

基于视频编辑模型的视频淡入、淡出和叠化的检测¹

陆海斌 章毓晋

(清华大学电子工程系 北京 100084)

摘要 该文基于视频编辑模型,推导出几个能用于检测镜头淡入、淡出和叠化的重要特征:方差、边界强度、有效平均梯度和双重色度差。通过比较,我们选取了方差序列和双重色度差特征作为检测依据,首先利用方差序列检测淡入、淡出和可能的叠化位置,以保证取得高的查全率,然后利用双重色度差序列确认叠化以提高检测准确率。由于检测对 DC 序列进行,算法计算时间复杂度比较低。

关键词 镜头检测,淡入,淡出,叠化

中图分类号 TN919.8, TP391.4

1 引言

将视频流切分成镜头是组织视频和建立视频索引的必要准备。镜头的变化可分为切变和渐变两种。常用的切变检测算法利用对应像素差异、直方图比较、边界变化以及统计运动信息等。文献 [1] 给出了这些算法的比较。较多存在于视频中的渐变主要有淡入 (Fade in)、淡出 (Fade out) 和叠化 (Dissolve) 等,对这些渐变的检测也很重要^[2-4]。文献 [2] 提出双阈值比较的算法;文献 [3] 计算间隔一定帧的对应像素差异,并用寻找“平台”的方法检测渐变。这两种算法虽然能达到很高的查全率,但由于对照相机和物体的运动过于敏感,误检多,造成检测准确率比较低。文献 [4] 的边界跟踪法要求在计算不相似度之前作全局运动补偿,因而时间复杂度高,准确率也不高。

本文提出的检测算法以 DC(Direct Current) 序列^[3] 为检测输入,在确保有高查全率的前提下,又借助确认步骤提高了检测的准确率。下面我们先给出视频编辑模型,然后利用该模型推导出一些能用于检测的重要特征,再利用这些特征检测淡入、淡出和叠化,并给出对真实视频的试验结果。

2 淡入、淡出和叠化检测算法

图 1 给出镜头检测的总体流程图(其中计算 σ , 检测 Fade, 检测 Dissolve 是本文的工作)。具体过程如下:首先从 DC 序列中检测切变位置,然后在相邻的切变位置之间检测是否有渐变存在。渐变检测分三步:首先计算每帧图像的方差;其次利用方差序列检测淡入、淡出的位置以及可能的叠化位置;最后再用双重色度差(Double Chromatic Difference, DCD) 特征^[5] 来确认叠化位置,以提高检测的准确率。

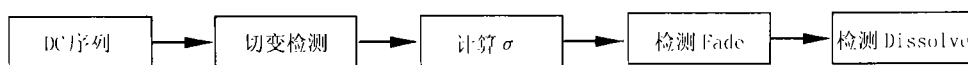


图 1 镜头检测流程图

2.1 视频编辑模型

根据视频编辑的实际情况,可得到常见的几种渐变,即淡入、淡出和叠化的近似模型为:

¹ 2000-05-18 收到, 2000-10-10 定稿

国家自然科学基金(69672029 和 60172025)、国家高技术研究发展计划(863-317-9604-05)资助

$f(x, y, t) = \alpha(t)p(x, y, t) + \beta(t)q(x, y, t)$, $0 \leq t \leq T$. 其中 $\alpha(t)$ 在渐变发生期间 $[0, T]$ 是单减函数, $\alpha(0) = 1$, $\alpha(T) = 0$; $\beta(t)$ 在渐变发生期间 $[0, T]$ 是单增函数, $\beta(0) = 0$, $\beta(T) = 1$. 对淡入, 有 $\alpha(t) = 0$, 于是: $f(x, y, t) = \beta(t)q(x, y, z)$, $0 \leq t \leq T$. 对淡出, 有 $\beta(t) = 0$, 于是: $f(x, y, t) = \alpha(t)p(x, y, t)$, $0 \leq t \leq T$. 叠化可看作淡入和淡出的组合.

