

# 一种基于非监督聚类的视频镜头边界检测算法

陈 静<sup>1</sup>, 刘 洋<sup>2</sup>

(1. 广东工业大学 物理与光电工程学院, 广东 广州 510006

2 广东工业大学 信息工程学院, 广东 广州 510006)

摘要: 在分析归纳现有的镜头边界检测技术的基础上, 提出了一种分层结构的基于非监督聚类的视频镜头边界检测算法. 实验结果表明所提出的新算法不仅对切变的检测取得了较高的查全率和准确率, 而且对目前技术仍不成熟的渐变检测也取得了较满意的结果. 另外针对不同类型的视频, 该算法的检测质量具有较强的鲁棒性.

关键词: 镜头边界检测; 矩保持聚类; 颜色聚合向量直方图; 闪光检测

中图分类号: TP391

文献标识码: A

文章编号: 1007-7162(2006)04-0076-06

随着多媒体技术和计算机网络的发展, 视频已经成为许多应用中不可缺少的部分, 视频点播、数字图书馆等视频服务已开始走进人们的生活. 人们对视频进行合理的组织和管理, 才能有效地访问和利用这些视频信息.

因为原始视频是由一系列的图像帧组成, 不便于管理和检索, 所以需要将其分割为独立的视频片段. 目前研究的基于内容的视频检索系统, 一般都是先将视频分割为独立的镜头, 然后对每个镜头选取关键帧, 后续步骤(视频特征提取、运动分析等)都可以在关键帧上完成. 所以如何准确、快速地检测出镜头边界, 进而分割出镜头是十分重要的研究课题.

本文在分析归纳现有的镜头边界检测方法的基础上, 提出了一种分层结构的基于非监督聚类的视频镜头边界检测算法. 实验结果表明所提出的新算法不仅具有较高的查全率和准确率, 且对不同的视频有稳定的检测效果.

## 1 现有镜头边界检测方法讨论

两个镜头进行切换形成镜头边界. 对于观察者来说, 是视频镜头的内容发生了某种意义上的变化, 即边界是由于视频内容的不连续造成的. 检测边界的任务即在于此, 好的算法检测出的边界应与这种视觉内容上的不连续一致. 可靠的边界检测方法一直是过去十几年来研究的重心<sup>[1]</sup>, 其中特征的选取, 以及如何利用特征量进行检测则是关键.

### 1.1 用于视频镜头边界检测的特征

基于像素对的特征是对视频内容不连续性的最简单度量, 及计算相邻帧各个对应像素之间的绝对差值:  $D = |P_i(x, y) - P_{i+1}(x, y)|$ , 其中  $P_i(x, y)$  和  $P_{i+1}(x, y)$  分别为两帧对应点的像素值. 这种做法受局部噪声影响显著, 鲁棒性较差. Zhang 等人<sup>[2]</sup>提出的改进方法可以去掉图像中的一些噪声, 并对小的物体运动和镜头运动起到补偿作用, 但对较大的镜头或物体的运动, 检测效果仍然不是很好.

收稿日期: 2005-12-23

作者简介: 陈静 (1980-), 女, 助教, 主要研究方向为多媒体信息处理.

直方图不考虑像素的位置信息, 而使用亮度和色彩的统计值, 因而基于直方图的特征抗噪能力比直接采用像素对特征强。但是, 由于直方图是全局的特征, 因此有可能两幅图像虽然有完全不同的内容, 但是却有相似甚至完全相同的直方图。

基于运动补偿的特征充分考虑了同一个镜头内对象及摄像机的运动情况和特征, 通过运动补偿等方法来减小对象和摄像机运动造成的镜头内不连续值的变化。有效的基于运动的特征需要计算光流和运动向量的分布和数量, 计算复杂度高。

另外一个有效的特征是图像的边缘。因为在发生镜头转换时, 新出现的边缘应远离旧边缘的位置, 消失的旧边缘也应远离新边缘的位置<sup>[3]</sup>。同样, 较大的计算量限制了该特征的应用。

