一种用于电视新闻节目的播音员镜头检测算法*

杨 娜, 罗航哉, 薛向阳

(复旦大学 计算机科学与工程系,上海 200433)
E-mail: xyxue@fudan.edu.cn
http://www.fudan.edu.cn

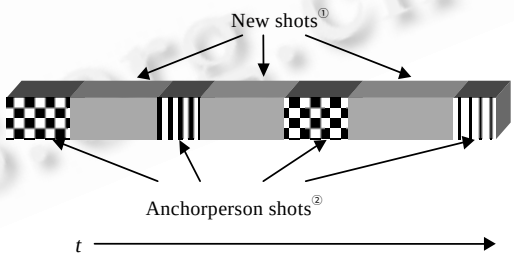
摘要: 提出一种播音员镜头自动检测算法.利用播音员镜头频繁出现的统计规律,基于自动聚类方法找出候选播音员镜头,然后根据播音员镜头出现的时空特征,用神经网络分类器对候选播音员镜头进行确认,从而实现对播音员镜头的检测.该算法是自动分析电视新闻节目内容的重要手段之一,对建立视频数据库索引是必不可少的.实验结果表明,该方法具有准确、快速的特点,可有效地应用于视频检索系统.

关 键 词: 视频信息检索;多媒体数据库;视频内容分析;播音员镜头检测;视频数据库索引
中图法分类号: TP391 文献标识码: A

近年来,随着存储成本的降低、传输速率的提高以及高效数据压缩技术的发展,因特网上视频信息大量涌现,人们访问并利用视频信息的需求日益增长.视频信息内容丰富、信息量大,如何对海量视频信息进行有效的索引、浏览和检索已成为当前信息检索领域非常重要的研究课题.

一般认为,视频内容可以表示成从低到高的 4 个层次,即关键帧-镜头-场景-视频.镜头是由一个摄像机从打开到关闭过程中拍摄的一串连续的帧序列.关键帧是从一个镜头中选取的具有代表性的图像帧.场景是表达同一语义信息的连续镜头序列,它是视频信息的最小语义单位.目前人们对镜头分割和关键帧提取[1,2]的研究已较成熟.由于镜头不能表达完整的语义,因此人们致力于视频语义自动分析的研究,并且已开发出一些原型系统.例如,美国 MITRE 公司信息系统研究中心开发的新闻节目导航系统[3].不过,目前系统在准确性和实时性上远远不能满足实际需求,原因在于快速、准确地进行视频语义分析是一个难题.

现实世界中,视频节目类型多种多样,常见的有电影、新闻、体育、商业广告节目等,它们均具有各自的特点.例如,新闻节目的结构在时间序列上较为固定,通常播音员镜头和新闻故事单元交替出现.图 1 给出了上海电视台新闻节目的一个例子.由于新闻故事内容多样,缺乏结构性,利用本身的镜头信息难以识别;而播音员镜头往往内容变化不大,运动较小,用一定的方法可以实现较准确的识别.播音员镜头可以作为新闻场景分割的边界,因此,对播音员镜头的检测成为新闻节目内容分析的重要手段之一.



①新闻故事单元,②播音员镜头.

Fig.1 The temporal structure of news program
图 1 新闻节目的结构

在文献[4]提出的新闻内容自动分析和索引系统中,采用模板匹配的方法实现播音员镜头检测,分 3 个步骤:

* 收稿日期: 2001-09-15; 修改日期: 2002-03-15
基金项目: 国家自然科学基金资助项目(60003017,69935010);国家高技术研究发展计划资助项目(2001AA114120);上海市科技启明星计划资助项目(01QD14013);上海市科技发展基金项目(015115044)
作者简介: 杨娜(1978–),女,河北辛集人,硕士生,主要研究领域为视频信息检索研究;罗航哉(1977–),男,四川泸县人,助教,主要研究领域为视频信息检索;薛向阳(1968–),男,江苏盐城人,教授,博士生导师,主要研究领域为多媒体信息检索.

