

一种稳健的新闻视频镜头检测方法

赵 镭, 彭天强, 刘胜涛, 李其斌

(解放军 66407 部队, 北京 100093)

摘 要: 根据人眼的视觉特性, 本文提出了一种基于 $K-L$ 变换的新闻视频镜头检测方法. 首先, 通过对原始 RGB 空间进行 $K-L$ 变换, 得到由主轴构成的图像参数模型; 其次, 运用基于滑动窗口的自适应阈值方法来比较相邻帧间参数模型的差异, 得到候选的镜头切换帧; 再次, 定义以候选帧为首帧的小窗口, 计算窗口内各帧与首帧的参数模型差异, 得到随时间变化的曲线, 称之为参数模型差异曲线 (Parameter Model Variation Curve), 简称 PMV 曲线; 最后, 通过分析曲线的特性进行镜头检测, 完成镜头切换方式的分类, 同时滤除掉闪光灯镜头. 实验结果和性能比较表明, 针对国内外新闻视频中出现的各种类型镜头切换算法取得了良好的检测效果, 并且对目标的快速运动、光照变化等具有一定的鲁棒性.

关键词: 镜头检测; 突变; 渐变; $K-L$ 变换; 视频检索

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2009) 02-0325-04

Video Shot Transition Detection Based on K-L Transform

ZHAO Ke ke, PENG Tian qiang, LIU Sheng tao, LI Qi bin

(PLA with Designation of 66407, Beijing 100093, China)

Abstract: We propose a new technique of shot transition detection based on K-L transform for news video according to the characteristics of human visual system. Firstly, the algorithm uses K-L transform of the RGB color space to produce image parameter modal that is consisted of eigenvectors. Secondly, the algorithm calculates the frame distance with the first and second principle axis' s direction differences in the two adjacent frames and weighting coefficients. Thirdly the algorithm uses a fixed length sliding window of current frame to detect candidate transition. Lastly, through calculating the differences between the head frame and the other frames in the local window, which is defined as the candidate transition frame, we get and analysis the model variance curve to locate and classify the shot transitions and at the same time disposes the flash shot. The experimental results show that the proposed algorithm outperforms the traditional methods, and is robust to illumination changes or camera motion.

Key words: shot boundary detection; abrupt transition; gradual transition; K-L transform; video retrieval

1 引言

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, 迫切需要自动地建立索引结构, 以进行新闻视频的组织、管理和检索^[1]. 镜头作为视频结构中的最小单元, 镜头检测成为新闻视频分析和检索的基础性工作. 镜头之间的转换方式主要有两大类: 突变 (Abrupt Transition) 和渐变 (Gradual Transition), 突变是镜头间的突然变化, 常在两帧图像间完成, 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.

目前, 突变镜头检测方法已比较成熟, 主要通过提取图像的低级特征, 包括像素、分块统计特性、颜色直方图、边缘等来衡量帧间差异, 当帧间差大于某个阈值时, 则认为有镜头突变^[2]. 其中最典型的是基于颜色直方图

的方法^[3], 在它的基础上, 又出现了很多改进算法^[4-7]. 但是, 直方图不含有空间信息, 如果具有不同目标的场景有近似的灰度或者颜色分布就容易造成漏检; 当镜头中存在目标快速运动或者光照变化的时候, 直方图比较方法也会放大这种差异, 导致误检.

渐变镜头类型多样, 主要包括淡入、淡出和叠化. 由于帧间差异变化不明显, 并且镜头内存在各种相机和目标的运动, 目前还没有一种有效、通用的检测方法. 文献 [8] 利用视频编辑模型来检测渐变, 还有一些其他的检测方法, 比如边缘变化检测^[9]、亮度方差变化检测^[10]、联合熵检测^[11]等. 由于这些方法计算量大, 只对某些特殊类型的渐变有效, 因此缺乏通用性.

