一种自动选取阈值的视频镜头边界检测算法

成 勇, 须 德
(北京交通大学计算机与信息技术学院, 北京 100044)

摘 要: 镜头边界检测是实现基于内容的视频检索的一个重要步骤. 文中介绍了现有的镜头边界检测的基本方法, 并针对其不足提出了一种自动选取阈值的、综合考虑颜色和空间特征的镜头边界检测算法. 该方法能较好地检测出镜头突变和物体运动以及光线变化等情况, 对渐变镜头也能达到检测的目的. 实验结果表明算法能够有效地检测出视频镜头边界.

关键词: 视频; 镜头边界检测; 自动阈值
中图分类号: TP391 文献标识码: A 文章编号: 0372-2112 (2004) 03-0508-04

A Method for Shot Boundary Detection Using Adaptive Threshold

CHENG Yong, XU De
(Department of Computer Science & Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: An important step in content-based video retrieval is the temporal segmentation of video. We give an overview of some existing shot boundary detection algorithms. After that, we propose an algorithm for shot boundary detection that integrates the spatial and color features of frames. Our method is not sensitive to brightness change and quick motion. Therefore, it can improve the precision of detecting shot boundaries. Finally we give the experimental results and draw the conclusion.

Key words: video; shot boundary detection; automatic threshold

1 引言

随着多媒体存储压缩技术和计算机网络的发展, 数字视频越来越广泛地存在于人们的生活空间中, 视频点播(VOD)、数字图书馆等视频服务也开始走进人们的生活. 对这些数字视频必须进行有效的管理和组织才能很好地对其进行分析和利用.

由于视频信号是由连续的图像组成的, 不便于直接管理和检索, 因此需要将其分割成相对独立的视频片段. 由视频信号组成可知, 视频一般由多段镜头拼接而成, 镜头内部各帧图像是连续变化的. 在目前研究的基于内容的视频检索系统中, 一般都是先将视频分割成独立的镜头, 然后对每个镜头选取代表帧来表示该镜头. 视频的特征提取、运动分析和检索等过程也都可以在代表帧上完成[1]. 因此, 镜头边界检测是基于内容的视频检索中十分重要的问题. 本文综合考虑了图像的颜色特征和空间特征, 提出了一种自适应阈值的非压缩域镜头边界检测算法, 根据差值的分布自动计算阈值, 因此具有较高的查全率和精确度.

2 镜头边界检测方法综述

在镜头边界检测方面, 目前已经有了许多研究[2~7]. 在镜头转换时, 视频数据会发生一系列的变化, 镜头边界检测方法就是通过比较前后两帧间的差异来寻找这些变化的规律. 根据所使用的视频特征的不同以及应用的视频对象的不同, 镜头边界检测算法能分成许多类[2], 其中有一些算法应用于压缩视频上, 但是大部分算法都是处理非压缩视频的. 基于检测两帧差异的度量标准, 算法主要能概括成三类: 基于象素比较、基于块比较和基于直方图比较的方法[3,4].

2.1 基于象素比较的方法

基于象素比较(Pixel Comparison)的方法, 又称为模板匹配(Template Matching), 通过比较前后两帧图像对应象素之间灰度或颜色的变化来检测镜头分割:

对于灰度图像:

$$D(i,i+1)=\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}|P_i(x,y)-P_{i+1}(x,y)|}{M\times N}$$

对于彩色图像:

$$D(i,i+1)=\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\sum_{c}|P_i(x,y,c)-P_{i+1}(x,y,c)|}{M\times N} \quad (1)$$

其中, i 和 i+1 分别代表第 i 和 i+1 帧, 它们的大小为 M×N; $P_i(x,y)$ 表示第 i 帧 (x,y) 位置的象素的灰度值; c 表示颜色分量的索引, 如 R、G、B 三基色等; $P_i(x,y,c)$ 表示第 i 帧位置 (x,y) 的象素的颜色分量值. $D(i,i+1)$ 表示第 i 帧与

收稿日期: 2002-12-26; 修回日期: 2003-05-28

第 i+ 1 帧之差, 如果该差值大于一个预定义的阈值, 则认为发生了镜头突变. 该方法的主要缺点是对噪声和镜头或物体运动十分敏感, 因为它严格地局限于象素的空间位置. 摄像头的任何移动, 都会使帧间差明显增大, 从而导致错误的镜头边界检测.