从上面的公式我们很容易得到: 在淡出结束时, 或淡入开始时, 图像的亮度方差接近 0. 这是一个很重要的特点, 可以用来实现鲁棒性高的淡入、淡出检测算法. 需要注意, 一幅图像既可以淡出到全黑, 也可以淡出到全白或其它单一颜色. 因此如果仅假定淡出到全黑, 就会造成漏检. 对淡入也是如此, 图像应可以从任意单一颜色淡入.

在后面的讨论中, 我们仅假设 $\alpha(t) + \beta(t) = 1$ 和 $p(x, y, t) = p(x, y)$, $q(x, y, t) = q(x, y)$. 前一个假设并没有限制 $\alpha(t)$ 为线性函数, 比较接近真实情况; 后一个假设对于没有物体和摄像机剧烈运动的一般渐变是合理的. 值得指出的是, 假设 $\alpha(t)$ 和 $\beta(t)$ 为线性函数, 比如, $\alpha(t) = 1 - t/T$, $\beta(t) = t/T$, 是不太合理的. 现有的渐变检测算法多利用这个线性函数, 但是实际影片的淡入、淡出和叠化的 $\alpha(t)$ 和 $\beta(t)$ 在多数情况下并不满足线性要求, 这样就会由于线性模型与实际不符合而造成查全率的降低. 另外, 由于我们选用 DC 序列检测渐变, 图像的平滑可以在一定程度上补偿物体和摄像机的小运动带来的影响.

2.2 可用于检测的特征

从视频编辑模型可以得出以下几个用于渐变的特征: (1) 边界强度; (2) 序列方差; (3) 有效平均梯度 (Effective Average Gradient, EAG)^[6]; (4) DCD.

如果令 $G(x, y)$ 代表 $f(x, y)$ 的梯度图, 则像素点 (x, y) 的边界强度可如下计算: $G^2[f(x, y)] = G_x^2(f) + G_y^2(f)$. 从视频编辑模型我们可以推导出: $G_x(f) = \alpha(t)G_x(p) + \beta(t)G_x(q)$, $G_y(f) = \alpha(t)G_y(p) + \beta(t)G_y(q)$. 如果令 $G^2(f) = \sum G^2(x, y)$, 则

$$\begin{aligned} G^2(f) &= \sum_{x,y} \{[\alpha(t)G_x(p) + \beta(t)G_x(q)]^2 + [\alpha(t)G_y(p) + \beta(t)G_y(q)]^2\} \\ &\approx [G^2(p) + G^2(q)]\alpha(t) - 2G^2(q)\alpha(t) + G^2(q) \end{aligned} \quad (1)$$

考虑到 $\sum G_x(p)G_x(q) \approx 0$, $\sum G_y(p)G_y(q) \approx 0$, 视频图像的边界强度序列在渐变期间会呈现抛物线形状, 其局部最小值位于: $\alpha(t) = G^2(q)/[G^2(p) + G^2(q)]$.

图像序列的均值为: $E[f(t)] = \alpha(t)E(p) + \beta(t)E(q) = \alpha(t)[E(p) - E(q)] + E(q)$, 而方差为:

$$\begin{aligned} \sigma^2(f) &= E(f - \bar{f})^2 = E(\alpha(t)p + \beta(t)q - \alpha(t)\bar{p} + \beta(t)\bar{q})^2 \\ &= \alpha^2(t)\sigma^2(p) + \beta^2(t)\sigma^2(q) + 2\alpha(t)\beta(t)E[(p - \bar{p})(q - \bar{q})] \\ &\approx \alpha^2(t)\sigma^2(p) + \beta^2(t)\sigma^2(q) \\ &= [\sigma^2(p) + \sigma^2(q)]\alpha^2(t) - 2\sigma^2(q)\alpha(t) + \sigma^2(q) \end{aligned} \quad (2)$$

可见, 图像序列的方差在渐变过程中会呈现抛物线的形状, 其局部最小值应发生在: $\alpha(t) = \sigma^2(q)/[\sigma^2(p) + \sigma^2(q)]$.