利用关键帧时间和空间结构的先验知识建立播音员镜头模板,然后把候选播音员镜头和模板进行匹配,根据某种相似度度量决定其是否为播音员镜头,最后根据整段新闻的时间信息确定真正的播音员镜头.该算法为了得到候选播音员镜头,利用播音员镜头运动较小的特点,采取逐帧比较的方法,找到相邻帧间变化小于一定阈值的镜头作为候选镜头.由于帧数量很多,所需计算量巨大.另外,利用模板检测播音员镜头的方法缺乏通用性,当播音员位置或播音室环境发生变化时,该模板便失去作用.

IBM 公司 T.J.Watson 研究中心的 C.Dorai 等人[5]采用特征和启发式规则相结合的方法实现播音员镜头检测.首先利用颜色特征对镜头聚类,得到候选播音员镜头类,再用人脸检测滤除不存在大尺度人脸的候选镜头类,然后根据镜头时间序列规则得到真正的播音员镜头.该方法的特点是速度快,但由于人脸检测准确率不高,规则也难以具有通用性,因此检测精确度较低.

本文提出一种新的播音员镜头检测算法.该算法先采用聚类方法得到类似播音员的候选镜头类,然后统计候选镜头类的时间和空间特征,再用分类方法确认最终的播音员镜头类.由于聚类时只对关键帧进行处理,后期的分类建立在镜头类特征基础上,因此计算量小,可实现快速检测.同时,根据镜头的统计信息对播音员镜头进行检测,不受衣着、背景等变化的影响,具有较强的通用性.下面我们将对该算法的基本原理及实验过程进行详细介绍.

1 播音员检测算法

在一段新闻节目中,播音员镜头的出现具有一定的规律性.首先,新闻通常以播音员镜头开始和结束,播音员镜头在新闻中多次出现(本文称重复出现的播音员镜头为同类镜头);其次,播音员镜头运动较小,一般只存在播音员的微小动作变化;再次,同一类播音员镜头之间非常相似,播音员衣着和播音室背景等一般不会改变太大;最后,播音员镜头比一般新闻镜头时间长,并且播音员镜头之间存在一定时间间隔.根据这些统计规律,首先利用聚类得到候选播音员镜头类,然后提取这些镜头类的时空特征,再用分类器将候选播音员镜头类分为播音员镜头类和非播音员镜头类.

1.1 检测候选播音员镜头

根据同一播音员镜头在新闻节目中重复出现的特点,基于视觉内容对镜头进行聚类.容易知道,播音员镜头类中包含多个镜头,根据类的大小可初步确定播音员镜头类.下面介绍聚类过程中的算法细节.

1.1.1 关键帧提取

在聚类过程中,用关键帧作为镜头代表.由于播音员镜头内部变化不大,少量关键帧即可表示镜头内容.提取关键帧的方法有多种[6],但是这些方法往往需要复杂计算,难以实现实时计算.在本文中,简单地取镜头第 n 帧和第 $L–n$ 帧作为关键帧(L 为镜头长度),这样既能消除淡入、淡出等渐变镜头切换的影响,又节省了计算量.

1.1.2 视觉特征提取

从关键帧中提取分块颜色直方图作为特征计算镜头之间的相似度.由于同类镜头在颜色分布上极为相似,同时直方图对微小运动不敏感,因此以颜色直方图为特征的同类播音员镜头间相似度很大.考虑到播音员帧的视觉内容在空间上的分布特征,图 2 是一帧典型的播音员图像.中间为播音员,右上角为新闻提要,下方为标题信息.新闻提要和标题信息在某些播音员镜头中不出现,而在另外一些播音员镜头中动态出现或改变.为减少这些信息对播音员镜头间距离计算的影响,采用分块方法,并利用优化方法得到各块权重.在实验中,关键帧被等分为 64 小块,每块分别提取 YUV 颜色空间的两个一维色度直方图作为描述该帧的视觉特征.

①播音员,②标题,③新闻提要,④背景.

Fig.2 The spatial structure of a typical anchorperson frame
图 2 典型播音员帧的空间结构

1.1.3 距离的计算

该算法采用 L_1 距离来计算两帧间对应的分块颜色直方图距离,两帧间距离为

$$d(i,j)=\sum_{l=1}^{n}\omega_l Dist(Hist_i(l),Hist_j(l)), \tag{1}$$

其中 n 为直方图个数,ω_l 为第 l 个直方图的权重,$Dist(Hist_i(l),Hist_j(l))$ 为第 i 帧和第 j 帧对应的第 l 个直方图的 L_1 距离.$\{\omega_l\}$ 可以由下面介绍的训练算法获得.