## 1.2 视频镜头边界检测算法

有了描述镜头边界的特征以后, 最常用的镜头转换检测方法便是基于阈值的方法。对于该法, 阈值的选取是关键。早期的很多算法采用全局阈值。由于视频序列的性质不是平稳的, 采用全局阈值不能适应这种不稳定性, 因此检测性能不够好。二次差分阈值设定法和滑动窗口法<sup>[4]</sup>都从不同侧面有效地改善了因摄像机和目标的剧烈运动造成误检和漏检的情况。对于渐变检测, 阈值设定最经典的算法是双阈值法<sup>[5]</sup>。这种方法的问题是镜头内的全局运动也会引起累加的帧差, 将影响到算法的检测性能。

用聚类的方法来进行镜头分割, 则可避免阈值的设定, 克服设定的阈值对输入视频内容变化敏感等缺点。Gtinsel等人<sup>[6]</sup>使用 k 均值算法, 根据相邻帧的颜色直方图来自动将帧差值分为场景变化和无场景变化两类。其最大优点在于去掉了阈值的设定并且利用多个特征来改善检测的可靠性。这种算法对各段视频序列有自适应的能力, 但对噪声的干扰比较敏感。

文献[7]中提出的基于数学模型的检测算法。只要模型建立得准确, 该法对渐变的检测往往能取得较好的效果。但是需要对每一种渐变过程建立数学模型, 而且建模的过程比较复杂。

## 2 基于非监督聚类的视频镜头边界检测算法

基于阈值的算法是以往采用最广泛的算法, 但是阈值的选取问题一直是影响检测效果的关键, 单一的阈值选取机制应用在不同性质的视频序列上, 不能得到稳定的检测效果, 而且这些阈值一般都来自实验和经验, 因而受人为主观因素的影响较大。另外, 基于模型的渐变检测算法, 需要对每种渐变建立模型, 有一定计算复杂度。基于聚类的算法, 因不需设定阈值, 且可以采用多特征, 所以对不同类型的视频序列有比较一致的检测效果。

1.2节中提到的现有的一些基于聚类的方法只区分出镜头转换帧和一般镜头帧两类, 并没有进一步区分出切变帧和渐变帧, 而且大部分算法只针对切变帧的检测。另外, 这些算法并没有考虑闪亮所造成的误检。根据现有算法的种种不足, 本文提出一种分层结构的基于非监督聚类的镜头边界检测算法。

### 2.1 特征选取

由于颜色直方图丢失了空间信息, 两幅空间信息完全不同但是直方图完全一样的图像将被判定为同样图像, 造成误识别, 因此本文提取颜色聚合向量 (Color Coherence Vector, CCV) 直方图<sup>[8]</sup>作为镜头边界初检特征。

CCV 直方图是基于空间连贯性对图像像素点进行区域分割的一种方法, 在一个大小可变的连续区域中, 如果一个像素点灰度值与这个区域相似, 则此像素点被称为空间连贯像素点。这样, CCV 直方图通过计算图像中色彩相似区域数目和每个相似像素点的数目, 来弥补传统直方

图丢失相似像素点空间位置造成的缺陷. 它同样具有一般直方图对物体运动不敏感的优点. 对于灰度图像, 记灰度值  $j$  的连贯像素点数目为  $\alpha_j$ , 不连贯像素的数目为  $\beta_j$ , 则一幅图像中 CCV 直方图表示为  $(\alpha_1 + \beta_1, \dots, \alpha_n + \beta_n)$ . 对于任意两幅图像  $I$  和  $I'$ , 可以使用 CCV 直方图来比较其相似性  $\Delta_{CCV} = \sum_{j=1}^n |(\alpha_j - \alpha'_j)| + |\beta_j - \beta'_j|$ . 而传统直方图是通过  $\Delta_{CH} = \sum_{j=1}^n |(\alpha_j + \beta_j) - (\alpha'_j + \beta'_j)|$  来比较  $I$  和  $I'$  之间的相似性.

基于像素对的特征对运动和噪声非常敏感, 直接用于镜头边界检测会造成较多误检. 而这里, 我们要区分摄像机镜头运动和渐变就正好利用该特征对运动敏感这一特点.