EAG 的定义如下^[6]: $EAG = TG/TP$, 其中 $TG = \sum G(x, y)$ 为梯度图的总梯度值, $TP = \sum p(x, y)$ 为非零梯度像素的总数, 这里 $p(x, y)$ 定义为

$$p(x, y) = \begin{cases} 1, & G(x, y) > 0 \\ 0, & G(x, y) = 0 \end{cases} \quad (3)$$

EAG 序列在渐变过程中呈现较明显的谷, 也可说, 通过计算 EAG 可以使抛物线的谷变得明显。

文献 [5] 提出的 DCD 特征能将叠化从变焦 (Zoom), 摇镜头 (Pan), 和划变 (Wipe) 中区分出来。借助视频编辑模型可进一步分析如下: $DCD(t) = \sum F(|[f(x, y, t_0) + f(x, y, t_N)]/2 - f(x, y, t)|)$, $t_0 < t < t_N$, 其中 t_0, t_N 分别是叠化过程中的两点; $F(\cdot)$ 是阈值函数。为方便令阈值为 0, 则

$$\begin{aligned} DCD(t) &= \sum_{x,y} \left| \frac{f(x, y, t_0) + f(x, y, t_N)}{2} - f(x, y, t) \right| \\ &= \sum_{x,y} \left| \frac{\alpha(t_0) + \alpha(t_N)}{2} - \alpha(t) \right| \times |p(x, y) - q(x, y)| \\ &= \left| \frac{\alpha(t_0) + \alpha(t_N)}{2} - \alpha(t) \right| \times \sum_{x,y} |p(x, y) - q(x, y)| \end{aligned} \quad (4)$$

因为 $\alpha(t)$ 是减函数, 所以 $DCD(t)$ 也出现抛物线形状, 这一特征可被用来确认叠化。在 $\alpha(t) + \beta(t) = 1$ 的假设下, 可以证明, 总存在一帧 t_m , $t_0 < t_m < t_N$ 使得: $[f(x, y, t_0) + f(x, y, t_N)]/2 = f(x, y, t_m)$, 即 $DCD(t_m) = 0$ 。

2.3 特征选择和算法讨论

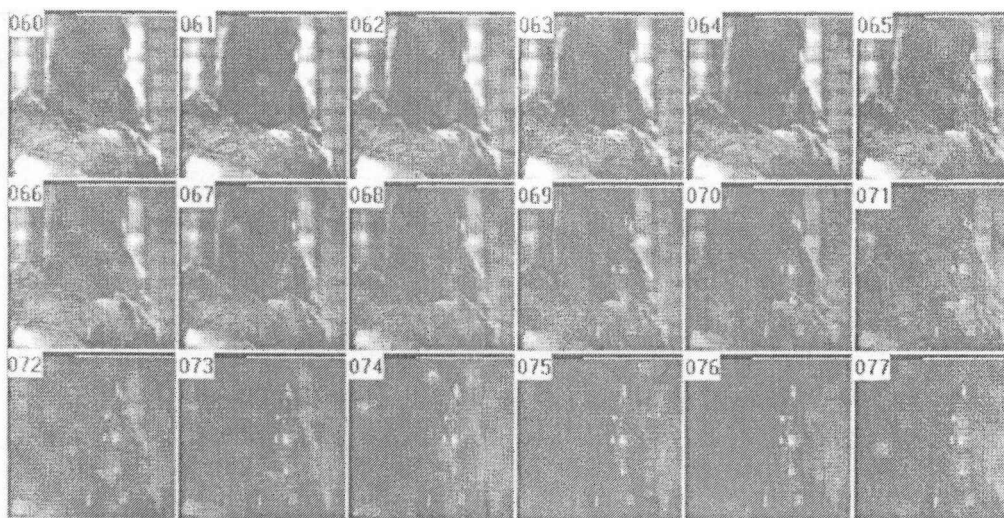
比较 (1), (2) 式, 我们可以看出叠化期间方差序列和边界强度序列都呈现出抛物线形状, 这个重要特征可被用来挑选出可能的叠化位置。当图像 p 或 q 的方差或边界强度变化较小时 (例如如图 2(a) 和图 2(b) 的片段), 抛物线的谷可能不很明显 (见图 2(c) 上半部分), 而 EAG 序列的谷较明显 (见图 2(c) 下半部分)。但是, 由于当叠化发生时, 总有一段抛物线出现, 而且局部最小值总存在, 且位于这段抛物线中的某处, 所以抛物线的谷不明显并不很影响检测可能的叠化位置。换句话说, 这几个特征都能被用来可靠地选出可能的叠化位置, 达到很高的查全率。

