

基于三步筛选的视频渐变镜头检测

王剑峰¹, 杜奎然²

(1. 重庆航天职业技术学院计算机工程系, 重庆 400021; 2. 重庆爱立信通讯有限公司, 重庆 400000)

摘要: 针对视频中的叠化与淡入淡出现象, 提出一种基于三步筛选的渐变镜头检测算法。提取视频帧的亮度和方差作为特征, 通过有限状态机实现初始渐变检测, 并计算视频帧的颜色、共生矩阵、运动特征, 从而进行三步筛选, 保证检测的准确性。对 TRECVID 视频进行实验, 结果表明, 该算法对渐变具有较好的检测性能, 对运动及闪光现象有较强的鲁棒性。

关键词: 镜头检测; 共生矩阵; 视频帧; 运动特征; 渐变

Video Gradual Transition Shot Detection Based on Three Steps Filter

WANG Jian-feng¹, DU Kui-ran²

(1. Dept. of Computer Engineering, Chongqing Aerospace Vocational Technology College, Chongqing 400021, China;

2. Chongqing Ericsson Communications Limited Company, Chongqing 400000, China)

[Abstract] Aiming at the dissolve and fade in fade out phenomenon of video, this paper proposes a video shot detection algorithm based on three steps filter. A finite state machine is used to search gradual transition candidates in intensity variance continuity signal, and verify those to get real gradual transitions, next, extract color, texture and motion feature of video frame to guarantee the precision of gradual shot detection. Experimental results based on the videos from TRECVID demonstrate good performance of the algorithm and its robustness against disturbances caused by object motion or flashlight.

[Key words] shot detection; co-occurrence matrix; video frame; motion feature; gradual transition

DOI: 10.3969/j.issn.1000-3428.2011.24.090

1 概述

随着通信技术, 特别是数据存储与传输技术的发展, 新闻视频可以非常容易地从电视、卫星、网络等众多渠道获得。为了使人们快速地获取需要的信息, 应自动建立索引结构, 以进行新闻视频的组织、管理和检索。镜头作为视频结构中的最小单元, 其检测成为新闻视频分析和检索的基础工作, 镜头之间的转换方式主要有两大类: 突变和渐变。突变是镜头间的突然变化, 常在 2 帧图像间完成, 渐变则是从一个镜头缓慢地变换到另一个镜头, 常延续十几或几十帧。

为对视频进行有效的组织, 需要将视频分解为其基本单元——镜头, 一个镜头是由一个摄像机连续拍摄得到的时间上连续的若干帧图组成^[1]。镜头边界的转换方式中突变占了大多数。针对突变的特点, 国内外许多研究者提出了不同的镜头边界检测算法^[2-10]。在镜头的边界检测中, 突变的检测现在已经不再是一个难题, 但渐变的检测还需要大量的工作才能达到更好的效果。为此, 本文提出一种基于三步筛选的渐变镜头边界检测算法。

2 基于三步筛选的视频渐变镜头检测

2.1 特征提取

2.1.1 颜色特征

由于渐变的多样性, 很难用一种特征和一个通用的方法检测出所有的渐变, 因此本文提出将多种特征结合, 最大可能地检测出所有的渐变, 并尽可能地去除误检。当叠化和淡入淡出发生时, 中间的过渡帧对比度有下降的趋势, 可以利用亮度方差特征来进行检测, 因为亮度方差特征就反映了帧的对比度特征。为了最大限度地准确地提取视频帧的亮度方差特征, 本文采用累加直方图求出视频帧的亮度均值(HM)和亮