本文使用模板匹配法计算相隔  $s$  帧的两帧的帧间差  $d(I_i, I_{i+s})$ :

$$d(I_i, I_{i+s}) = \frac{\sum_{x=0}^{x < W} \sum_{y=0}^{y < H} |I_i(x, y) - I_{i+s}(x, y)|}{W \times H},$$

考虑到渐变或相机运动一般都会持续十几帧到几十帧, 本文取  $s = 5$  式中,  $I_i(x, y)$  为帧  $I_i$  在  $(x, y)$  位置的像素值,  $W, H$  分别为帧的宽度和高度. 由于模板匹配方法对运动很敏感, 所以当摄像机镜头运动时, 帧间差会很大; 而渐变是由一幅图像慢慢过渡到另一幅图像, 过渡过程中的图像都具有较强的相似性, 所以帧间差比较小.

## 2.2 聚类算法

本文使用的矩保持聚类算法是 Tsa 在文献 [9] 中提出的分析方法. 该法从解决矩保持方程得到聚类中心, 然后把各个样本数据按距离最近的原则分配到各个类中. 当聚类数目是 2、3 或 4 时, 该算法的聚类中心寻找时间呈线形分布.

近年来, 矩保持法在模式识别与图像处理方面的应用日益成熟, 还有越来越多的学者将它引入到基于内容的视频检索领域当中. 在此, 作者在前人的基础之上将它应用到镜头边界检测之中, 取得了较为满意的效果.

## 2.3 算法实现

具体的算法流程描述如下

- 1) 提取帧间颜色聚合向量 (CCV) 直方图差作为输入特征数据, 通过矩保持聚类, 得到三个视频帧类: 非转换类 (NC)、准切变类 (SC) 和准渐变类 (ST).
- 2) 对准切变类预检测, 将一部分连续帧归入准渐变类中, 剩下的标记为准切变帧.
- 3) 对准渐变类, 作相关预处理后, 用模板匹配法提取其相隔  $s$  帧的两帧帧间差作为输入特征数据, 再次聚类, 得到渐变类 (T) 和镜头运动类 (M). 本文取  $s = 5$
- 4) 对 2) 中得到的准切变帧, 引入闪光检测机制, 检测出闪光帧, 得到切变帧.
- 5) 以 3) 得到渐变类和 4) 得到的切变帧作为镜头边界, 作初步镜头分割, 提取镜头关键帧.
- 6) 采取融合策略去除误检, 确定最终的切变帧和渐变帧, 实现对视频镜头的分割.

## 3 实验结果及分析

本文采用多媒体信息检索中的通用指标——查全率和准确率来衡量镜头检测的效果, 计算公式如下

$$\text{Precision} = \frac{N_{\text{correct}}}{N_{\text{correct}} + N_{\text{false}}}, \quad \text{Recall} = \frac{N_{\text{correct}}}{N_{\text{correct}} + N_{\text{miss}}}$$

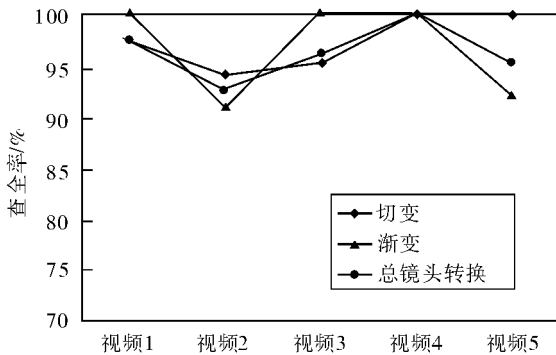
其中, Precision 是准确率, Recall 是查全率,  $N_{correct}$  是正确检测到的镜头个数,  $N_{false}$  是错误检测出的镜头个数,  $N_{miss}$  是漏掉未检出的镜头个数。