212　基于块比较的方法

基于块比较的方法将每帧图像分成 b 块, 然后比较相邻帧的对应块. 这样, 第 i 和 i+ 1 帧之间的差异就可以定义为:

$$D(i, i+1) = \sum_{k=1}^{b} c_k DP(i, i+1, k) \tag{2}$$

其中 c_k 是预定义的第 k 块的系数, $DP(i, i+1, k)$ 是第 i, i+ 1 两帧的第 k 块的比较结果. 相对应的块之间差异定义为:

$$DP(i, i+1, k) = \begin{cases} 1, & \text{如果 } K_k > T_1 \\ 0, & \text{其他} \end{cases} \tag{3}$$

其中 T_1 是块间比较的阈值; K_k 为似然率(likehood ratio):

$$K_k = \frac{\left[\frac{R_{k,i} + R_{k,i+1}}{2} + \left(\frac{L_{k,i} - L_{k,i+1}}{2}\right)^2\right]^2}{R_{k,i} \cdot R_{k,i+1}} \tag{4}$$

其中 $L_{k,i}$, $R_{k,i}$ 分别是第 i 帧内第 k 块的均值和方差. 当发生变化的块的数量达到一定程度时, 则认为发生了镜头切换.

与模板匹配方法相比较, 基于块比较的方法利用了图像的局部特征来降低镜头和物体运动的影响. 因为比较的粒度较大, 运动因素的影响较小, 基于块的方法能消除慢速小物体的运动对检测的影响, 但是它的计算复杂度要高一些. 一旦前后两个对应块内容不同, 但是灰度相同, 该方法就会产生漏检.

213　基于直方图比较的方法

直方图法是使用得最多的计算帧间差的方法, 它不考虑象素的位置信息, 而使用象素亮度和色彩的统计值, 因此与基于象素比较的方法相比, 降低了对噪声和镜头或物体运动的敏感性. 直方图比较有如下两种方法:

直方图比较　　$$D(i, i+1) = \sum_{j=1}^{n} |H_i(j) - H_{i+1}(j)|$$

V2 直方图比较　　$$D(i, i+1) = \sum_{j=1}^{n} \frac{[H_i(j) - H_{i+1}(j)]^2}{H_{i+1}(j)} \tag{5}$$

其中, $H_i(j)$, $H_{i+1}(j)$ 分别是帧 i, i+ 1 的直方图在灰度(彩色)级 j 上的值; n 是灰度(彩色)级的数量. 如果两帧之差 $D(i, i+1)$ 大于一个阈值 T, 则认为发生了镜头切换. 基于直方图的方法能取得比较好的效果, 但是当两帧图像具有相似的直方图或是镜头内光线突然发生变化时该方法会误检.

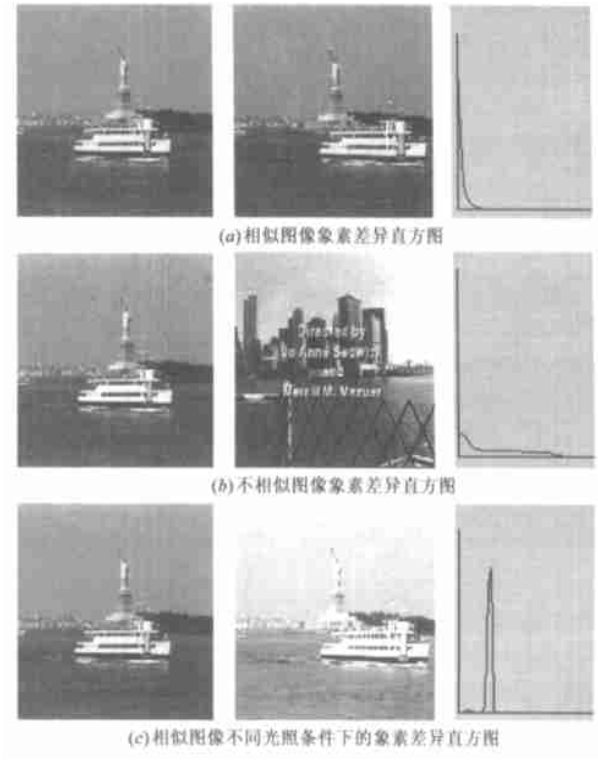
上面介绍了一般镜头边界检测所采用的基本方法, 在一些研究中对这些方法进行了部分改进[5, 6]. 如双重直方图比较方法就是通过两个阈值来检测突变和渐变[7]. 但是, 目前视频帧间差计算方法主要采用的还是几种经典方式.