本文算法选用方差序列, 而不用边界强度特征是因为前者时间复杂度小, 而且 DC 序列仅是原图的低频部分, 并不适合用来检测边界。没有采用 EAG 序列, 主要是因为用它的效果与用方差序列差不多, 而用后者已经能有效地选出可能的叠化位置。另一方面, 因为物体和摄像机的运动也有可能造成类似用方差序列得到的抛物线形状, 从而造成误检, 导致检测准确率不高。所以在叠化检测中进一步利用 DCD 特征作确认以提高检测准确率。

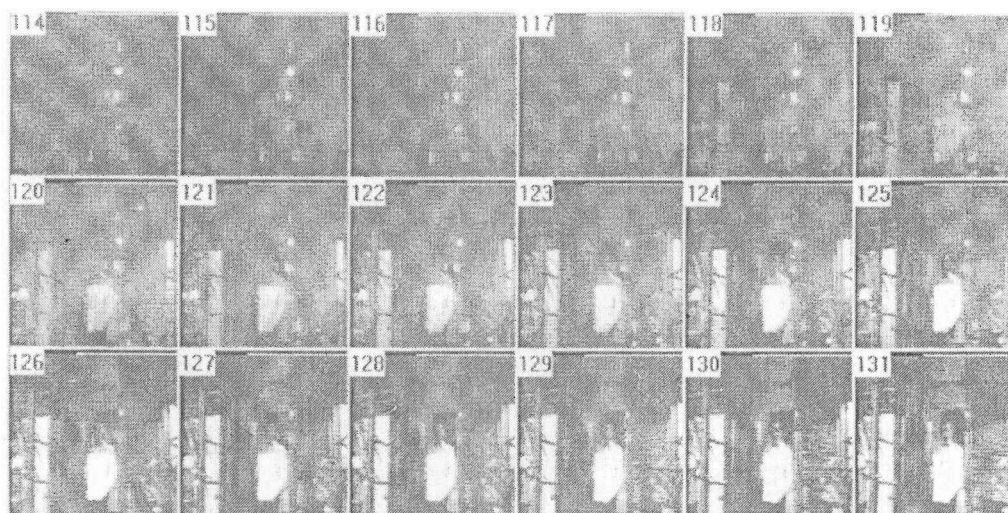
从 (4) 式可以看到, 任意选择叠化过程中的某两点间计算 DCD, 都能得出抛物线形状的 DCD 序列。也就是说, 为计算 DCD 序列, 没有必要准确地定位叠化的起点和终点。另一方面, 我们在实验中发现, 叠化起点和终点的定位常难以做到很精确, 常有几帧的偏差。另外, 由于物体或摄像机的小运动常常是不可避免的, 那么当叠化起点和终点间隔较远时, 就会有较大的运动积累, 导致前面对 $p(x, y, t) = p(x, y), q(x, y, t) = q(x, y)$ 的假设不成立, 从而影响 DCD 序列的抛物线形状。鉴于此, 我们在叠化起终点区域 $[0, T]$ 间有目的地选择 t_0, t_N , 使得 $0 \leq t_0 \leq t_N \leq T$, 且相距较近。这可以在一定程度上补偿上述的小运动积累问题。参见图 3 的示意图, 其中曲线是方差序列, 3 处可能的叠化位置在图中分别用 1, 2, 3 表示, 1, 2 的情况对应图 2(c), 黑点表示局部最大值和局部最小值位置。图 3 中实箭头指示所选的 t_0 位置, 虚箭头指示所选的 t_N 位置。

