

视频镜头边缘检测算法研究与比较

牛 颢¹, 丁春利¹, 张 明¹, 唐 博²

(1. 四川省计算机研究院, 成都 610041; 2. 四川大学 计算机学院, 成都 610064)

摘 要: 视频镜头边缘检测中的直方图、像素差异等算法只针对有限的测试数据进行了范围比较狭窄的测试, 对于算法的适用领域和范围没有进行测试和说明。为了解决这一问题, 设计实现了三种视频镜头检测算法, 并构造了丰富的测试视频数据集, 针对测试结果进行了实现算法的性能分析以及提出了改进建议。

关键词: 视频检索; 镜头检测; 颜色直方图; 边缘改变比率

中图分类号: TP391.4

文献标志码: A

文章编号: 1001-3695(2013)02-0631-04

doi:10.3969/j.issn.1001-3695.2013.02.087

Research and comparison of video shot detection techniques

NIU Hao¹, DING Chun-li¹, ZHANG Ming¹, TANG Bo²

(1. Sichuan Institute of Computer Sciences, Chengdu 610041, China; 2. College of Computer Science, Sichuan University, Chengdu 610064, China)

Abstract: Color histogram and pixel differences algorithms only test in a limited material. In order to research these algorithms' used areas, this paper designed and implemented three algorithms and constructed a test video set. After test, it analyzed these algorithms' performance. And it concludes that a algorithm's performance requires a trade-off between recall and precision that must be guided by the user's demand.

Key words: video retrieval; shot boundary detection; color histogram; edge change ratio

0 引言

镜头是视频检索领域最关键的研究问题之一。镜头^[1,2]是指一个摄像机在同一个场景拍摄到的连续的帧图像,它是进行视频检索的基本元素,将视频分成时间和空间上最小的逻辑单元。镜头检测技术^[3]是基于内容的视频检索技术的核心技术之一,它为后续的视频抽象、高语义层次的视频^[4]分割、视频修复提供了基础。镜头检测技术是在基于内容的静态图像检索的基础上发展起来的。镜头检测的关键问题之一是找到合适的描述视频内容的特征,通过帧间特征值的差异来判断是否存在镜头转换。国内外已有镜头检测算法,如离散余弦变换、块匹配方法,但这些算法的使用有十分严格的局限性。而通用的镜头检测算法有基于像素差异、基于直方图差异和基于边缘改变比率差异三种。本文通过研究上述三种通用视频检测算法在不同测试数据上的镜头检测结果,分析其召回率与准确率,为不同视频内容选取最有效的镜头检测算法提供依据。

1 镜头检测算法简介及实现

一个镜头是指摄像机中未被打乱的连续的帧,因此,一个由两个人从不同视角交替的视频序列可能由很多镜头组成。场景是指一个或一组连续聚焦于一个物体或一个焦点的镜头。随着有效的在线数字视频的增多,对视频内容自动分析分类技术的需求越来越明确,很多研究主要集中于对摄像头、镜头边缘的自动检测。

1.1 基本术语介绍

镜头之间有很多不同类型的转换方式,具体有以下类型:

a) 切入。连续两帧之间很突然的镜头转换,它在一个帧的时间进行转换,如图1所示。



图1 切入镜头示例

b) 淡入淡出。明亮度缓慢地转变,通常是一个帧全部是黑色像素点开始或者结束,如图2所示。



图2 淡入淡出镜头示例

c) 溶解。第一帧图像里面的图片慢慢衰减,而第二帧图片缓慢变亮,直到第二帧的图片出现,如图3所示。

d) 消除。第二帧图像的像素按照一种固定的模式替代第

收稿日期: 2012-05-23; 修回日期: 2012-07-12

作者简介: 牛颢(1983-),男,工程师,主要研究方向为网络优化、系统设计(13438368112@gmail.com);丁春利(1983-),男,工程师,主要研究方向为视频检索及信息系统;张明(1978-),男,高级工程师,主要研究方向为网络设计;唐博(1990-),男,本科,主要研究方向为视频检索。

