

一种基于有限自动机的渐变镜头检测算法^{*})

佟子健 袁进辉 郑武杰 林福宗 张 钹

(清华大学计算机科学与技术系智能技术与系统国家重点实验室 北京 100084)

摘 要 渐变镜头检测算法分为两个方面:渐变边界帧的判定和边界帧的组合。前者判断某一帧是否符合渐变边界帧的条件,后者判断一段包含边界帧的视像是否是渐变。以往的算法侧重解决边界帧的判定,忽视了边界帧的组合。本文定义了渐变检测容忍度的概念,并提出了一种基于有限自动机的渐变镜头检测方法,利用了自动机多状态的记忆性,提高了算法的适应性和鲁棒性。在 TRECVID2004 的 SBD 项目中,本渐变镜头检测系统取得了渐变检测性能第一的好成绩。

关键词 镜头边界检测,边界帧组合,容忍度,有限自动机

A New Approach for Gradual Transition Detection Based on Finite-State Automata

TONG Zi-Jian YUAN Jin-Hui ZHENG Wu-Jie LIN Fu-Zong ZHANG Bo

(State Key Laboratory of Intelligent Technology & Systems, Department of Computer Science & Technology, Tsinghua University, Beijing 100084)

Abstract Video shot boundary detection is an essential step for video browsing and retrieval. Almost every shot boundary detection algorithm detects cut translation without difficulty, but gradual translation is regarded as a hard problem. Gradual translation detection comprises boundary frame determination and boundary frames combination. The former is to determine that a frames is or isn't a gradual boundary frame, and the latter is to determine that a segment which composes boundary frames and normal frames is or isn't a gradual transition. The approaches focus always on boundary frame determination, and ignore boundary frames combination before. In this paper, we define the tolerance in gradual transition detection, a new conception about boundary frames combination. And we present a new approach for Gradual Transition Detection, which uses Finite-State Automata to obtain proper adaptability and good performance. In TRECVID2004, the approach obtains the best performance in gradual transition detection.

Keywords Shot boundary detection, Boundary frame combination, Tolerance, Finite-state automata

1 引言

镜头是视像序列的基本元素,其边界检测是视像内容分析和基于内容的视像检索的基础。镜头的边界可分为两类:突变和渐变。目前,突变的检测效果比较好,而渐变边界的检测效果并不理想^[1]。这是因为,相比突变边界,渐变边界有如下特点:

- 1) 长度的不确定性。渐变的长度从几帧到上百帧不等,这要求检测算法能够考察整个变化的过程。
- 2) 变化类型的多样性。渐变可以分为融入融出、淡入淡出、划入划出等,这要求检测算法有较强的鲁棒性。
- 3) 变化的平缓性。渐变的变化比较平缓,帧间差异比较小,甚至和正常帧间没有明显差异,这要求检测算法能容忍渐变过程中少量平缓变化的帧。

渐变检测的算法主要分为两个方面:判断视像中的一帧是否是渐变边界帧,称为边界帧的判定;判定一段包含边界帧的视像是否是渐变,称为边界帧的组合。以往的检测算法侧重于解决边界帧的判定问题,却没有考虑如何将一组边界帧组合成渐变边界。本文将分这两个方面阐述我们的工作。第2节分析双阈值方法和自适应阈值方法,提出了我们的基于运动的自适应方法。第3节提出渐变检测容忍度的概念,并

且给出了具有较好容忍度的基于有限自动机的渐变镜头检测算法。第4节提供算法的评价数据、标准和测试结果。

2 边界帧判定

判断一个帧是否是渐变边界帧,是渐变检测的基础,也是以往研究的重点。

2.1 前人的工作

早期的检测渐变边界的方法是双阈值法^[2]。双阈值法针对渐变边界的检测设定了一个门限阈值 T_l 和一个长度控制阈值 T_h 。当帧间点对点灰度差值高于 T_h 时,判定为突变。渐变过程中的每一个帧间差都必须高于 T_l 。将遇到的第一个帧间差高于 T_l 且低于 T_h 的帧判定为渐变的起始帧;然后在比较帧间差和 T_l 的同时计算和起始帧间的多帧差。当多帧差高于 T_h 时,判定为渐变结束。由于视频中渐变的种类、变化速度和变化程度各不相同,所以很难确定一个能适应各种渐变的固定阈值 T_l 。在实际应用中,为了适应渐变的多样性和平缓性特点,系统只能降低 T_l 。而与此同时,为了防止渐变的无限蔓延,需要 T_h 控制渐变的长度。但是,这种使用长度控制阈值 T_h 截断渐变边界的方法过于草率,而且无法找到一个合适的阈值。