度方差(HV)。若一幅图像的某一特征的一般直方图为:

$$H(P) = (h_{x1}, h_{x2}, \dots, h_{xn})$$

则该特征的累加直方图能根据式(1)计算:

$$G(P) = (G_{x1}, G_{x2}, \dots, G_{xi}, \dots, G_{xn}) \quad (1)$$

其中, $G_{xi} = \sum_{j=1}^i h_{xj}$, j 为整数。从而亮度均值为:

$$HM = \sum_{i=0}^M i \times p(i) \quad (2)$$

其中, M 为直方图柄的个数; $p(i)$ 为第 i 个柄对应的累加值。则亮度累加直方图方差为:

$$HV = \sum_{i=0}^M (i \times p(i) - HM)^2 \quad (3)$$

2.1.2 灰度共生矩阵

尽管对纹理没有正式的定义, 但在直觉上, 它提供了对诸如平滑度、粗糙度和规律性等特性的度量。纹理特征提取的一种有效方法是以灰度级的空间相关矩阵即灰度共生矩阵为基础, 进行二次特征提取。灰度共生矩阵被定义为从灰度为 i 的像素点, 离开某个固定步长位置关系的点上的灰度为 j 的概率, 即 $p(i, j)$ 等于集合:

$$\{(x, y) | f(x, y) = i, f(x + DX, Y + DY) = j; x, y = 0, 1, \dots, N-1\} \quad (4)$$

基金项目: 2010 年重庆市高等教育教学改革基金资助重大项目(101404); 2011 年重庆市教委科学技术研究基金资助项目(KJ113201); 2010 年中国航天科技集团第七研究院青年科技创新基金资助项目([2010]776)

作者简介: 王剑峰(1982—), 男, 讲师, 主研方向: 图像处理; 杜奎然, 工程师

收稿日期: 2011-06-07 **E-mail:** zxr0709@163.com

的元素个数, $i, j = 0, 1, \dots, L-1$, 其中, x, y 是图像中像素坐标; $f(x, y)$ 是该点灰度级; L 是灰度级数目; DX, DY 是偏移量。在实际计算操作中, 反映固定的位置关系的参数 DX, DY 的选择既包括偏移量, 又包括方向的选择。一般方向只取东西、西北-东南、南北、东北-西南 4 个方向。因此, DX, DY 的选择只取以下 4 种值:

$$\begin{cases} DX = d, DY = 0 \\ DX = d, DY = d \\ DX = 0, DY = d \\ DX = -d, DY = d \end{cases} \quad (5)$$

而一幅图像的灰度级的数目一般是 256, 这对于灰度共生矩阵的计算来说太大, 会耗费大量的时间。为了解决这一问题, 本文预先将灰度级压缩至 16 级, 即式(4)中的 L 为 16。矩阵中的每个元素为该行列所代表像素灰度差在某种条件下出现的次数, 由此可知, 此共生矩阵是对称矩阵。共生矩阵的压缩示意图如图 1 所示。

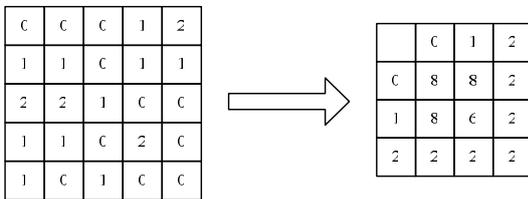


图 1 共生矩阵的压缩示意图

一般来说, 用 4 个最重要的参数来刻画共生矩阵的特性, 即纹理一致性的统计量、纹理方差统计量、纹理熵的统计量、纹理灰度相关性的统计量。纹理特征提取有 3 个参数需要确定:

- (1)扫描图像的方向, 有 $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$ 和 315° 。
- (2)步长可为 1、2、3、4、5 等。
- (3)对共生矩阵特征的计算。

实验证明, 角度 45° 、步长为 1, 计算共生矩阵的纹理方差得到的特征曲线对渐变检测而言是最好的, 本文采用这 3 个参数。

2.1.3 运动特征

在渐变检测中, 镜头内目标快速运动容易带来干扰, 因此误检为突变^[11]。为了更好地解决这个问题, 本文提出了利用运动信息确定阈值, 从而形成自适应阈值。其功能可以总结为: 运动强度越大, 检测为突变的可能性越小^[12]。