一个镜头的像素,如一行从右边界开始一次取代的像素点的模式,如图4所示。



图3 溶解镜头示例



图4 消除镜头示例

1.2 镜头检测算法实现

镜头检测算法早期主要集中在切入镜头转换的检查,近期更多的工作集中于检测渐变转换。主要应用于镜头边缘检测的算法有像素差异、直方图比较、边缘改变比例以及压缩差异和运动向量^[3]。而镜头检测算法最关键的问题是如何选择一个好的特征和度量来度量帧间差^[5],即帧与帧之间的视觉相似性。主要的计算帧间相似度的方法是抽取反映视觉内容的特征^[6]。

1.2.1 基于像素灰度特征

最直接反映视觉内容的元素就是每个像素的灰度或亮度值,因此度量相邻帧之间差异最简单的方法就是计算第 k 帧和第 $k+1$ 帧中所有像素的灰度或亮度的差值。于是第 k 和 $k+1$ 帧的帧间差 $Z(k, k+1)$ 可表示为

$$Z(k, k+1) = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y |I_k(x, y) - I_{k+1}(x, y)| \quad (1)$$

其中: X 、 Y 、 $I_k(x, y)$ 、 $I_{k+1}(x, y)$ 分别是图像的宽度、高度、第 k 帧中 (x, y) 像素的灰度值、第 $k+1$ 帧中 (x, y) 像素的灰度值。这种方法的一个缺点就是不能区分大区域内的小变化和小区域内的大变化,这通常会造成镜头的无检测。为区分上述两种变化,对此种算法的一种改进就是只计算灰度变化达到一定阈值的像素个数,即帧间差 $Z(k, k+1)$ 可表示为

$$Z(k, k+1) = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y D_{k, k+1}(x, y) \quad (2)$$

其中:

$$D_{k, k+1}(x, y) = \begin{cases} 1 & \text{if } |I_k(x, y) - I_{k+1}(x, y)| > T_1 \\ 0 & \text{else} \end{cases} \quad (3)$$

1.2.2 基于直方图差异

同一镜头内的相邻帧一般都有全局的视觉相同的元素,也就是同一镜头内相邻的帧具有相似的颜色空间分布;反之,不同镜头的相邻帧之间的颜色空间分布相似度很低。反映在直方图上就是:同一镜头内相邻帧之间的直方图差异较小;不同镜头中帧之间的直方图差异较大。显然理论上还是会存在这种情况:视觉内容并不相似,但是直方图却相差较小,但是在实际的视频序列中,出现这种情况的概率是非常小的。相对来说,基于直方图差异^[7]是比较简单也比较有效的方法,被广泛

采用,但是基于直方图的方法并没有考虑图像中像素的空间信息,因此对于较缓慢的物体运动并不敏感,能减少物体运动带来的无检测^[8]。

假设 $H_k(i)$ 是第 k 帧中第 i 个灰度级直方图的值, i 的范围为 $[0, N]$,其中 N 为灰度级数。通常对于彩色图像会采用三个直方图来表示,分别是红色、绿色和蓝色直方图。因此本文实现带权直方图。

第 k 帧和第 $k+1$ 帧的帧间差为

$$Z(k, k+1) = \sum_{i=0}^{N-1} |H_k(i) - H_{k+1}(i)| \quad (4)$$

在帧图像中,有些颜色分量所占比重比较大,在计算帧间差时此颜色分量应给予较大的权值。因此有了带权直方图:

$$Z(k, k+1) = \frac{r}{s} Z_{\text{red}}(k, k+1) +$$

$$\frac{g}{s} Z_{\text{green}}(k, k+1) + \frac{b}{s} Z_{\text{blue}}(k, k+1) \quad (5)$$

$$s = (r + g + b) / 3 \quad (6)$$

其中: r 、 g 、 b 分别代表图像中红色分量、绿色分量以及蓝色分量的亮度值。

1.2.3 基于边缘改变比例

时间轴上帧图像的视觉差异通常会伴随着图像内容结构上的不连续性。例如,处于不同镜头的连续两帧,上一帧内的图像的边缘必然不会出现在下一帧图像中。所以边缘也可以作为衡量视觉差异的特征。

边缘变化率(ECR)定义如下:

$$\text{ECR}_k = \max(X_k^{\text{in}} / \sigma_k, X_{k-1}^{\text{out}} / \sigma_{k-1}) \quad (7)$$