3　自动选取阈值的镜头边界检测算法

从以上几种系统的分析中可以看到, 镜头分割基本上都是采用直方图相似度或帧间灰度差的方法来进行的. 但是, 单一的判别准则不能有效地检测镜头边界, 需要将多种特征综合起来考虑. 在前面介绍的系统中都需要预先确定阈值, 使得算法的灵活性和适应性都受了较大的限制, 可能在一段视频中行之有效的参数在另一段视频中却效果变得很差. 在镜头转换时, 相邻帧间将会产生整个颜色组成的显著变化或物体位置的显著变化、或者两种情况都有. 因此, 我们综合考虑了图像间的空间特征和颜色特征, 提出了一种自适应阈值的非压缩域镜头边界检测算法.

311　参数选择

在一个镜头内部, 连续的两帧图像之间的象素差值通常都是很小的, 因此对这些差值进行统计形成的直方图也都大多分布在低值区中. 而如果连续的两帧之间发生了镜头转换, 它们之间象素的差异值的分布就会比较均匀, 统计得到的直方图会均匀地分布在各个值区中. 图 1 中(a)和(b)表示了这两种情况下的象素差异直方图的分布. 从图中我们可以看出, 如果两帧图像十分相似, 则它们的差异直方图的方差将会变得很大, 而如果它们不相似, 方差就会很小. 因此, 我们可以利用差异直方图的方差来检测镜头边界. 此外, 如果两个相似的帧之间只是光照条件发生了变化(如闪光灯), 那么它们相应象素的变化值都会增大, 导致象素差异直方图整体平移, 但是总体分布都很集中, 方差依然会很大, 这类似于同等光照条件的相似帧的比较情况. 图 1(c)表示了这种情况.



(a)相似图像象素差异直方图

(b)不相似图像象素差异直方图

(c)相似图像不同光照条件下的象素差异直方图

图 1(c)　相似图象不同光照条件下的象素差异直方图

象素差异直方图方差(VDHM)的计算方法如下:

$$V(i, i+1) = \frac{\sum_{i=1}^{n} (DH_{i,i+1}(j) - \overline{DH})^2}{n} \tag{6}$$

其中, $DH_{i,i+1}(j)$ 是图像 i 和 i+1 的归一化象素差异直方图在差异级别 j 上的值, $\overline{DH}$ 是 $DH_{i,i+1}(j)$ 的平均值, n 是差异直方图的级别数量. 可以看出, V(i, i+1) 的值在 0 和 1 之间.

象素差异直方图方差反映的是图像间的空间差异, 我们利用分块直方图方法(BHDM) 来计算图像间的颜色差异. 由于镜头内存在物体运动或噪声等影响, 容易造成同属于一个镜头的图像间的部分分块的差异值很大, 因此, 我们可以通过舍弃差异最大的块以降低物体运动或噪声带来的影响:

$$D(i,i+1)=\frac{\left[\sum_{k=1}^{r}DB(i,i+1,k)\right]-\mathrm{Max}(DB(i,i+1,k))}{r-1}$$

$$DB(i,i+1,k)=\frac{\sum_{j=1}^{n}|H_{i,k}(j)-H_{i+1,k}(j)|}{n} \qquad (7)$$

其中, $H_{i,k}(j)$ 是第 i 帧的第 k 块的归一化直方图在灰度级 j 上的值, n 是灰度级的数量; r 是该帧的分块的总数. Max(DB(i, i+1, k)) 是所有对应块匹配中最大的差异值.

3.2 检测镜头突变

这样, 通过为每两帧计算其分块的直方图差(BHDM) D(i, i+1) 与象素差异直方图方差(VDHM) V(i, i+1), 每两帧之间的比较值可以用一个二维特征向量(D(i, i+1), V(i, i+1)) 表示, 每个特征值都是归一化的. 理想状态下, 发生镜头转换的两帧间的比较值将具有较大的 D(i, i+1) 值和较小的 V(i, i+1) 值. 光照变化时, 两帧比较值 D(i, i+1) 和 V(i, i+1) 都较大; 而发生镜头或物体运动时, 两帧比较值 D(i, i+1) 和 V(i, i+1) 都较小(图 3 表示了这些情况). 因此, 如果镜头发生突变, 两帧间的分块直方图差将超过一个阈值 T_h 且它们的象素差异直方图方差将小于另一个阈值 T_v. 在我们的方法中, 采用了局部阈值计算的方法来检测镜头边界. 我们设置一个大小为 W 的滑动窗口来处理连续 W 帧间的比较值. 在滑动窗口 W 中, 我们为分块直方图差和象素差异统计方差分别计算其均值 D、V. 利用这两个均值分别计算 T_h(均值 D 的 3 倍) 和 T_v(均值 V 的三分之一).