在 MPEG-I/II 编码过程中, 视频序列由一个个图像组(Group of Picture, GOP)构成, 一个典型的图像组为 IBBPBBP BBPBBPBB。在算法中, 首先提取各帧宏块的运动矢量(帧内编码宏块可认为运动矢量为 0), 然后将所有的运动矢量分布于 4 个象限。统计运动矢量最多的象限内的宏块个数 max_mb , 它已能反映帧内目标的运动情况, 为了形成自适应阈值, 需要对其进行规一化, 设总的宏块个数为 $total_mb$, 则可得运动强度定义为:

$$MI = \frac{max_mb}{total_mb} \quad (6)$$

被检测为突变的可能性随着运动强度的增强而减弱, 通过对多个视频进行分析, 对每一个运动强度下找出理想的阈值, 得到拟合数据。从拟合数据形状上看大致与高斯函数曲线类似, 因而采用高斯函数进行高斯曲线拟合, 事实也证明通过高斯拟合比多项式拟合等其他拟合方式能获得更小的拟

合误差, 最后到阈值函数为:

$$T_c = \alpha e^{-\frac{MI}{\beta}} \quad (7)$$

实验结果表明当 α 值为 0.15, β 值为 0.1 时效果最佳。

当第 t 帧相对应的 Fisher 判别式的值低于 T_c , 则被标记为突变, 否则不是突变点, 是渐变的切入点, 即:

$$P(x, y, t) = \begin{cases} CUT & F(t) \leq T_c \\ non_CUT & F(t) \geq T_c \end{cases} \quad (8)$$

2.2 检测策略

由前文可知, 叠化和淡入淡出占到渐变的 95% 以上, 而叠化和淡出淡入发生时, 过渡帧表现为对比度明显下降, 因此, 最好利用视频帧的颜色信息即亮度和方差进行检测。本算法采用 3 步检测法, 主要思想如下:

- (1)通过有限状态机对亮度方差特征进行抛物线的搜索, 确定候选渐变。
- (2)使用纹理特征(灰度共生矩阵)进一步确认, 若无明显变化, 则丢掉候选渐变; 反之进入(3)。
- (3)使用运动特征再次检测, 当运动强度过大, 丢掉该候选渐变。

2.2.1 亮度方差曲线渐变检测

2 个镜头之间的过渡帧可以看作是前后 2 个镜头的线性组合。因此, 对于亮度特征有:

$$I_z = \alpha I_x + (1-\alpha)I_y \quad (9)$$

其中, $\{\alpha_i\} \in (0,1)$ 且递减。对于亮度方差特征有:

$$V_z = \alpha^2 \sigma_x^2 + (1-\alpha^2) \sigma_y^2 \quad (10)$$

其中, $\sigma_z^2, \sigma_x^2, \sigma_y^2$ 分别是过渡帧、起始帧和终止帧的亮度方差值。由上式可知, 当发生渐变时, 帧亮度方差特征值呈二次曲线形状, 但反之则不然, 即当亮度方差曲线出现抛物线时并不一定是由渐变引起, 也可能因目标运动而起。在此将曲线中的抛物线检测出来后作为候选的渐变, 再通过第 2 和第 3 阶段的特征消除候选中的误检, 得到准确的渐变。对于亮度方差曲线中的抛物线识别, 算法引入了有限状态机进行识别。有限状态机包括 4 个状态: 初始状态(S0), 上升状态(S1), 下降状态(S2), 终止状态(S3)。状态的定义如表 1 所示, 状态之间的转换关系如图 2 所示。

表 1 状态的定义

状态名	状态定义
S0	初始状态
S1	曲线下降状态
S2	曲线上升状态
S3	已检测到波谷
S4	叠化确认状态

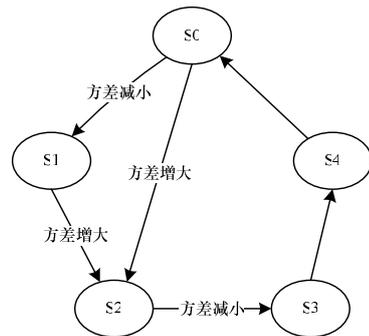


图 2 状态之间的转换关系

为了防止误检出毛刺, 在搜索过程中, 需要附加 2 个限

制条件:

(1)发生状态转换时, 还应考虑在前几帧时是不是仍然保持这种趋势。如果短暂的趋势变化不应发生状态变化。

(2)为了避免将微小的变化也检测出来, 在上升或下降时, 变化的幅值须加以限制。

2.2.2 纹理特征渐变检测

经过前阶段的筛选, 去除部分误检, 同时保留真正的渐变, 但还是存在部分误检, 这些误检从本质上讲很难仅用颜色特征消除它们。因此, 需要借助纹理特征进一步去除。渐变前后的纹理特征应呈现较大变化, 在前面提取的纹理特征曲线上, 比较渐变候选中的 X 和 Y 点的纹理值 λ_x 和 λ_y , 当 $|\lambda_x - \lambda_y| > T_t$ (实验证明 T_t 选择 20 为最佳) 时, 该渐变候选被保留, 否则丢弃。

2.2.3 淡出淡入检测

值得注意的是, 渐变中的淡出淡入中的过渡帧并不是由 2 个镜头线性组合而成的, 不满足式(9)和式(10), 不能用判别叠化的方法来进行判别。不过, 淡出淡入中一定会有单色的过渡帧, 单色帧的方差特征值几乎接近于 0, 这一点从曲线中可以看出。所以对波谷谷点的方差特征值用一个固定门限进行判断就可以精确地检测出淡出淡入, 即:

$$X \sim Y = \begin{cases} FOI & \sigma_z^2 < T_F \\ Non_FOI & \sigma_z^2 > T_F \end{cases} \quad (11)$$

在通过上述方式找出候选的渐变后, 结合这些帧的纹理特征。当纹理特征出现较大变化时, 便留下这些渐变候选。

在最后的第 3 个阶段中, 求出这个时间段内帧的平均运动强度, 然后与该时间段前后帧的运动强度进行比较, 当这个平均运动强度大大超过前后帧的运动强度时, 认为这些特征异常是因运动而引起, 否则认为是真正的渐变。

3 实验结果及分析

实验选取了大量的视频来测试算法性能。视频来源主要为 TRECVID2006 和 TRECVID2007 中的各类视频: 其中 A、B 为新闻, 时间分别为 1 min; C 为电视纪录片, 时间为 45 s; D、E 为电影片段(主要是一些汽车运动场景), 时间为 50 s; F 为动画, 时间为 25 s。为了进行比较, 本文对以上实验数据分别采用文献[5]中的宏块判别法、文献[9]中的基于信息理论的联合熵方法、文献[13]的基于特征信息的边界检测法和本文提出的方法分别进行镜头检测。镜头检测算法一般用查全率(R)和查准率(P)来进行性能评估。为了综合量化算法的性能, 定义综合评价指标 F :

$$F = 2 \times \frac{P \times R}{P + R} \quad (12)$$

本文算法和文献[5,9,13]的检测结果如表 2 所示。

表 2 4 种算法的检测结果对比

视频	本文算法			文献[5]算法			文献[9]算法			文献[13]算法		
	P	R	F	P	R	F	P	R	F	P	R	F
A	0.84	0.79	0.81	0.71	0.69	0.66	0.78	0.71	0.74	0.58	0.63	0.51
B	0.87	0.81	0.83	0.77	0.71	0.73	0.67	0.68	0.67	0.68	0.75	0.71
C	0.71	0.63	0.66	0.63	0.55	0.49	0.59	0.56	0.57	0.61	0.61	0.61
D	0.82	0.83	0.82	0.69	0.70	0.69	0.68	0.71	0.69	0.64	0.70	0.66
E	0.79	0.83	0.81	0.73	0.72	0.72	0.72	0.64	0.67	0.65	0.63	0.63
F	0.77	0.84	0.80	0.66	0.69	0.67	0.61	0.62	0.61	0.70	0.67	0.68