设镜头 s 的两个关键帧为 KF_1^s 和 KF_2^s,则镜头 s_1 和 s_2 之间的距离定义为

$$D(s_1,s_2)=\min(d(KF_1^{s_1},KF_1^{s_2}),d(KF_1^{s_1},KF_2^{s_2}),d(KF_2^{s_1},KF_1^{s_2}),d(KF_2^{s_1},KF_2^{s_2})). \tag{2}$$

1.1.4 聚类算法

根据新闻节目中同一组播音员镜头之间非常相似、不同组播音员镜头以及非播音员镜头之间相似度较小的特点,采用一种快速聚类方法,该方法具有计算量小、度快的特点.基本步骤如下:

(1) 设一段视频中的镜头集合为 $\Theta=\{s_1,s_2,...,s_n\}$,利用式(2)计算两两镜头之间的距离,得到距离矩阵 $\Delta=\left[d_{i,j}\middle|1\le i,j\le n,d_{ij}=d(s_i,s_j)\right]$.

(2) 对于给定的阈值 T,若 $d_{i,j}<T$,则镜头 i 和 j 属于同一类.这样可以把 Θ 分为若干个类.阈值 T 可由训练数据计算得到.

(3) 定义包含的镜头个数不小于 η (在实验中取 $\eta=2$)的类为候选播音员镜头类;候选播音员镜头类中包含的镜头为候选播音员镜头.根据这一定义,得到候选播音员镜头类和候选播音员镜头.

1.1.5 权矩阵和阈值的计算

在聚类算法中,人工确定参数 $\{\omega_l\}$ 和 T 是困难的.这里提出一种通过手工标注自动确定参数的方法.

1.1.5.1 训练数据的手工标注

为训练出优化的参数,需对训练数据进行手工标注以得到镜头之间的分类信息.由于只是对播音员镜头进行识别,非播音员镜头的聚类错误对此没有影响,因此在训练中只采用播音员镜头的聚类信息,不考虑非播音员镜头,以减少手工标注的工作量.标注时,首先找出训练数据中所有播音员镜头,然后根据训练数据在视觉特征上的相似性对播音员镜头进行归类,最终得到若干类播音员镜头.

1.1.5.2 权矩阵的计算

假设训练数据中有 M 类手工标注的播音员镜头:$\overline{\Theta}=\{\overline{\Theta}_1,\overline{\Theta}_2,...,\overline{\Theta}_M\}$,每个类中播音员镜头数分别为 $m_1,m_2,...,m_M$.播音员镜头类内距离较小将有助于在聚类过程中把相似的播音员镜头聚为同一类,并避免了把非播音员镜头和播音员镜头归为一类.为了使参加训练的各播音员镜头类内距离和最小,采用如下数学模型计算最佳权矩阵.

$$\min L=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\ne j}}^{m_k}D(s_i,s_j). \tag{3}$$

如果直接将由(2)定义的距离函数代入式(3),则问题变得相当复杂.为使求解过程简单,首先用距离函数 $d(KF_1^{s_i},KF_1^{s_j})$ 替换式(3)中的 $D(s_i,s_j)$:

$$\min \tilde{L}=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\ne j}}^{m_k}d(KF_1^{s_i},KF_1^{s_j}). \tag{4}$$

式(4)可以用 Lagrange 乘数法求解.而式(3)的求解可以在式(4)的求解基础上得到.下面用 Lagrange 乘数法求解式(4),并进一步给出式(3)的求解方法,并从数学上证明该解法的正确性.