为了验证本文提出的算法的检测性能对不同类型的序列是稳定的, 作者采集了包括广告、动画、MTV、电影片段和体育等五大类各种各样的视频序列。视频 1 是两部电影的片段, 视频 2 是两段 MTV, 视频 3 是篮球和足球的片段, 视频 4 和视频 5 分别是三段广告和一个动画片段。此外, 还将本文的方法与同样使用矩保持聚类算法的文献 [10] 中的方法做了对比实验。文献 [10] 中使用一般颜色直方图作为特征, 算法不区分镜头的切变和渐变。

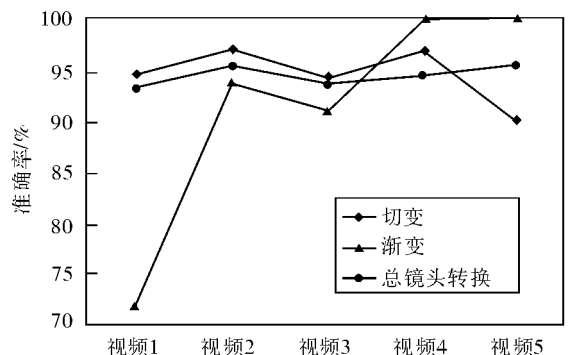
实验结果如表 1 所示。实验查全率和准确率的统计结果如图 1 所示。

表 1 实验结果

视频段	视频描述	本文算法						文献 [10] 的算法		
		切变			渐变			检出	误检	漏检
		检出	误检	漏检	检出	误检	漏检			
视频 1	103 切变 5 渐变	106	6	3	7	2	0	111	13	10
视频 2	66 切变 64 渐变	64	2	4	62	4	6	132	15	13
视频 3	84 切变 20 渐变	85	5	4	22	2	0	109	16	11
视频 4	31 切变 3 渐变	32	1	0	3	0	0	37	4	1
视频 5	18 切变 25 渐变	20	2	0	23	0	2	46	5	2



(a) 各类镜头的查全率



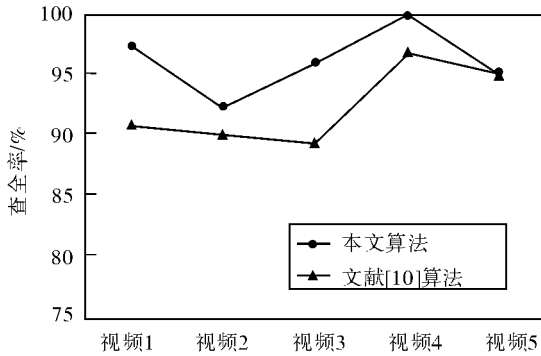
(b) 各类镜头的准确率

图 1 本文算法的检测效果

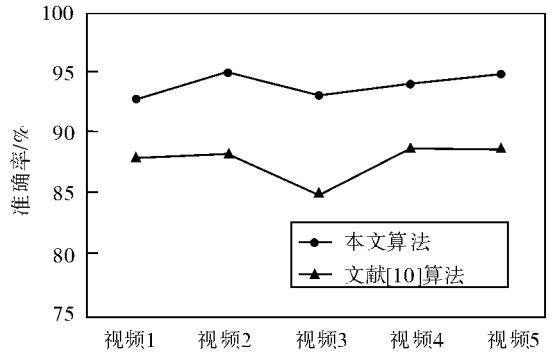
从表 1 和图 1 看到, 本文算法对于不同视频有较一致的检测效果, 并且取得了较高的平均查全率和准确率。图 1 中视频 1 的渐变检测准确率严重偏离平均水平, 是由于该视频段所包含的渐变镜头数少, 因运动造成的误检所占比重就相对较大。

实验中, 用本文算法进行镜头边界检测时, 对切变的误检大多是由于对象的突然出现, 大物体在镜头前移动等因素引起的; 对切变的漏检是因为切变前后两帧的画面相似, 或只有物体的空间位置变化, 另外一些暗镜头间的切变也被漏检了; 对渐变的误检因素比较多, 包括背景变化小的摄像机镜头运动, 对象的快速非平移运动, 光照的逐渐变化等; 对渐变的漏检则主要是因为渐变前后镜头的关键帧相似, 被合并成同一个镜头了。