新闻视频是一类特殊的视频, 存在着大量闪光灯事件, 传统的方法容易将它们与镜头切换相混淆, 造成误检. 稳健的镜头检测方法能够克服视频中目标运动和光照变化带来的影响, 真正地体现出镜头切换在视觉内容

上的差异,同时具有较低的计算复杂度.一幅图像可以由一组基函数来描述,内容上相似的图像对应的基函数也是相似的^[12],换句话说,如果图像内容发生了变化,基函数也会发生相应的变化.利用这一特性,本文提出了一种基于K-L变换的新闻视频镜头检测方法.新方法改变了传统的依靠图像低级特征进行镜头检测的方式,将各类型的镜头切换统一于一种模型下,避免了模型选择和繁琐的参数调整,不仅可以检测出突变和渐变,还能够对渐变的类型做出分类.实验结果表明,该算法是有效的,并且该算法对目标的快速运动、光照变化等具有一定的鲁棒性,性能优于基于直方图的双重阈值法^[7]和基于信息理论的方法^[11].

2 K-L 变换

K-L 变换,又称 Hotelling 变换或主分量方法,是一种基于目标统计特性的最佳正交变换,广泛应用于图像压缩、图像重建、模式识别等领域^[13].从降维的角度上讲,这一方法就是寻找由观察数据向量组成的向量组中主要分量的子集,使变换矢量能量更集中.相应的基向量组满足正交性,由它定义的子空间去除了数据的相关性,因此,基向量组就构成了描述图像空间的参数模型,包含了原始图像的空间信息,集中表达了图像的某些细节特征.

K-L 变换的数学原理如下:

设 n 维随机矢量 $x = (x_1, x_2, \dots, x_n)^T$, 其均值矢量 $\bar{x} = E[x]$, 相关阵 $R_x = E[xx^T]$, 协方差阵 $C_x = E[(x - \bar{x})(x - \bar{x})^T]$, x 经标准正交矩阵 T^T 正交变换后成为为矢量 $y = (y_1, y_2, \dots, y_n)^T$, 即

$$y = T^T x = (t_1, t_2, \dots, t_n)^T x \quad (1)$$

这里,标准正交矩阵 $T = (t_1, t_2, \dots, t_n)$ 由求解协方差阵 C_x 的特征向量得到,从坐标变换的角度来看,将特征向量 t_i 视为第 i 主轴, $i = 1, 2, \dots, n$, 最佳变换就是依据 C_x 寻找最佳主轴的变换,第 i 主轴与 C_x 的第 i 个最大特征值对应,这个特征值正是变换后分量 y_i 的方差,反映了该方向上的数据变化剧烈程度.

3 帧间差值度量

本文是将图像帧内的像素作为观察数据向量,在 RGB 颜色空间用三维矢量 (r, g, b) 表示,接着计算这 m' 个三维矢量的协方差矩阵, m' 为图像帧的像素总数,然后利用 K-L 变换求取特征值以及特征向量,这些特征向量,也称为基向量就构成了描述图像的参数模型.根据人眼的视觉特性,一个镜头内的图像帧如果没有太大的视觉差异,它们的参数模型中对应特征向量的方向是基本一致的.计算对应特征向量之间的方向差异,并对不同的特征向量设置不同的加权系数,以加权后的平均距离作为两帧图像之间的差值.这样计算帧间差值,更

好地体现了不同镜头的帧图像间的视觉内容差别,并且由于只对 3×3 的协方差矩阵进行 K-L 变换,计算速度也可以接受.综合考虑到计算的复杂度和检测结果的精度,我们只利用第一、第二主轴进行计算.

相邻两帧对应主轴的方向差异,可以用两个向量夹角余弦来表示.这样计算简单,满足计算速度的需要.

$$S(x, z) = (x^T \cdot z) / (\|x\| * \|z\|) \quad (2)$$

x 和 z 分别为相邻帧对应主轴, $S(x, z)$ 表示了主轴之间的方向差异.计算出相邻两帧对应第一、第二主轴之间的方向差异,分别记为 S_1 和 S_2 , 则两帧差值由加权系数和主轴方向差异计算得到:

$$DF = w_1 S_1 + w_2 S_2 \quad (3)$$

这里通过试验,取 $w_1 = 0.7, w_2 = 0.3$.

4 镜头切换检测

在传统的镜头检测方法中,常先检测切变再检测渐变.在渐变检测中,也常先检测某些类型的渐变,再检测其他类型的渐变.这就带来了选择不同的检测模型和复杂的参数设置的问题,因此,我们将不同类型的镜头检测统一到同一模型下,提出了一种新的二阶段镜头切换检测策略.在第一阶段,利用基于因果的滑动窗口法检测出候选边界帧,在第二阶段,定义以候选镜头切换边界为首帧的小窗口,获得 PMV 曲线并分析其特性进行镜头检测.