公式**(4)**的 **Lagrange** 乘数解法

$$\min \tilde{L}=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\ne j}}^{m_k}d(KF_1^{s_i},KF_1^{s_j})=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\ne j}}^{m_k}\sum_{l=1}^{n}\omega_l Dist(Hist_l(KF_1^{s_i}),Hist_l(KF_1^{s_j})). \tag{5}$$

取约束条件

$$\sum_{l=1}^{n}\frac{1}{\omega_l}=1, \tag{6}$$

则
$$\tilde{L}=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k}\sum_{l=1}^{n}\omega_l Dist(Hist_l(KF_1^{s_i}),Hist_l(KF_1^{s_j}))-\lambda\left(\sum_{l=1}^{n}\frac{1}{\omega_l}-1\right).$$

令 $d_l=Dist(Hist_l(KF_1^{s_i}),Hist_l(KF_1^{s_j}))$,下面求解 ω_l:

$$\frac{\partial L}{\partial\omega_l}=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k}d_l+\lambda\left(\frac{1}{\omega_l^2}\right)=0. \tag{7}$$

经推导可求出

$$\omega_l=\sum_{g=1}^{n}\frac{\sqrt{f_g}}{\sqrt{f_l}},\text{其中 } f_g=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k}d_g. \tag{8}$$

根据 Lagrange 乘数法理论,上面过程得到的权矩阵 $\{\omega_l\}$ 不一定是 $\tilde{L}$ 的极值点.若二次型是正定的,那么目标函数 $\tilde{L}$ 在点 $\{\omega_l\}$ 取约束极小值.在附录中将给出 $\Delta\tilde{L}$ 是正定的证明.

$$\Delta\tilde{L}=(\Delta\omega_1,...,\Delta\omega_n)\begin{pmatrix}\frac{\partial^2 L}{\partial\omega_1^2} & \frac{\partial^2 L}{\partial\omega_1\partial\omega_2} & \cdots & \frac{\partial^2 L}{\partial\omega_1\partial\omega_n}\\ \frac{\partial^2 L}{\partial\omega_2\partial\omega_1} & \frac{\partial^2 L}{\partial\omega_2^2} & \cdots & \frac{\partial^2 L}{\partial\omega_2\partial\omega_n}\\ \cdots & \cdots & \cdots & \\ \frac{\partial^2 L}{\partial\omega_n\partial\omega_1} & \frac{\partial^2 L}{\partial\omega_n\partial\omega_2} & \cdots & \frac{\partial^2 L}{\partial\omega_n^2}\end{pmatrix}\begin{pmatrix}\Delta\omega_1\\ \cdots\\ \Delta\omega_n\end{pmatrix}. \tag{9}$$

公式**(3)**的求解

当距离函数取为式(2)时,不能直接用上面的解法.取定一组满足约束式(6)的权矩阵初值 $\{\omega_l^0\}$,则 $D^0(s_i,s_j)=D(\{\omega_l^0\},s_i,s_j)$ 为确定常数,可以计算 $\hat{L}^0=L(D^0(s_i,s_j))$.并且,对任何一对镜头 (s_i,s_j) 来说,使

$$D^0(s_i,s_j)=d^0(KF_{a^0}^{s_i},KF_{b^0}^{s_j}) \tag{10}$$

的两个关键帧 $KF_{a^0}^{s_i}$ 和 $KF_{b^0}^{s_j}$ 也随之确定.其中 $a^0=a(\{\omega_l^0\},s_i,s_j)$; $b^0=b(\{\omega_l^0\},s_i,s_j)$; $a^0,b^0=1$ 或 2 为关键帧的编号.

定义

$$d_l^0=Dist(Hist_l(KF_{a^0}^{s_i}),Hist_l(KF_{b^0}^{s_j})), \tag{11}$$

从而可以通过式(8)计算出新的 $\{\omega_l^1\}$.

定义

$$TL^1=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k}\sum_{l=1}^{n}\omega_l^1 Dist\left(Hist_l(KF_{a^0}^{s_i}),Hist_l(KF_{b^0}^{s_j})\right)=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k}\tilde{d}^1(KF_{a^0}^{s_i},KF_{b^0}^{s_j}).$$

根据附录的证明可知 $TL^1\le\hat{L}^0$,而

$$D^1(s_i,s_j)=\min\left\{\begin{array}{ll}\tilde{d}^1(KF_{a^0}^{s_i},KF_{b^0}^{s_j}), & \tilde{d}^1(KF_{|a^0-2|+1}^{s_i},KF_{b^0}^{s_j}),\\ \tilde{d}^1(KF_{a^0}^{s_i},KF_{|b^0-2|+1}^{s_j}), & \tilde{d}^1(KF_{|a^0-2|+1}^{s_i},KF_{|b^0-2|+1}^{s_j})\end{array}\right\}$$
$$\le\tilde{d}^1(KF_{a^0}^{s_i},KF_{b^0}^{s_j}),$$