从表 2 可以看出, 本文算法相对于文献[5,9,13]提出的方

法在检测性能上都较好, 主要是由于有限状态机几乎能搜索到所有因渐变引起的方差曲线谷点, 进一步结合亮度方差及纹理、运动特征可去除一些不是因渐变引起的波谷, 最大限度地防止漏检。在对视频段 C 的检测中, 几种算法的检测性能都较低, 主要是由于视频段 C 是一些纪录片, 叠化发生得较多, 镜头的拍动较多, 光照较复杂; 在对视频段 A、B、D、E 的检测中, 体现了对运动场景的较好的检测效果, 查全率和查准率都在 80% 以上; 本算法发生误检主要是由大物体快速切换引起的; 视频中的淡出淡入变化用前述的方法几乎能全部准确无误地检测出。另外本文检测算法的整体性能较其他几种渐变算法平均提高了约 25%。

4 结束语

本文通过三步筛选策略对渐变镜头进行了检测, 在查全率和查准率方面较传统的方法都有了明显提高, 适合一些对实时性要求较高的视频检索系统。在今后的研究中, 将进一步对帧速较快、码流较大的视频素材进行检测研究。

参考文献

- [1] Hanielic A. Shot-boundary Detection: Unraveled and Resolved[J]. IEEE Trans. on Circuits and Systems for Video Technology, 2002, 12(2): 90-105.
- [2] Quenot G M, Moraru D, Besacier L. CLIPS at TRECVID: Shot Boundary Detection and Feature Detection[C]//Proc. of TRECVID'03. Gaithersburg, Maryland, USA: [s. n.], 2003.
- [3] Zhang Jiang, Kankanhalli A, Smoliar S. Automatic Partitioning of Video[J]. Multimedia Systems, 2008, 1(1): 10-28.
- [4] Pickering M J, Ruger S M. Multi-timescale Video Shot-change Detection[C]//Proc. of the 10th Text Retrieval Conference. Gaithersburg, Maryland, USA: [s. n.], 2001.
- [5] Li Zhenming, Jiang Jianmin, Xiao Guoqiang, et al. An Effective and Fast Scene Change Detection Algorithm for MPEG Compressed Videos[C]//Proc. of ICIP'06. Porto, Portugal: [s. n.], 2006.
- [6] Bai Xuesheng, Zhang Ziyin, Xu Guangyou, et al. Analysis of Digital Video Effect Shot Transition Detection Algorithm[J]. Journal of Software, 2002, 13(7): 1278-1283.
- [7] Zabih R, Miller J, Mai K. A Feature-based Algorithm for Detecting and Classifying Scene Breaks[C]//Proc. of ACM Multimedia Conference. San Francisco, USA: [s. n.], 2007.
- [8] Hirzalla N. Automatic Cut and Camera Operation Detection for Video[C]//Proc. of International Conf. on Consumer Electronics. Rosemont, USA: [s. n.], 2005.
- [9] 施游, 黄少年, 张友生. 基于交互信息量和联合熵的镜头检测算法[J]. 计算机工程与应用, 2006, 42(30): 54-56.
- [10] 李江, 孙立军. 基于凸包裁剪的行人视频检测算法[J]. 计算机工程, 2010, 36(2): 173-175.
- [11] Bell A J, Sejnowski T J. The Independent Components of Natural Scenes are Edge Filters[J]. Vision Research, 2008, 37: 3327-3338.
- [12] Han K J, Tewfik A H. Eigen-based Video Segmentation and Indexing[C]//Proc. of ICIP'07. San Antonio, Texas, USA: [s. n.], 2007.
- [13] 李玉峰, 赵政. 基于特征信息的镜头检测研究[J]. 计算机工程与应用, 2009, 45(6): 168-169.

编辑 顾姣健