4.1 候选边界帧检测

我们注意到,一个镜头内部的各帧之间的差异不会太大,相邻帧间差值相对稳定在某一个范围之内,即围绕某个均值上下波动,而当发生镜头转换时,无论是何种类型,相邻帧间差值的这种稳定性将被打破,据此,在进行候选镜头边界检测时,只考虑当前帧之前有限长度内的帧间差值信息.这个长度通常是前一镜头结束后的第 1 帧至当前帧前一帧,即第 $l-1$ 帧,可以认为这是一个可变长度的滑动窗口.根据第 3 节中介绍的方法计算这一窗口内的相邻帧间差值,并得到平均值 $Mean$, 如果当前计算得到的相邻帧间差值与这一平均差值之间差异达到一定程度,就认为第 $l-1$ 帧为候选边界帧,实际上它对应着一个镜头的可能结束帧.这种改进,实际上是将滑动窗口检测算法的双边窗口变成了单边窗口,我们称之为基于因果的滑动窗口法^[14].

4.2 镜头切换类型检测

根据对不同类型新闻视频的大量统计,发现突变、短渐变、淡入淡出、叠化这 4 种类型占到了总数的 95% 以上,而除此之外的镜头切换,如旋转、擦除、上拉、下拉等等,只占了较少的比例.另外新闻视频中还存在着闪光灯镜头,对镜头检测造成干扰.我们认为,无论是突变,还是各种类型的渐变,从切换结果来看,镜头内容发

生了较大的变化,这也是与闪光灯镜头的本质区别;从切换过程来看,镜头内容发生变化的持续时间不一致.因此,在确定了候选边界帧之后,我们定义了以候选帧为首帧的小窗口,窗长为 N ,计算窗口内各帧与首帧的参数模型差异,记为 $DF_1, DF_2, \dots, DF_{N-1}$,得到随时间变化的曲线,称之为参数模型差异曲线(Parameter Model Variation Curve),简称 PMV 曲线.图 1 给出了不同类型镜头切换的 PMV 曲线分布情况.

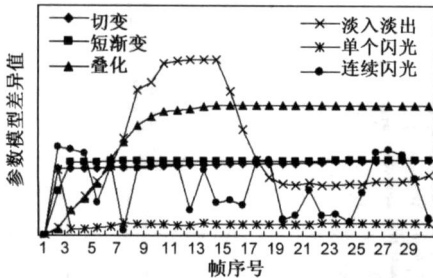


图 1 不同类型镜头 PMV 曲线

可以看出,我们用不同颜色的 PMV 曲线来代表不同类型的镜头切换和闪光灯镜头.对于镜头切换,PMV 曲线分布基本上都是先有一个上升的过程,然后稳定在某一较高值的附近,就像“梯形平台”的一个边,然而不同类型的 PMV 曲线的上升持续时间不同,从小到大依次为突变、短渐变、叠化和淡入淡出,而且代表淡入淡出镜头的曲线在尾端会出现下降趋势,这就为我们区分不同的镜头切换提供了很好的判断条件.对于容易造成检测干扰的闪光灯镜头,我们将其细分为单个闪光和连续闪光两种情形.观察单个闪光 PMV 曲线,发现当出现上升后,迅速衰减到某一较低值附近并保持稳定,这与闪光帧前后的图像内容无变化这一物理特性是一致的,也保证了能够将单个闪光镜头和容易混淆的突变和条形短渐变能够区分开.连续闪光通常有多个闪光灯叠加,造成内容上没有太大变换的图像帧发生了极不规则的亮度变化,反映在 PMV 曲线上是差异值高低随时间交替变化,在闪光结束后稳定在某一较低值,因此检测出曲线上不规则变化区域,再加上区域前后帧参数模型差异的判断,便能分析出连续闪光的存在.

这样,镜头切换类型检测统一到了一种模式下,即对 PMV 曲线的特性进行分析,区分不同类型的镜头切换,并有效排除闪光灯镜头的干扰.我们提出上升趋势持续帧数(Ascend Frame Number,简称 AFN)特征,其计算公式为:

$$AFN = k \quad \text{若 } DF_i > a \times DF_{i-1}, 1 \leq i \leq k \quad (4)$$

其中 $DF_0 = 0$, a 为一预先设定的比例系数.由 AFN 的大小 k 决定下一步的处理:

(1) 当 $k = 1$ 时,比较窗口内其它 $DF_i, 2 \leq i \leq N - 1$ 与帧差均值 $Mean$ 的大小,若 $DF_i \geq b \times Mean$, b 为一预

先设定的比例系数,则认为该处是一个突变镜头.

