Helal A S. Scene change detection techniques for video databases systems. Multimedia systems, 1998(6): 186~195
- Hampapur A, Jain R, Weymouth T E. Production model- based digital video segmentation, Multimedia Tools and Applications, 1995, 1(1): 9~46
- Alattar A M. Wipe scene change detector for use with video compression algorithm and MPEG-7. IEEE Transaction on Consumer Electrics, 1998, 44(1): 43~51
- 祁卫, 钟玉琢. 基于 MPEG 国际标准压缩视频流的镜头切分算法. 清华大学学报(自然科学版),1997,37(9): 50~54
- Swekdens W. The lifting scheme; A construction of second generation wavelets [J]. SIAM J Math Anal, 1997
- Calderbank A R, Daubechies I, Swekdens W. Wavelet Transforms that map Integers to Integers [J]. Appl Comput Harmon Anal, 1998, 7(5): 332~369
- 朱兴全, 薛向阳, 吴立德. 一种自动门限选取的视频 Shot 分割方法. 计算机研究与发展, 2000, 37(1): 80~85
- Gao Xinbo, Tang Xiaou. Unsupervised Video-Shot Segmentation and Model-Free Anchorperson Detection for News Video Story Parsing. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(9): 765~776
- Zabih R, Miller J, Mai K. A feature-based algorithm for detecting and classifying production effects. Multimedia Systems, 1999, 7:119~128
- 金红,周源华. 用 Hausdorff 距离进行视频镜头边界检测. 数字视频, 2002, 11(221): 12~14
- Stricker M, Orengo M. Similarity of color images. SPIE Storage and Retrieval for Image and Video Databases,1995 III, 2420: 381~392
- Li Y, Gao X B, Ji H B. A 3D Wavelet Based Spatial-temporal Approach for Video Watermarking. In:Proc. of the 5th International Conf. on Computational Intelligence and Multimedia Applications, 2003. 260~265
- Wu Min, Wolf W, Liu Bede. An Algorithm for Wipe Detection. IEEE Inter Conf on Image Processing(ICIP'98), Chicago, 1998 (1114)