所以,

$$\hat{L}^1=\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k}D^1\left(s_i,s_j\right)\le\sum_{k=1}^{M}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k}\tilde{d}^1\left(KF_{a^0}^{s_i},KF_{b^0}^{s_j}\right)=TL^1\le\hat{L}^0.$$

于是可以得到如下的迭代算法:

(a) 取 $\omega_l^0=n,1\le l\le n$, $t=1$,可求得 $\hat{L}^0$;

(b) 根据式(10)求解 a^{t-1},b^{t-1};

(c) 根据式(11)求解 $d_l^{t-1},1\le l\le n$;

(d) 根据式(8)求解 $\omega_l^t,1\le l\le n$;

(e) 根据式(3)求 $\hat{L}^t$;

(f) 如果 $\left(\hat{L}^t-\hat{L}^{t-1}\right)\le\varepsilon$ 或 $t\ge N$,其中 ε 为给定正数,N 为给定正整数,则停止,否则 $t\leftarrow t+1$,转步骤(b).

从上面推导过程可知,这个迭代算法是收敛的.

1.1.5.3 阈值的计算

定义 D_k 为手工标注的第 k 个播音员镜头类中任意两镜头间的最大距离.取阈值 $T=a\max\{D_1,...,D_M\}$,其中 a 为常数,M 为播音员镜头类的数目.

1.2 候选播音员镜头分类

某些非播音员镜头在新闻中也会重复出现,因而被检测为候选播音员镜头.然而,播音员镜头和非播音员镜头的出现在时间和空间上的统计规律明显不同,可以采用统计方法提取候选播音员镜头类的时空特征,并利用分类器进行分类,最后得到播音员镜头类.

1.2.1 特征提取

在新闻节目中,播音员镜头和非播音员镜头在时间和空间特征上的不同表现在于:播音员镜头重复出现次数较多,而非播音员镜头重复出现次数较少;同类的相邻播音员镜头间存在一定的时间间隔,而同样多次重复出现的天气预报等非播音员镜头则经常连续出现;通常播音员镜头需持续一定时间,以至少播报一条新闻,故同类播音员镜头的总长度较长;播音室环境与一般采访场所相比,光线稳定,故播音员镜头关键帧的对比度变化范围不大,而非播音员镜头关键帧则无此特性.基于上述差异,我们选取:(1) 类中镜头数目;(2) 类中镜头总长度;(3) 类中任意两镜头间最大间隔;(4) 类中相邻镜头最小间隔;(5) 类中相邻镜头平均间隔;(6) 类中各镜头亮度分量的平均对比度[7]作为候选播音员镜头类的特征.综合考虑这几种特征,将有效实现播音员镜头和非播音员镜头的判别.

1.2.2 分类器选择

模式识别中有多种分类工具,包括几何分类器、支撑矢量机、神经网络分类器、模糊分类器等.各种分类器适用于不同的领域,分类效果亦存在差异.选择合适的分类器是影响分类效果的因素之一.本文采用 3 层感知器[8]作为分类器,它是一种非线性分类器,可以产生任意复杂的判决区,实验证明,它对候选播音员镜头类具有很好的分类性能.

2 算法评价

用 8.5 个小时的新闻节目作为实验数据(来自上海电视台的 MPEG-1 视频流),其中 3 小时节目作为训练数据集,剩余 5.5 小时节目作为测试数据集.所有节目的镜头边界信息已在实验前由镜头边界检测系统自动获得.在实验过程中,首先利用训练数据计算出聚类阈值 T 和权矩阵 W.然后对测试数据提取每个镜头的关键帧,根据关键帧的视觉特征对镜头进行聚类,得到候选播音员镜头类.在分类之前,对训练数据进行聚类,并对得到的候选播音员镜头类进行手工标注,作为训练数据对分类器进行训练.然后,用训练好的分类器对每个测试数据集中的候选镜头类进行判决.图 3 的两帧图像分别从训练数据集中典型的播音员镜头类和非播音员镜头类中获得.这两类镜头在时间特征上有各自的特点,用 3 层感知器可以很好地区分.用 $F=\dfrac{(\beta^2+1)RP}{\beta^2R+P}$ 衡量播音员镜头的检测性能.其中 R 为召回率,P 为精度,β 为常数.