(2) 当 $k = 2$ 时,比较窗口内其它 $DF_i, 3 \leq i \leq N - 1$ 与 DF_k 的大小,若它们都与 DF_k 相近,则认为该处是一个短渐变.

(3) 当 $3 \leq k \leq N - 1$ 时,比较窗口内其它 $DF_i, 4 \leq i \leq N - 1$ 与 DF_k 的大小,若它们都与 DF_k 相近,则认为该处是一个叠化镜头;若它们满足(1) $DF_i, k < i < p$ 与 DF_k 相近;(2) $DF_i \leq a \times DF_k, p < i < N$,则认为该处是一个淡入淡出镜头.

(4) 不符合以上条件的情况,包括闪光灯镜头、噪声等,按我们以前的工作^[15]中的算法进行处理.

5 实验结果与分析

为了验证算法的有效性,我们从自己建立的新闻视频数据库中选取比较有代表性的几类新闻节目作为实验对象.其中包括:1 个小时 CCTV1 的新闻联播,其镜头边界以突变为主,基本没有渐变;1 个小时 CCTV4 的中国新闻,其镜头边界以突变和短渐变为主,其它渐变类型很少;1 个小时美国 CNN 新闻和 1 个小时的英国 BBC 新闻,其镜头边界复杂,上面介绍的多种类型全部包括.以上测试数据约为 4 个小时,共计有 2811 个镜头,其中实际的主持人镜头为 106 个.在测试前我们手工标注出了样本的所有镜头边界,作为方法检测结果的标准参照.

为了进行比较,我们对以上实验数据分别采用文献[7]中的二次帧差法、文献[11]中的基于信息理论的联合熵方法和本文提出的方法分别进行镜头检测.用查准率和查全率两个指标来评估镜头切换的检测结果.

$$\text{查准率} = \frac{\text{正确检出数}}{\text{正确检出数} + \text{误检数}}$$

$$\text{查全率} = \frac{\text{正确检出数}}{\text{正确检出数} + \text{漏检数}}$$

三种方法对各类新闻视频的检测结果如下表 1 所示,同时在图 2 给出了各方法的平均查全率和查准率.

表 1 不同检测算法的检测结果对比

新闻视频	检测算法	镜头数	检出数	误检数	漏检数	查全率 %	查准率 %
CCTV1	文献[7]	702	712	51	41	94.2	92.8
	文献[11]		706	41	31	95.5	95.0
	本文方法		703	19	20	97.3	97.2
CCTV4	文献[7]	760	730	140	170	77.7	80.9
	文献[11]		770	110	100	86.8	85.7
	本文方法		765	27	32	96.4	95.8
CNN	文献[7]	932	1100	378	210	77.5	65.7
	文献[11]		1000	293	225	75.9	70.7
	本文方法		970	145	107	88.6	85.1
BBC	文献[7]	925	1086	366	205	77.8	66.3
	文献[11]		992	280	213	77.0	71.8
	本文方法		930	137	132	85.8	85.3

分析表 1、图 2 的结果可以看出,本文提出的基于

$K-L$ 变换的镜头检测算法改变了利用图像低级特征来比较帧间差异的检测方式,能够对多种类型的镜头切换进行检测,

而且查全率和查准率也有较大的提高,相比文献[7]中的二次帧差法,本文算法将查准率平均提高了 14.4%,查全率提高了 10.2%,相比文献[11]中的联合熵法,本文算法将查准率平均提高了 10.0%,查全率提高了 8.4%,这对于视频媒体的基于内容的实时检索和处理是大有好处的。