在双阈值方法的基础上,文[3]提出了基于滑动窗口的自

^{*}) 本课题得到国家自然科学基金(60135010, 60321002)、国家重点基础研究(973)(2004CB318108)资助。

适应阈值。在被检测帧的附近设置一个长度为 N 的滑动窗口,根据窗口内帧的特征值确定阈值 T_i 。自适应阈值 T_i 的方法提高了渐变门限阈值 T_i 的适应性,无需使用 T_n 控制渐变长度。不过,由于要求渐变中每一帧的帧间差都要高于门限阈值 T_i ,所以如果在渐变过程中有少数一两帧的帧间差低于 T_i ,这种方法就会将渐变截断。

2.2 我们的边界帧判定方法

渐变检测算法需要解决的一个主要问题是边界帧的判定。在帧图像的 RGB 空间上定义帧间差:对每一帧图像的 R, G, B 三维分别做直方图统计,每一维 16 个量化等级,用直方图最小级差法计算各维的差值,最后取平均值作为帧间差(公式 1、2)。对于帧 F_i, R_i, G_i, B_i 分别表示它的 RGB 三维直方向量。对于两帧 F_i 和 F_j ,

$$\begin{cases} D_R = 1 - \frac{\sum_{k=1}^N \min(R_i(k), R_j(k))}{N} \\ D_G = 1 - \frac{\sum_{k=1}^N \min(G_i(k), G_j(k))}{N} \\ D_B = 1 - \frac{\sum_{k=1}^N \min(B_i(k), B_j(k))}{N} \end{cases} \quad (1)$$

$$Diff_{Hist} = \frac{1}{3} (D_R + D_G + D_B) \quad (2)$$

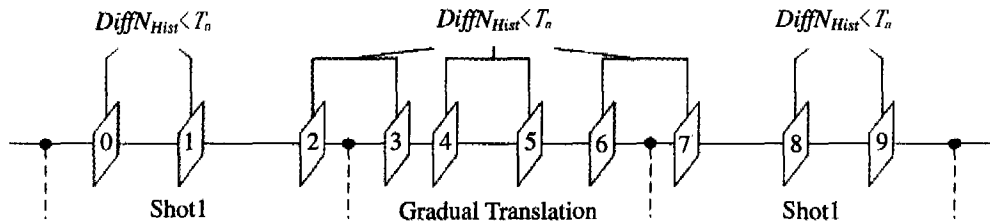


图 1 n 帧差原理示意图

3 边界帧组合

完成了渐变帧的判定后,视像序列可以看作是 0 和 1 组成的序列,其中 0 代表非边界帧,1 代表边界帧。面对这个二值的序列,如何准确地检测出渐变过程是边界帧组合所研究的问题。

3.1 渐变检测的容忍度

由于渐变变化过程的平缓性,渐变过程中经常出现帧间

其中 N 是直方图的维数, D_R, D_G, D_B 分别表示由直方图最小级差法计算出的 RGB 三维的差值, $Diff_{Hist}$ 是帧间差。

我们采用了基于运动的自适应阈值方法。假设当视像中出现大面积高速运动的物体时,帧间颜色直方图差值也相应地提高,因此也应当相应地提高直方图特征的阈值。

$$Diff_{Hist} > \alpha M_X + \beta M_Y \quad (3)$$

其中 $Diff_{Hist}$ 是直方图帧间差, M_X 和 M_Y 是每一帧的宏块(Macro Block)在水平和垂直方向的运动矢量^[5]绝对值的均值,它反映了这一帧相比上一参考帧的运动强度。在每一帧的运动特征的基础上构造阈值,一方面体现了阈值的自适应特性,另一方面降低了剧烈运动对检测的干扰。