其中: X_k^{in} 、 σ_k 、 X_{k-1}^{out} 、 σ_{k-1} 分别为在第 k 帧中进入边缘的像素数、第 k 帧中所有边缘的像素数、第 $k-1$ 帧从图像中消失的边缘的像素数、第 $k-1$ 帧所有边缘的像素数。部分文献中使用Canny算子进行边缘检测,为了使边缘特征能抵抗物体运动的干扰,通常在帧图像中出现的边缘像素,如果在后续相邻帧图像中一定范围内出现,则不认为该边缘像素为进入或消失的边缘像素。

2 镜头检测算法测试集

本章详细描述所选择用于测试的算法以及测试数据的选择。因为上述三个算法的源代码并不完全可用,因此在本文中全部自主实现以保证进行一致性测试。所有的这些算法都是用C语言实现并且运行在Windows系统。由于测试程序将产生很多的数据图片,本文采用MATLAB来组织相关图片进行测试分析。为了实施对所实现算法的全面测试,本文选择混合视频作为测试数据,具体的视频数据类型描述如下:

a) 电视节目。这段材料包含168 s《亮剑》电视剧的情节,共计121个镜头转换,其中有2个切入镜头转换、119个渐变镜头转换。

b) 新闻节目。这段材料是CCTV-1的新闻联播节目,共计299 s,其中包含86个切入镜头和24个渐变镜头转换,共110个镜头转换。

c) 电影。这段材料包含《窃听风云2》中间的173 s,总共有4 325帧,其中有30个切入镜头转换、137个渐变镜头转换。

d) 音乐视频。这段材料是《我愿意》的完整视频,共计274 s、4 110帧,其中有99个切入镜头、43个渐变镜头。

e) 其他视频。是电脑屏幕使用视频,总共450 s、4 500帧,

其中有切入镜头 82 个、141 个渐变镜头。

f)短视频。总共 8.833 s、共计 265 帧,其中有切入转换 6 个、渐变转换 10 个。

上述测试数据的镜头边界定位和镜头转换类型都是由人工分析出来的,具体信息如表 1 所示。

表 1 视频类型

视频类型	总帧数	切入	渐变
电视	$168 \times 25 = 4\ 200$	2	119
电影	$173 \times 25 = 4\ 325$	30	137
新闻	$299 \times 25 = 7\ 475$	86	24
音乐视频	$274 \times 15 = 4\ 110$	17	124
其他视频	$450 \times 10 = 4\ 500$	82	141
短视频	$8.833 \times 30 = 265$	6	10

3 测试与分析

3.1 衡量标准和测试结果

本文采用召回率和准确率^[3,7,9]作为评价标准。召回率和准确率通常被用于信息检索领域。召回率是指被检索到的需要检索的项的百分比。准确率是指被准确检索到的项所占检索到的项的百分比。具体定义如下:

召回率 = 正确数 / (正确数 + 丢失数)

准确率 = 正确数 / (正确数 + 错误数)

单纯通过召回率和准确率的值很难界定算法的优劣。例如,一个自动视频索引系统使用人工操作得到很高的召回率,而通过选取视频中分钟的关键帧得到很高的准确率。所以在本文中,通过考察召回率、准确率以及阈值来比较算法的优劣。这个比较实验产生了很大的数据量,在此只作总结分析。

三种算法在不同测试数据上的测试结果如表 2~7 所示。

表 2 电视数据测试结果

算法	召回率	准确率
像素灰度差异	53/121	53/76
直方图差异	71/121	71/99
边缘改变比例	94/121	94/131

表 3 电影数据测试结果

算法	召回率	准确率
像素灰度差异	73/167	73/142
直方图差异	63/167	63/117
边缘改变比例	112/167	112/151

表 4 新闻数据测试结果

算法	召回率	准确率
像素灰度差异	94/110	94/101
直方图差异	75/110	75/96
边缘改变比例	83/110	83/130

表 5 音乐视频测试结果

算法	召回率	准确率
像素灰度差异	77/141	77/134
直方图差异	78/141	78/127
边缘改变比例	90/141	90/136

表 6 其他视频测试结果

算法	召回率	准确率
像素灰度差异	91/223	91/121
直方图差异	85/223	85/143
边缘改变比例	96/223	96/160

表 7 短视频测试结果

算法	召回率	准确率
像素灰度差异	6/16	6/11
直方图差异	8/16	8/9
边缘改变比例	9/16	9/15

3.2 召回率分析

依照 3.1 节中的测试数据可得:不同测试数据在不同算法中召回率如图 5 所示。

分析图 5,可得如下结论:

a)新闻视频在像素灰度差异算法和直方图差异算法中都取得最高的召回率,分别为 0.855 和 0.682。边缘变化比例算法也取得了很好的结果。这是因为新闻视频的内容切入镜头