3 实验结果

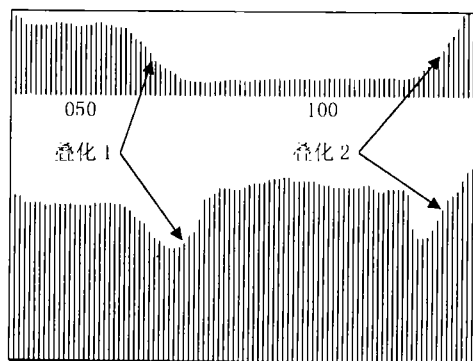
我们选择了几段真实视频 (包含 12 个淡入, 11 个淡出) 来测试淡入、淡出检测算法的有效性, 其中一段视频如图 4(a) 所示, 该视频对应的方差序列和 EAG 序列见图 4(b); 可以看出,



(a) 从 Casablanca 中检测出的一个叠化片段 (后化入的画面强度较弱)



(b) 从 Casablanca 中检测出的一个叠化片段 (先化出的画面强度较弱)



(c) 上半部分是方差序列 (边界强度序列也有类似结果), 下半部分是 EAG 序列
图 2

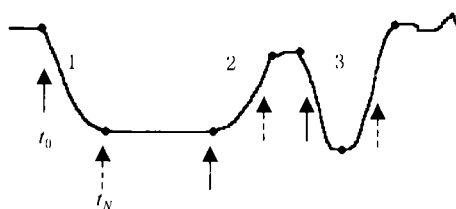
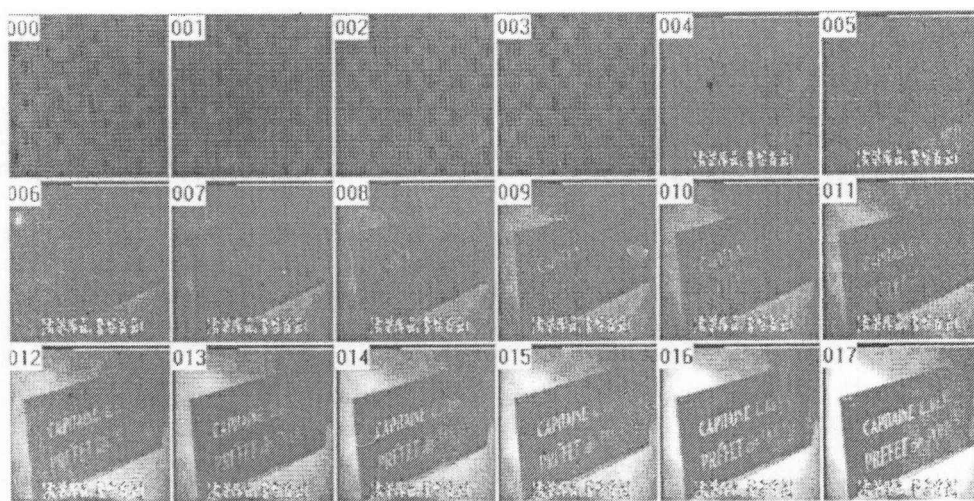
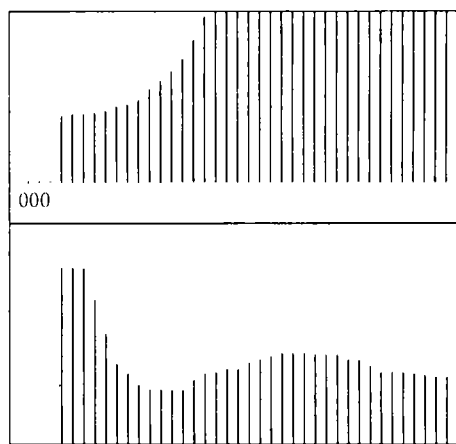