除了需要满足式(3)以外,边界帧还需要满足 7 帧差^[4]的门限条件:

$$Diff_{7_{Hist}} > T_{7_{frame}} \quad (4)$$

在图 1 中,帧 0 和 1 在一个镜头内部,镜头场景基本没有变化,帧间差异较小,所以 n 帧差较小,低于 n 帧差阈值 T_n ; 帧 2 和 3 由于分别处在 shot1 和 GT 中,两帧间差异较大,高于阈值;帧 4 和 5 在 GT 内,由于 GT 是一个逐渐积累的变化过程,帧 4 和 5 之间有较大的差异,高于阈值。这样,和渐变相关的帧就被检测出来了。在我们的系统中,取 $N=7$ 。

差异很小的帧,这些帧不满足边界帧的检测条件,所以被检测成非边界帧。但是,这些非边界帧和它前后的边界帧作为一个整体属于同一个渐变过程。以往的检测算法没有考虑这种问题,它要求在渐变过程中的所有帧都能符合渐变边界帧的条件。这样,一个完整的渐变将会被截断,甚至被认为不是渐变过程,使得算法的鲁棒性较差。为了解决这个问题,本文提出了渐变检测中容忍度的概念,并且提出了一种具有容忍度的基于自动机的边界帧组合方法。

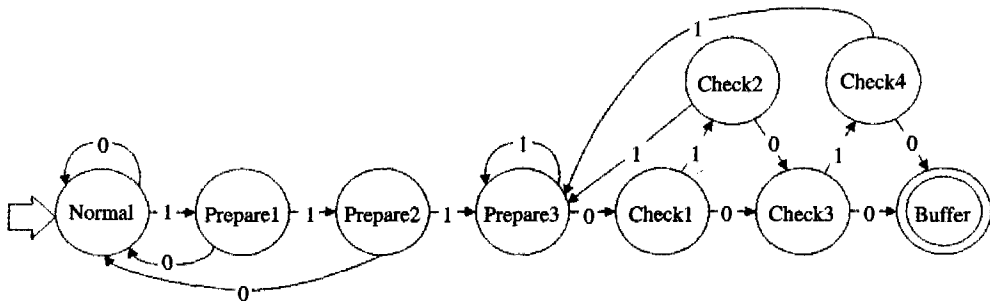


图 2 有限自动机状态图

一个渐变过程中允许最多连续出现 N 个非渐变帧,称 N 为渐变检测的容忍度。

以往的渐变检测算法均要求渐变过程中的每一帧都符合渐变边界帧的条件,所以容忍度为零,算法适应性很差。具有适当的容忍度将会提高算法的适应能力。当然,容忍度并不

是越高越好。

3.2 有限自动机模型

基于自动机的状态模型很好地解决了边界帧合成边界的问题。图 2 中信号 0 表示当前帧为非边界帧,信号 1 表示当前帧为边界帧。Normal 状态是正常状态。Buffer 状态是渐

变声明的准备状态,在该状态,如果待声明的渐变首尾帧差超过阈值 T_{total} ,则声明为渐变,否则返回 Normal 状态。在检测到 3 次较高的帧间差(信号 1)之后,自动机到达 Prepare3 状态。在此之后,自动机可以容忍连续出现 2 个非渐变帧(信号 0),渐变检测的容忍度为 2。

有限自动机有多个具有复杂意义的状态,利用这种状态记忆功能,允许渐变中连续出现 2 个变化平缓的帧(非渐变帧),提高了检测的适应性和鲁棒性。

3.3 渐变确认

当自动机跳转到 Buffer 状态时,确认了待声明渐变的首尾帧。如果首尾帧间差超过阈值 T_{total} ,则声明为渐变,否则返回 Normal 状态。 T_{total} 可以考察渐变变化的整个过程,防止了由于出现一系列剧烈运动所造成的误判。

4 性能评价

4.1 数据集

NIST(美国国家标准技术研究所)为 TRECVID2004 镜头边界检测子项目提供了 12 段用于测试的 mpeg 格式视像^[5],内容是 ABC(美国广播公司)和 CNN(美国有线电视网)1998 年 10~12 月间的新闻视像。突变和渐变在测试数据中分布的比例分别为 57.7%和 36.5%。