(a) Key frame from anchorperson shot (b) Key frame from non-anchorperson shot

(a) 播音员镜头类 (b) 非播音员镜头类

Fig.3 Key frames from potential anchorperson shots

图 3 候选播音员镜头类

表 1 列出了播音员镜头检测的结果,可以看出,该算法具有较高的检测性能.造成检测错误的原因是某些包含镜头数较少的播音员镜头类在时间特征上与非播音员镜头类非常相似,从而导致误分;而某些非播音员镜头重复出现,其时空特征与播音员镜头相似,造成分类错误.N_{total} 表示数据集中实际播音员镜头数;N_{idenp} 表示自动检测到的正确的播音员镜头数;N_{false} 表示虚警数;N_{mis} 表示漏检的播音员镜头数;Recall 表示召回率(N_{idenp}/N_{total});Precision 表示精度(N_{idenp}/(N_{idenp}+N_{false})).

Table 1 Experimental results in detecting anchorperson shots ($\alpha=1.1, \beta=1.0$)

表 1 播音员检测算法实验结果($\alpha=1.1, \beta=1.0$)

Video①	N_{total}	N_{idenp}	N_{false}	N_{mis}	Recall (%)	Precision (%)	F (%)
Training data set②(3 h)	184	184	0	0	100	100	100
Testing data set③(5.5 h)	223	221	4	2	99.1	98.22	98.66

①视频,②训练数据集,③测试数据集.

图 4 为各种检测错误的播音员镜头.图 4(a)的播音员镜头所属类中只包含两个镜头,并且两个镜头的间隔较小,它在特征上与某些非播音员镜头类似,因而被错误判决为非播音员镜头.图 4(b)为采访镜头,该类镜头的镜头长度与镜头间隔等特征都非常类似于播音员镜头类,故难以正确判决.


(a) Missed by classification (b) False positive

(a) 漏检的播音员镜头类 (b) 采访镜头

Fig.4 Examples of false detection

图 4 错误检测实例

在聚类过程中选取不同的阈值会得到不同的聚类结果,从而对播音员镜头检测造成影响.图 5 给出了播音员镜头检测在不同阈值下的结果.阈值较小时,同一类播音员镜头可能被分成几个类,这些小类独立地参与分类器的训练和分类过程,大部分被正确判决;当阈值较大时,相似的播音员镜头类可能被聚为一类,这些类同样可以被正确地判决.但是当阈值过大时,聚类过程将无法区分播音员镜头和非播音员镜头.实验表明,当α值大于 1.2 时,训练数据中的某些播音员镜头将和非播音员镜头聚为同一类,此时的训练将失去意义;当阈值过小时,聚类产生的镜头类特征不足以体现播音员镜头类与非播音员镜头类的不同,因此当α取(0.4,1.2)范围内的值时,性能较稳定,α在(1.0,1.2)范围内时性能达到最优.

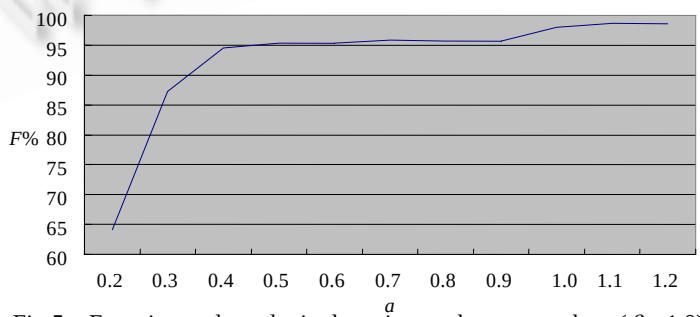


Fig.5 Experimental results in detecting anchorperson shots ($\beta=1.0$)

图 5 播音员镜头检测结果 ($\beta=1.0$)

实验还表明,本文提出的算法速度非常快.通过对全部 8.5 小时的电视新闻节目作速度测试,共需约 12 分钟(不包括自动镜头分割时间),机器配置为 733MHz P3 CPU 128MB.

