第 36 卷 第 9 期 计 算 机 工 程 2010 年 5 月
Vol.36 No.9 Computer Engineering May 2010
·多媒体技术及应用· 文章编号：1000—3428(2010)09—0230—02 文献标识码：A 中图分类号：TP391

基于聚类的镜头边界检测算法

许文竹，徐立鸿
(同济大学电信学院，上海 200092)

摘 要：镜头边界检测是基于内容视频检索的重要组成部分。为从不同类型的视频中有效地检测出视频镜头边界，提出一种视频镜头边界检测算法。通过视频帧图像的颜色特征，得到视频的相似性矩阵，根据突变镜头和渐变镜头在 Affinity Propagation 聚类结果中的不同特点，运用双阈值法检测镜头边界。实验结果表明，该算法从视频的本身信息分布出发，能自动快速地检测出镜头边界。

关键词：视频检索；镜头边界；视频镜头；聚类

Shot Boundary Detection Algorithm Based on Clustering

XU Wen-zhu, XU Li-hong
(School of Electronics and Information, Tongji University, Shanghai 200092)

【Abstract】 Shot boundary detection is important component of content-based video indexing. For detecting shot boundary efficiently from different video, this paper presents an efficient method for shot boundary detection. It gets the shot comparability matrix by histogram differences, according to the different characteristics of abrupt change and gradual change shot in Affinity Propagation clustering results, detect shot change by twin-comparison. Experiment results show that the algorithm can detect shot boundary automatically and fast from video information distribution.

【Key words】 video indexing; shot boundary; video shot; clustering

1 概述

随着多媒体技术和网络技术的高速发展，产生大量的视频文档，视频数据信息越来越多。传统的基于关键词描述的视频检索因其描述能力有限、工作劳动力大、主观性强等很多客观因素，已经不能适应海量视频检索的需要。因此，如何能对视频数据库进行快速地检索和访问，越来越受到人们的重视。镜头边界的检测是实现视频自动快速检索的一个重要工作，已经成为基于内容的视频检索的重要组成部分。

2 镜头边界检测的一般算法

传统的镜头边界检测算法[1]大致可分为以下 3 类：(1)模板匹配法。该方法通过 2 帧图像之间的灰度或颜色来检测镜头分割，这种方法对噪声、镜头运动非常敏感，从而导致误检测。(2)基于图像边缘的方法[2]。该算法通过帧图像出现的边缘变化检测镜头切变，其算法计算量大，且容易出现误检测。(3)基于直方图的方法[3] 。通过帧图像直方图的差异判断 2 帧图像是否存在着镜头的切变，该算法不考虑图像像素点的位置信息，因此，不能反映图像的整体内容。尽管在镜头边界检测技术上已有许多研究成果，然而寻求一个能自动快速地检测出镜头边界且稳健的算法仍然是视频检索领域的一个难题。

本文提出一种新的镜头边界检测算法，能够根据视频本身内容的分布，自动快速地检测出视频镜头边界，且算法稳健。该算法首先通过视频帧的颜色特征，计算出视频帧图像的相似性矩阵，然后根据突变镜头和渐变镜头在 Affinity Propagation 聚类结果中的不同特点，通过双阈值法自动快速地检测出视频镜头。该算法不需要用户提供聚类个数或者聚类中心等参数，克服了运用 K 均值算法时，用户难以确定聚类个数和初始聚类中心及初始聚类中心的随机性而带来结果的不稳定性，能根据视频内容的复杂度和视频内容信息分布，从视频数据本身的结构出发，快速地检测出视频镜头边界。

3 特征提取

颜色是图像内容最基本的元素，也是图像最直观的特征。选择一个符合人眼视觉特性的颜色空间对于视频检索至关重要[4]。*HSV* 颜色空间反映了人眼视觉观察彩色的方式，同时也有利于图像处理。*HSV* 颜色空间有 2 个重要的特点：(1)V 分量(即亮度)与彩色信息无关；(2)H 分量(即色调)和 S 分量(即饱和度)与人感受彩色的方式紧密相连。这些特点使得 *HSV* 颜色空间非常适合于借助人眼视觉系统感知彩色特性的图像处理方法。