4.2 评价标准

NIST 为镜头边界检测给出了一种标准的评估方法^[6],主要以镜头变换的查全率(recall)和查准率(precision)来进行评估。为了进一步评估渐变的检测效果,还增加了帧级别上的查全率(frame recall)和查准率(frame precision)。具体计算方法如下:

$$\text{查全率} = \frac{\text{正确检测到的变换数目}}{\text{参考的变换总数}}$$

$$\text{查准率} = \frac{\text{正确检测到的变换数目}}{\text{所报告的变换总数}}$$

$$\text{综合} = \frac{\text{查全率} + \text{查准率}}{2}$$

$$\text{帧查全率} = \frac{\text{正确检测到的属于变换过程的帧数}}{\text{参考的属于变换过程的帧数}}$$

$$\text{帧查准率} = \frac{\text{正确检测到的属于变换过程的帧数}}{\text{所报告的属于变换过程的帧数}}$$

$$\text{帧综合} = \frac{\text{帧查全率} + \text{帧查准率}}{2}$$

4.3 检测结果

根据 NIST 公布的各个参赛队检测结果,thuai(清华大学智能控制与系统国家重点实验室)的渐变检测结果排在第一位(见表 1,每队提交 10 组结果)。

表 1 NIST 公布的渐变检测前 20 组结果

参赛队	查全率	查准率	综合	帧查全率	帧查准率	帧综合
thuai17	0.846	0.775	0.808945	0.82	0.848	0.833765
thuai16	0.827	0.79	0.808077	0.824	0.848	0.835828
thuai15	0.792	0.82	0.805757	0.824	0.86	0.841615
thuai14	0.803	0.807	0.804995	0.829	0.848	0.838392
thuai19	0.821	0.789	0.804682	0.817	0.859	0.837474
thuai08	0.801	0.801	0.801	0.838	0.848	0.84297
thuai10	0.809	0.79	0.799387	0.839	0.835	0.836995
thuai07	0.808	0.791	0.79941	0.839	0.836	0.837497
thuai02	0.808	0.791	0.79941	0.839	0.836	0.837497
thuai05	0.811	0.787	0.79882	0.839	0.836	0.837497
it07	0.772	0.819	0.794806	0.743	0.83	0.784094
it09	0.746	0.844	0.79198	0.762	0.83	0.794548
it05	0.798	0.782	0.789919	0.73	0.831	0.777233
FS10_07	0.784	0.795	0.789462	0.757	0.735	0.745838
FS10_08	0.738	0.839	0.785266	0.713	0.778	0.744083
it03	0.828	0.738	0.780414	0.715	0.832	0.769076
FS10_06	0.799	0.76	0.779012	0.781	0.694	0.734934
FS05_07	0.709	0.847	0.77188	0.701	0.783	0.739735
it06	0.753	0.802	0.776728	0.794	0.71	0.749654
FS10_05	0.808	0.743	0.774138	0.802	0.658	0.722899

结论 本文提出了一种鲁棒的基于有限自动机的渐变镜头检测方法。我们采用 RGB 颜色直方图计算帧间差和 7 帧间差,利用有限自动机的记忆性能增加了检测的容忍程度,结合基于运动特征的自适应阈值提高了检测的鲁棒性,较好地解决了边界帧合成的问题。在 NIST 提供的 Trecvid2004 测试数据集上,本方法取得了较好的效果。

参考文献

- 1 Hanjalic A. Shot Boundary Detection: Unraveled and resolved? IEEE Transactions on Circuits and Systems for Video Technolo-

gy, 2002, 12:90~105

- 2 Zhang H, Kankanhalli A, Smoliar S W. Automatic partitioning of full-motion video. Multimedia System, 1993, 1: 10~28
- 3 Ma Yu-Fei, Sheng Jia, Chen Yuan, et al. MSR-Asia at TREC-10 Video Track: Shot Boundary Detection Task. In: Proceedings of TRECVID2003, 2003
- 4 Gilvarry J. Extraction of Motion Vectors from an MPEG Stream: [Technical Report]. School of Electronic Engineering of Dublin City University, 1999
- 5 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,2000,137(11)
- 6 Guidelines for the TRECVID 2004 Evaluation. <http://www.nlpir.nist.gov/projects/tv2004/tv2004.html>
- 7 Alan F S, Over P, Taban R. The TREC-2001 Video Track Report. TREC2001