利用这两个阈值, 分别检测分块直方图差和象素差异统计方差, 可以得到滑动窗口 W 中可能的突变镜头边界 δ_h 和 δ_v. 由于光照会使差异直方图算法失效, 大物体运动会使差异直方图方差算法失效, 所以在这里我们取 $\delta_h \cap \delta_v$ 作为真正的突变镜头集, 并且得到光照集合 $\delta_h - (\delta_h \cap \delta_v)$ 和大物体运动集合 $\delta_v - (\delta_h \cap \delta_v)$.

3.3 检测镜头渐变

在确定滑动窗口内的突变镜头集、光照集合、大物体运动集合后, 我们利用这三个集合来确定渐变镜头的阈值, 方法如下:

设 Y(i, i+1) = D(i, i+1) @(1- V(i, i+1))

(1) 计算滑动窗口内的每一个 Y(i, i+1) 值, 利用直方图的方法来选取阈值 T_y, 取直方图中的第一个趋零点为该阈值. 如图 2 所示为 a. 得到所有大于该阈值 T_y 的集合 C = {Y(i, i+1)}, 集合中包含有突变检测时得到的镜头边界.

(2) 从集合 C 中筛选掉突变检测中得到的集合, 得到剩余的集合 $C_1 = C - \delta_h \cap \delta_v$.

(3) 对集合 C_1 进行判断, 如果在该集合中, 存在几个连续帧的帧间差值, 则判断该处为渐变镜头边界, 否则认为是噪声或是不确定因素造成的帧内信息变化.

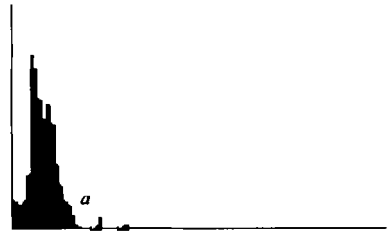


图 2 帧间差直方图

(4) 完成了上述检测后, 清空滑动窗口, 继续下 W 帧的检测. 如果剩余的帧数小于 W/2, 则计算剩余所有的帧.

本算法与传统的镜头变换检测算法比较起来, 对光照和大物体运动具有相当的鲁棒性, 也解决了镜头阈值选取不灵活等问题. 由于采用了两种参数, 象素差异直方图的计算量比较大, 因此它比传统的单一参数的算法都要慢, 但是通常镜头检测都是离线进行的, 因此速度问题相对显得次要.

4 实验

对镜头边界检测结果的评价方法一般使用查全率(Recall) 和查准率(Precision) 这两个参数, 它们的定义如下:

查全率= 正确检出数/(正确检出数+ 漏检数)

查准率= 正确检出数/(正确检出数+ 误检数)

本文也采用了这种评价标准, 查全率和查准率越高说明检测方法越好.

实验选取了五段视频, 包括运动片断、故事片和新闻片, 一共有一万余帧图像. 图像大小均为 352* 288, 24 位真彩色. 在这些视频片断中, 包含有大约 70 个镜头. 我们设置滑动窗口大小为 300, 利用我们提出的自动选取阈值的镜头边界检测算法能取得较好的效果. 图 3 显示的是部分帧间差的实验数据, 包含有光线变化(flash), 镜头突变(shot)、渐变(gradual)