比较多且新闻视频内容结构化相对明显,如图 6 所示。

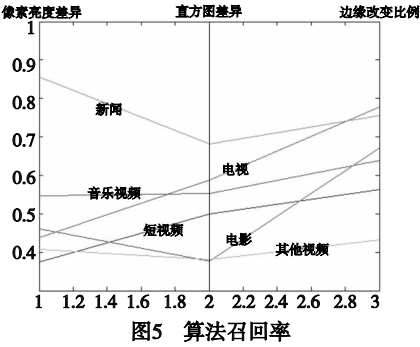


图5 算法召回率



图6 新闻镜头检测结果

b)对于电视测试数据,因为选取的测试数据是一段抗日战争的片段,视频中不同镜头的背景像素的亮度相似性很高。所以用像素灰度算法进行镜头检测召回率很低。而在调整阈值之后,对于同样的视频内容,电视节目通过直方图算法测试得到的召回率有明显的提高,如图 7 所示。



图7 电视镜头检测结果示意及相应直方图

c)对于电影视频数据,因为所选择视频的实际窗口大小为 1440 × 1280,而真实反映视频变化的窗口大小是 790 × 350。因此背景对视频镜头边缘检测的影响很大,直方图方法检测效果很差,如图 8 所示。

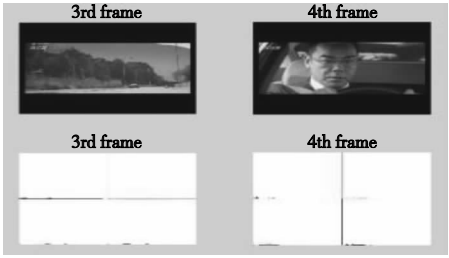


图8 电影镜头检测结果示意及相应直方图

3.3 准确率分析

依照 3.1 节中的测试数据可得:不同测试数据在不同算法中的准确率如图 9 所示。

分析图 9 可得如下结论:

a)对于短视频测试数据,因为阈值选择苛刻,所以直方图

算法得到的准确率很高。如图 10 所示,左图为阈值为 50 时检测的帧数,右图为阈值为 60 时选择的帧数,很明显,阈值越高,检测到的镜头越少,准确率也越高。

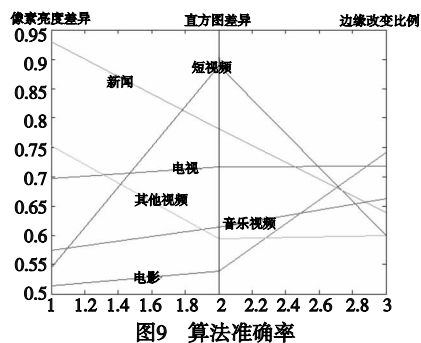


图9 算法准确率

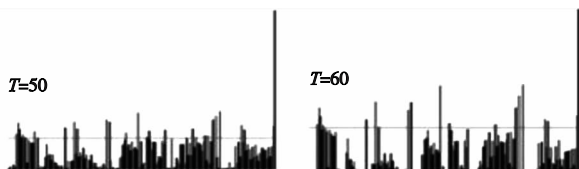


图10 阈值对镜头检测结果的影响

b) 对于音乐视频,因为渐变镜头最多,所以采用像素灰度差异和直方图差异算法时,得到的准确率并不是很高。而依据视频内容,采用边缘改变比例算法能得到较为理想的准确率。如图 11 所示。