6 结论

本文提出了一种稳健的基于 $K-L$ 变换的镜头检测方法,该方法采用了分级的思想进行检测。首先,不用计算图像的低级特征,而是通过 $K-L$ 变换构造图像的参数模型,根据当前帧前面有限数量的特征向量方向差异的平均值自适应地计算镜头边界检测阈值,通过计算特征向量的方向差异有效地消除了边界检测对镜头内目标、摄像机运动以及光照变化的敏感性,来获取候选镜头边界帧。然后,以候选镜头边界帧为首帧设定窗口,构造 PMV 曲线,通过提取 AFN 特征和窗口内参数模型特征差异间的比较,完成镜头切换的确认和分类。实验证明了该方法的有效性和适应性,适合于应用在一些需要对视频进行实时处理的场合。

参考文献:

- [1] 章毓晋. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003. 5-12.
- [2] A Hanjalic. Shot boundary detection: unraveled and resolved[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(2): 90-105.
- [3] JS Boreczky, LA Rowe. Comparison of video shot boundary detection techniques[J]. Journal of Electronic Imaging, 1996, 5(2): 122-128.
- [4] G M Quenot, Daniel Moraru, Laurent Besacier. CLIPS at TRECVID: Shot boundary detection and feature detection[A]. In Proceedings of the TRECVID 2003 Workshop[C]. Gaithersburg, Maryland, USA, 2006. 35-40.
- [5] M J Pickering, SM Ruger. Multi timescale video shot change detection[A]. In NIST Special Publication 500-250: Proceedings of the Tenth Text Retrieval Conference[C]. Gaithersburg, Maryland, USA, 2001. 275-278.
- [6] A Smeaton, W Kraaij, P Over. TRECVID 2003 An introduction[A]. In Proceedings of the TRECVID 2003 Workshop[C].

Gaithersburg, Maryland, USA, 2003. 1-10.

- [7] 周艺华, 曹元大, 张洪欣. 一种基于二次帧差的突变镜头检测方法[J]. 计算机工程与应用, 2005, 31(6): 22-25.
Zhou Yi hua, Cao Yuan da, Zhang Hong xin. A quadratic difference based method for abrupt shot change detection[J]. Computer Engineering and Applications, 2005, 31(6): 22-25. (in Chinese)
- [8] 白雪生, 张子银, 等. 数字视频特技镜头转换检测算法的分析[J]. 软件学报, 2002, 13(7): 1278-1283.
Bai Xue sheng, Zhang Zi yin, et al. Analysis of digital video effect shot transition detection algorithm[J]. Journal of Software, 2002, 13(7): 1278-1283. (in Chinese)
- [9] R Zabih, J Miller, K Mai. A feature based algorithm for detecting and classifying scene breaks[A]. Proc. ACM Multimedia 95[C]. San Francisco, CA, November, 1993. 189-200.
- [10] N Hirzalla, A Karmouch. Automatic cut and camera operation detection for video[A]. Proceedings of International Conference on Consumer Electronics[C]. Rosemont, IL, USA, 1995. 290-291.
- [11] 施游, 黄少年, 张友生. 基于交互信息量和联合熵的镜头检测方法[J]. 计算机工程与应用, 2006, 42(30): 54-56.
Shi You, Huang Shaonian, Zhang You sheng. A mutual information and joint entropy based method for shot change detection[J]. Computer Engineering and Applications, 2006, 42(30): 54-56. (in Chinese)
- [12] AJ Bell, TJ Sejnowski. The independent components of natural scenes are edge filters[J]. Vision Research, 1997, 37(23): 3327-3338.
- [13] 孙即祥, 等. 模式识别中的特征提取与计算机视觉不变量[M]. 北京: 国防工业出版社, 2001. 57-59.
- [14] 彭波, 李弼程. 一种因果的突变镜头检测方法[J]. 计算机工程与应用, 2004, 40(5): 91-92.
Peng Bo, Li Bi cheng. A cause and effect algorithm for cut shot change detection[J]. Computer Engineering and Applications, 2004, 40(5): 91-92. (in Chinese)
- [15] 彭天强, 李弼程. 一种抗闪光灯的新闻视频镜头检测方法[J]. 信息工程大学学报, 2007, 8(4): 483-486.
PENG Tian qiang, Li Bi cheng. Effective news video shot detection method for flash events[J]. Journal of Information Engineering University 2007, 8(4): 483-486. (in Chinese)

作者简介:



赵课课 男, 1981 年出生于安徽省淮北市, 1999 年进入信息工程大学信息工程学院。硕士研究生, 研究方向为基于内容的视频检索, 模式识别。E-mail: zkk11@163.com