在确定采用 *HSV* 空间模型后，将色调 H 空间、饱和度 S 空间、亮度 V 空间 3 个分量按照人的颜色感知进行非等间隔的量化，其中，色调 H 空间被分成 8 份，饱和度 S 空间被分成 3 份，亮度 V 空间被分成 3 份，因此，整个 *HSV* 空间被分为 72 个子空间(8×3×3)，把这 3 个颜色分量(H, S, V)通过式(1)合成一维特征矢量：

$$l=9H+3S+V \tag{1}$$

通过统计帧图像在各个子空间上的像素个数，可得到该帧的 *HSV* 颜色直方图矢量，该矢量有 72 个分量，即 0~71。在得到每个帧图像的 *HSV* 直方图后，需要选择合适的距离度量表示帧图像之间直方图的差异[5]，可以简单地用绝对值距离表示，这样既计算简单，又满足计算速度要求。

$$D(\boldsymbol{H}_i,\boldsymbol{H}_j)=\sum_{k=0}^{71}|\boldsymbol{H}_i(k)-\boldsymbol{H}_j(k)| \tag{2}$$

其中，$\boldsymbol{H}_i$ 是第 i 帧图像的直方图矢量；$\boldsymbol{H}_j$ 是第 j 帧图像的直方图矢量；$D(\boldsymbol{H}_i,\boldsymbol{H}_j)$ 是第 i 帧图像和第 j 帧图像之间的直方

作者简介：许文竹(1979－)，女，博士研究生，主研方向：基于内容的视频检索；徐立鸿，教授、博士生导师

收稿日期：2009-12-10 **E-mail：**xuwenzhu@163.com

图距离。

这里定义视频帧之间的相似性矩阵 $\boldsymbol{S}$ 为视频帧直方图之间距离的负数，即 2 帧图像直方图距离越大，其相似性越小，距离越小，其相似性越大。

$$\boldsymbol{S}(\boldsymbol{H}_i,\boldsymbol{H}_j)=-D(\boldsymbol{H}_i,\boldsymbol{H}_j) \tag{3}$$

其中，$\boldsymbol{S}(\boldsymbol{H}_i,\boldsymbol{H}_j)$ 表示第 i 帧图像和第 j 帧图像之间的相似性。

4 Affinity Propagation聚类镜头边界的检测

Affinity Propagation 聚类是由文献[6]提出的一种新的聚类算法，其优势在于算法快速、有效。它与 K 均值算法都是属于 K 中心聚类方法。然而，K 均值算法需要用户指定聚类个数以及初始聚类中心，而且对初始聚类中心的选择敏感[7]，不同的初始聚类中心会导致不同的聚类结果。Affinity Propagation 聚类算法则克服了这些缺点，其迭代过程不断搜索合适的聚类中心，同时也使得聚类的适应度函数最大化，能给出比较准确的聚类结果，而且运算速度快。

Affinity Propagation 聚类算法以所有样本相似度组成的样本相似性矩阵 $\boldsymbol{S}(N\times N)$为基础，其中，$N$ 为样本数。首先将数据集的全部样本点都视为候选的聚类中心；然后在循环迭代过程中，各个样本点相互竞争最终的聚类中心。用 $R(i,k)$ 表示样本点 k 对样本点 i 的吸引度，即样本点 k 适合作为样本点 i 的类代表的度，$A(i,k)$ 表示样本点 i 对样本点 k 的归属度，即表示样本点 i 选择样本点 k 作为其类代表的适合程度。$R(i,k)$ 与 $A(i,k)$ 越大，说明样本点 k 作为最终聚类中心的可能性越大。这样，通过每次迭代循环进行消息的传递，不断更新样本点的吸引度和归属度，最终形成高质量的类代表和对应的聚类。

样本吸引度和归属度的计算公式为

$$R(i,k)=\boldsymbol{S}(i,k)-\max_{j}\{A(k,j)+\boldsymbol{S}(i,j)\}$$
$$j\in\{1,2,\cdots,N\},\ j\neq k \tag{4}$$

$$A(i,k)=\min\{0,R(k,k)+\sum_j\{\max(0,R(j,k))\}$$
$$j\in\{1,2,\cdots,N\},\ j\neq k,\ j\neq i \tag{5}$$