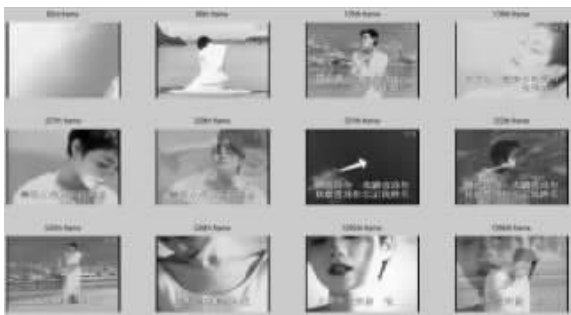


图11 渐变镜头对检测结果的影响

3.4 召回率和准确率相互关系

依照 3.1 节中的测试数据可得:不同测试数据在不同算法中准确率和召回率的相互关系如图 12 ~ 14 所示。

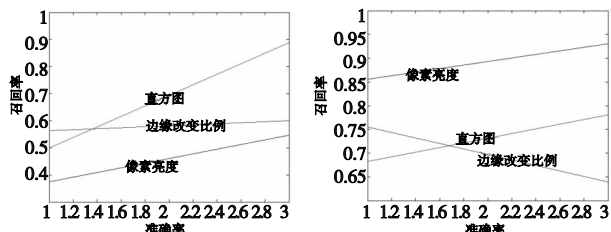


图12 短视频的召回率和准确率

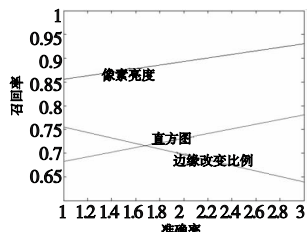


图13 电影视频的召回率和准确率

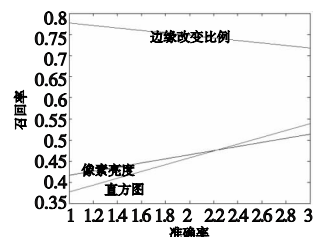


图14 电视的召回率和准确率

分析图 12 ~ 14 可得到如下结论:

a) 直方图算法取得的召回率虽然不是最高但是准确率是

这三种算法中最高的。

b) 牺牲准确率可以得到很高的召回率。

c) 对于给定的阈值,视频的内容对召回率和准确率的影响很大。

d) 切入镜头的比例和结构化的视频内容能显著提高直方图和像素亮度算法的视频镜头召回率和准确率。

e) 对于大部分算法准确率均高于召回率,而良好的算法性能取决于阈值选择和召回率及准确率的折中。

f) 越严格的阈值选择,可以得到越高的准确率,同时召回率会受到影响而降低。

4 结束语

本文设计并实现了基于像素点差异、直方图差异和边缘改变比率三种视频镜头检测算法,并构造了丰富的测试视频数据集。针对测试结果进行了这三种算法的性能及其分别适应领域的分析,并对这三种算法的检测效果提出了改进建议。

通过比较和分析这三种算法在不同测试数据集上的测试结果,以及结合阈值选取、镜头检测结果的召回率和准确率得到:这三种算法在符合其特征检测视频上都取得了很好的视频检测结果;当结合这三种算法时能产生比这三种中任何一种都要好的结果。通常,简单的算法胜过更加复杂的算法性能。这些复杂的算法对阈值设定比较敏感。在本文中,这些暗含的参数没有详细描述。

参考文献:

- [1] ULGES A, SCHULZE C, KOCH M, *et al.* Learning automatic concept detectors from online video[J]. *Computer Vision and Image Understanding*, 2010, 114(4): 429-438.
- [2] 秦剑鹏, 符茂胜, 涂铮铮, 等. 基于颜色直方图变化率的视频镜头检测[J]. *计算机应用与软件*, 2011, 28(4): 17-20.
- [3] BORECZKY J S, ROWE L A. Comparison of video shot boundary detection techniques[J]. *Journal of Electronic Imaging*, 1996, 5(2): 170-179.
- [4] CHANG S F, JIANG W, YANAGAWA A, *et al.* Columbia University TRECVID2007 high-level feature extraction [C]//Proc of TRECVID Workshop. 2007.
- [5] ZABIH R, MILLER J, MAI K. A feature-based algorithm for detecting and classifying production effects[J]. *Multimedia Systems*, 1999, 7(2): 119-128.
- [6] BORTH D, ULGES A, SCHULZE C, *et al.* Keyframe extraction for video tagging and summarization [C]//Proc of Informatiktage. 2008: 45-48.
- [7] ZHANG H J, KANKANHALLI A, SMOLIAR S W. Automatic partitioning of full-motion video[J]. *Multimedia Systems*, 1993, 1(1): 10-28.
- [8] BOLLE R M, YEO B L, YARMG M M. Content-based digital video retrieval [C]//Proc of International Broadcasting Convention. 1997: 160-165.
- [9] LEFEVER S, INCENT V. Efficient and robust shot change detection [J]. *Journal of Real-Time Image Processing*, 2007, 2(1): 23-34.