图 2 显示, 用本文算法进行镜头边界检测所得到的平均查全率和平均准确率都较文献 [10] 算法有了显著提高。



(a) 查全率的对比



(b) 准确率的对比

图 2 本文算法与文献 [10] 算法检测效果比较

## 4 结束语

本文在分析现有的镜头检测技术及其不足的基础上, 针对基于非监督聚类的方法, 提出了分层结构的、多种特征联合检测的新思路以及具体的实现方案. 用 VC6.0 开发实现了桌面级的视频镜头边界检测系统, 从而验证了提出的新算法不仅具有较高的查全率和准确率, 且对不同类型的视频的检测质量具有较强的鲁棒性. 但是检测算法没有完全消除大物体运动带来的干扰, 对一些暗镜头间的切换也无法有效地检测, 这也是我们下一步要研究解决的问题.

## 参考文献:

- [1] Koprinska J, Carrato S. Temporal video segmentation: a survey[J]. Signal Processing Image Communication, 2001, 16(5): 477-500.
- [2] Zhang H, Wu JH, Zhong D, et al. An integrated system for content-based video retrieval and browsing[J]. Pattern Recognition, 1997, 30(4): 643-657.
- [3] Zabih R, Miller JM, et al. A feature-based algorithm for detecting and classifying scene breaks[C]. ACM Multimedia 95. San Francisco CA, 1995. 189-200.
- [4] Yeo BL, Liu B. Rapid scene analysis on compressed video[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1995, 5(6): 533-544.
- [5] Zhang H, Kankanhalli A, Smoliar S W. Automatic partitioning of full-motion video[J]. Multimedia Systems, 1993, 1(1): 10-28.
- [6] Gensel B, Fernan AM, Tekap AM. Temporal video segmentation using unsupervised clustering and semantic object tracking[J]. Journal of Electronic Imaging, 1998, 7(3): 592-604.
- [7] Hanapur A, Jain R, Weymouth T. Production model based digital video segmentation[J]. Multimedia Tools and Applications, 1995, 1(1): 9-46.
- [8] Pass G, Zabih R. Histogram refinement for content-based image retrieval[C]. Proceedings of 3rd IEEE Workshop on Applications of Computer Vision, 1996. 96-102.
- [9] Tsai W S. Moment-preserving thresholding: A new approach. Comp Vision Graphics Image Process[J], 1985, 29(3): 377-393.
- [10] Lo C C, Wang S J. A histogram-based moment-preserving clustering algorithm for video segmentation[J]. Pattern Recognition Letters, 2003, 24(14): 2209-2218.

## A Method of Video Shot Boundary Detection Using Unsupervised Clustering

CHEN Jing<sup>1</sup>, LU Yang<sup>2</sup>

(1. Faculty of Applied Physics, Guangdong University of Technology, Guangzhou 510006, China)

2. Faculty of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract** According to the analysis of the existing shot boundary detection algorithms, a new algorithm based on moment-preserving clustering is proposed. Simulation results indicate that the proposed algorithm can not only improve the recall and the precision of detecting shot boundaries, but also provide satisfactory results of gradual shot boundaries detection. In addition, the algorithm is robust to various types of video sequence.

**Key words** shot boundary detection; moment-preserving clustering; CCV histogram; flashlight detection

---

(上接第 63 页)

## Typical Application of Fuzzy Theory Used in Equipment Fault Diagnosis Based on ANN

LONG Xiang, QIAN Zhibo

(College of Marine, Northwestern Polytechnical University, Xi'an 710072, China)

**Abstract** This paper gives the way of using fuzzy theory to establish ANN (artificial neural network) for single-fault and multi-faults diagnosis based on one set of symptoms and multifault diagnosis based on more sets of symptoms, respectively. In addition, each ANN established here is a typical kind of fuzzy neural networks for fault diagnosis. The paper sets forth how to establish an available fault diagnosis system based on ANN while in practice, therefore, it may have some extent reference values and may be instructive in fault diagnosis field to construct fault diagnosis system based on ANN.

**Key words** artificial neural network (ANN); fault diagnosis; fuzzy theory; fuzzy neural network