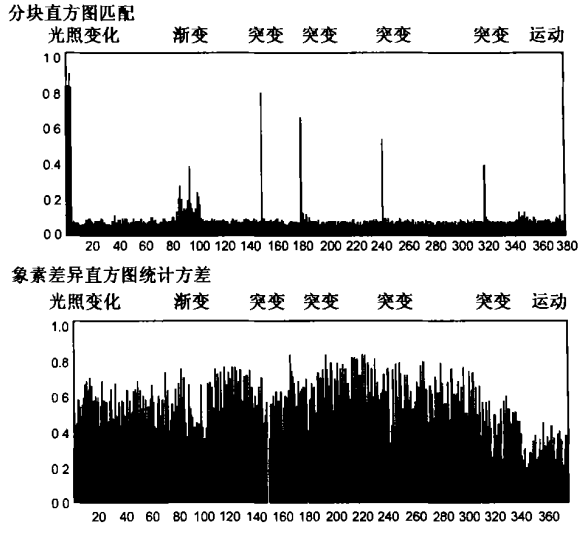


图 3 自动选取阈值的算法部分实验结果

和大物体运动(motion)等情况. 表 1 表示的是这五段视频的检测结果. 表 2 表示了用传统的几种方法的得到的检测结果, 可以看出, 自动选取阈值的方法能在综合性能上达到更好的效果.

表 1　自动选取阈值的视频镜头边界检测实验结果

视频片断	视频描述	帧数	突变	渐变	漏检	误检	光照变化	大物体运动	镜头检测结果统计	
									查准率	查全率
Train	0 突变,0 渐变	241	0	0	0	0	0	1	94.1%	95.5%
Act1 of Family Album U.S.A Ñ	10 突变, 1 渐变	1200	10	3	0	2	5	1		
Act2 of Family Album U.S.A Ñ	9 突变,2 渐变	2000	9	3	0	1	1	0		
Act3 of Family Album U.S.A Ò	43 突变, 2 渐变	5900	41	2	3	1	8	0		
Lion	0 突变,0 渐变	172	0	0	0	0	1	0		

表 2　自动选取阈值的视频镜头边界检测方法与传统镜头检测方法实验结果比较

	模版匹配	颜色直方图匹配	X^2 直方图匹配	自动选取阈值
查准率	781 48%	851 50%	89155%	941 1%
查全率	100%	951 16%	96177%	951 5%

5　结论

镜头边界检测是实现基于内容的视频检索的一个重要步骤. 本文介绍了现有的镜头边界检测的基本方法, 并针对其不足提出了一种自动选取阈值的镜头边界检测算法, 它考虑了图像的颜色特征和空间特征, 综合利用象素差值的统计方差, 直方图差异等多种方法, 根据差值的分布自动计算阈值. 该方法能较好地检测出镜头突变和物体运动以及光线变化等情况, 对渐变镜头也能达到检测的目标, 但是不能识别其类型. 镜头边界检测只是在物体特征上对视频进行分段, 进一步可以对分割出来的镜头进行语义上的分析聚类, 形成高层的视频结构, 以提供视频基于语义的检索.

参考文献:

[1] A K Elmagarmid, H Jiang, A helal, A Joshi, M Ahmed. Video database systems [M]. Boston: Kluwer Academic Publishers, 1997.

[2] Irena Koprinska, Sergio Carrato. Temporal video segmentation: A survey [J]. Signal Processing: Image Communication, 2001, 16(5): 477-500.

[3] 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 [J]. 计算机研究与发展, 2000, 37(1): 80-85.

[4] Rainer Lienhart. Reliable dissolve detection [A]. Storage and Retrieval for Media Databases 2001, Proc SPIE 4315 [C]. San Jose: Society of Photo2Optical Instrumentation Engineers, Jan 2001. 219-230.

[5] M R Naphade, R Mehrotra, et al. A high performance shot boundary detection algorithm using multiple cues [A]. Proc IEEE International Conference on Image Processing [C]. Chicago, USA: IEEE Computer Society, Oct 1998, Volume 2. 884-887.

[6] Kien A Hua, JungHwan Oh, Khanh Vu. Non2linear approach to shot boundary detection [A]. IS&T/SPIE Conference on Multimedia Computing and Networking 2001 [C]. San Jose: Society of Photo2Optical Instrumentation Engineers, Jan 2001. 1-12.

[7] H J Zhang, A Kankanhalli, S W Smoliar. Automatic partitioning of full2motion video [J]. Multimedia Systems 1993, 1(1): 10-28.

作者简介:

成　勇　男, 1977 年 9 月出生于江西省南昌市, 现为北京交通大学博士生, 主要研究方向为视频数据库技术.

须　德　男, 1944 年 2 月出生于江苏省常州市, 现为北京交通大学教授、博士生导师, 从事数据库系统及其应用、多媒体技术等方面的研究工作.