图 3 t_0, t_N 的选择示意

EAG 序列在方差序列为近似零的地方仍然有一定的数值，可见在检测淡入、淡出方法中使用方差序列更合适。实验中我们对以上淡入、淡出的检测取得了 100% 的查全率和准确率。



(a) 一个淡入片段



(b) 与图 2(a) 视频片段对应的方差序列 (上半部分) 和 EAG 序列 (下半部分)

图 4

作为叠化检测的数据，我们从影片“Casablanca”中选取了 10 段视频，每段包含若干个切变和一个叠化；又从影片“CONAN- the barbarian”中选择了一段包含 9 个叠化和若干个

切变的视频段, 其中有不少物体的运动。叠化检测的实验结果见表 1。

表 1 叠化检测结果

影片	总数	检测出	正确	查全率	准确率
Casablanca	10	11	9	90%	81%
CONAN	9	10	9	100%	90%

4 结 论

本文基于淡入、淡出和叠化的视频编辑模型, 推导出可以用于检测这几种渐变的若干特征: 边界强度、序列方差、EAG 以及 DCD 特征。利用序列方差来检测淡入、淡出已取得了满意的检测效果。在叠化检测方面, 通过先利用方差序列提取可能的叠化位置, 再用 DCD 特征加以确认的方法, 也取得了较好的结果。由于算法对 DC 序列进行检测, 因此时间复杂度比较低, 有实用价值。

参 考 文 献

- [1] J. Boreczky, L. Rowe, Comparison of video shot boundary detection techniques, SPIE, 1996, 2670, 122-128.
- [2] H. J. Zhang, A. Kankanhalli, S. W. Smoliar, Automatic partitioning of full-motion video, ACM Multimedia Systems, 1993, 1, 10-28.
- [3] B. L. Yeo, B. D. Liu, Rapid scene analysis on compressed video, IEEE Trans. on CASVT, 1995, 5(6), 533-544.
- [4] R. Zabih, J. Miller, K. Mai, A feature-based algorithm for detecting and classifying production effects, Multimedia Systems, 1999, 7(2), 119-28.
- [5] H. Yu, G. Bozdagi, S. Harrington, Feature-based video segmentation, IICIP'97, Santa Barbara, USA, 1997, 498-501.
- [6] Y. J. Zhang, J. J. Gerbrands, Transition region determination based thresholding, Pattern Recognition Letters, 1991, 12(1), 13-23.

FADE AND DISSOLVE DETECTION BASED ON VIDEO EDITING MODEL

Lu Haibin Zhang Yujin

(Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract Based on video editing model, several important characteristics that can be used for detecting fade and dissolve, such as the variance, edge intensity, Effective Average Gradient (EAG) and Double Chromatic Difference (DCD) of video sequence are derived. After comparison, the variances are used to achieve robust fade detection. It is approved that in terms of recall rate for dissolve detection, the variance, edge intensity and EAG have similar efficiency. So, the variances are used for selection of possible positions of dissolve transitions, which guarantees high recall rate, and the DCD are then used for further confirmation to improve precision rate. Since the detection is carried out on DC sequence, the computation complexity of this method is relatively low.

Key words Shot detection, Fade in, Fade out, Dissolve

陆海斌: 男, 1973 年生, 博士生, 所学专业为图像处理。

章毓晋: 男, 1954 年生, 教授, 博士生导师, 现主要从事图像工程(图像处理、图像分析、图像理解及其工程应用)的研究和教学工作, 已发表论文 160 篇, 著书 5 部。