偏向参数 $p(i)$ 和阻尼因子 l 是 Affinity Propagation 聚类的 2 个参数，偏向参数 $p(i)$ 表示样本点 i 被选为聚类中心的倾向性，置于相似性矩阵 $\boldsymbol{S}$ 对角线上，因此，当 $i=k$ 时：

$$R(k,k)=p(k)-\max_{j}\{A(k,j)+\boldsymbol{S}(k,j)\} \tag{6}$$

阻尼因子 l 反映了本次迭代过程中 R_i 和 A_i 与上一次 R_{i-1} 和 A_{i-1} 之间的关系，即：

$$R_i=R_i\times(1-l)+R_{i-1}\times l \tag{7}$$

$$A_i=A_i\times(1-l)+A_{i-1}\times l \tag{8}$$

本文将偏向参数 $p(k)$ ($k\in\{1,2,\cdots,N\}$)设定为相似性矩阵 $\boldsymbol{S}$ 元素的中值，阻尼因子 l 文本取值为 0.5。

本文算法主要流程如下：

(1)定义镜头边界 C 为空数列，滑动窗大小 W=300。

(2)初始化偏向参数 $p=S_{midia}$，定义最大循环次数 $mcount$= 60 000，收敛条件为聚类中心 50 次循环无变化，初始化吸引度 R_i 和归属度 A_i 为 0。

(3)按照式(4)~式(8)进行循环迭代，不断更新吸引度 R_i 和归属度 A_i，算法产生 k 个聚类中心。

(4)判断是否收敛，若收敛或者循环次数到达最大循环次数，则转入流程(5)，否则转流程(3)。

(5)输出聚类中心和聚类结果，即对应的帧号及其所属的类别。当帧号所属的类别发生变化时，查看对应的帧间差 d_i，当 $d_i>T_h$ 时，则认为此处发生镜头切变，或者存在着闪光灯，查看此处帧的前后帧之间的相似度，若相似，则认为是闪光灯的影响，否则认为是镜头发生切变，把其对应的帧号放入数组 C 中，当存在至少 2 个连续的或趋于连续(即帧号所属类别改变间隔小于 4)的帧间差 $d_i>T_l$，则认为此处可能发生渐变，也可能是噪声影响或者物体运动，查看此处帧与其相隔 4 帧的相似性是否呈现单调性递减，容忍度为 1 帧(防噪声干扰)，若是则认为此处发生渐变，放入数组 C 中，其中，T_h 是一个较大的值，T_l 是一个较小的值，转入流程(6)。

(6)判断视频帧是否到达视频结束帧，若是，则算法结束，否则以流程(5)中得到的镜头边界数列 C 中最后一个镜头边界为起始帧，向后滑动 W=300 帧，转入流程(2)，若起始帧距离视频结束帧不足 300 帧，则向后滑动至视频结束帧，转入流程(2)。

5 实验结果

为了验证上述算法的有效性，本文采用不同类型的实验视频，包括 MTV、新闻、体育节目等不同类型的视频，视频序列从几百帧到几万帧，视频帧大小各不相同。其实验数据集基本情况如表 1 所示。

表 1 实验数据集

样本	帧数	突变镜头数	渐变镜头数
MTV	2 053	26	57
新闻	557	23	0
体育	14 962	45	23

通过查全率和查准率这 2 个参数作为镜头边界检测的评价标准，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)

查全率和查准率越高，说明检测的算法越好，为比较算法的有效性，把通过本算法检测的结果与自适应阈值算法的检测结果做比较，通过本算法得到的实验结果如表 2 所示。通过自适应阈值检测算法的结果如表 3 所示。

表 2 本文算法实验结果

样本	正确检出数	漏检数	误检数	查全率/(%)	查准率/(%)
MTV	74	9	7	89.2	91.4
新闻	22	1	0	95.7	100.0
体育	63	5	8	92.6	88.7

表 3 自适应阈值算法实验结果

样本	正确检出数	漏检数	误检数	查全率/(%)	查准率/(%)
MTV	71	12	16	85.5	81.6
新闻	22	1	3	95.7	88.0
体育	61	7	21	89.7	74.4

通过算法检测结果可以看出，通过本算法检测的查全率和查准率都有一定的提高。对于 MTV 视频，漏检很多，原因是在 MTV 视频中，部分发生渐变镜头时，前后镜头背景相似，从而导致了漏检。体育视频误检很多是因为部分视频片段在目标对象剧烈运动的同时发生了摄像机的快速移动，从而导致了误检测。

6 结束语

本文提出基于 Affinity Propagation 聚类的镜头边界检测算法，克服运用 K 均值算法时难以确定聚类个数以及初始聚类中心的随机性带来结果的不稳定性，本算法能够根据视频本身的信息分布，自动检测出镜头边界，在一系列不同类型的视频片断上进行实验和分析，证明了该算法的有效性和快速性。

(下转第 237 页)