3 结论与将来工作

本文提出了一种用于数字电视新闻节目播音员镜头检测的快速有效算法,该算法在镜头分割的基础上,先由镜头关键帧的视觉特征聚类得到候选播音员镜头类,再利用新闻节目中播音员镜头出现的时间特征对候选镜头自动分类,以得到真正的播音员镜头.检测结果可作为新闻内容分析的依据,以便于用户对新闻的层次浏览和检索.算法思想亦可用于娱乐、谈话节目中主持人的检测,并在此基础上作节目的内容分析.由于本算法的聚类基于镜头的关键帧,大大减少了计算量,可实现快速检测.在分类过程中,利用了镜头的时间序列特征,检测精度很高.但是,在某些情况下,这些特征还不能充分刻画播音员镜头类与非播音员镜头类的差异,造成分类错误.今后我们将对播音员镜头作伴音内容分析,综合伴音特征来提高检测性能.

References:

[1] Idris, F., Panchanathan, S. Review of image and video indexing techniques. Visual computation and image representation, 1997,8(2):146~166.

[2] Rui, Y, Huang, T.S., Mehrotra, S. Constructing table-of-content for videos. Multimedia Systems, 1999,7(5):359~368.

[3] Merlino, A., Morey, D., Maybury, M. Broadcast news navigation using story segmentation. In: Proceedings of the 5th ACM International Conference on Multimedia'97. Seattle: ACM Press, 1977. 381~391. http://www.informatik.uni-trier.de/~ley /db/conf /mm/mm97.html.

[4] Zhang, H.J., Gong, Y., Smoliar, S.W., *et al*. Automatic parsing of news video. In: Proceedings of the International Conference on Multimedia Computing and Systems. 1994. 45~54. http://www.informatik.uni-trier.de/~ley/db/conf/icmcs /icmcs94.html.

[5] Shearer, K., Dorai, C., Venkatesh, S. Incorporating domain knowledge with video and voice data analysis in news broadcasts. In: Proceedings of the 1st International Workshop on Multimedia Data Mining MDM/KDD 2000. Boston, 2000. 46~53. http://www.cs.ualberta.ca/~zaiane/mdm_kdd2000.

[6] Zhang, Y., Rui, Y., Huang, T.S., *et al*. Adaptive key frame extraction using unsupervised clustering. In: Proceedings of the IEEE International Conference on Image Processing. 1998. 886~870. http://citeseer.nj.nec.com/47138.html.

[7] Tamura, H., Mori, S. Textural features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics, 1978,8(6):460~472.

[8] Bian, Zhao-qi, Zhang, Xue-gong. Pattern Recognition. 2nd ed., Beijing: Tsinghua University Press, 2000. 254~257 (in Chinese).

附中文参考文献:

[8] 边肇祺,张学工.模式识别.第 2 版,北京:清华大学出版社,2000.254~257.

附录

公式(9)正定性的证明.

$$\Delta\tilde{L} = (\Delta\omega_1,...,\Delta\omega_n)\begin{pmatrix} \frac{\partial^2 L}{\partial\omega_1^2} & \frac{\partial^2 L}{\partial\omega_1\partial\omega_2} & \cdots & \frac{\partial^2 L}{\partial\omega_1\partial\omega_n} \\ \frac{\partial^2 L}{\partial\omega_2\partial\omega_1} & \frac{\partial^2 L}{\partial\omega_2^2} & \cdots & \frac{\partial^2 L}{\partial\omega_2\partial\omega_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 L}{\partial\omega_n\partial\omega_1} & \frac{\partial^2 L}{\partial\omega_n\partial\omega_2} & \cdots & \frac{\partial^2 L}{\partial\omega_n^2} \end{pmatrix}\begin{pmatrix} \Delta\omega_1 \\ \cdots \\ \Delta\omega_n \end{pmatrix},$$

令 $$M=\begin{pmatrix}\frac{\partial^2 L}{\partial\omega_1^2} & \frac{\partial^2 L}{\partial\omega_1\partial\omega_2} & \cdots & \frac{\partial^2 L}{\partial\omega_1\partial\omega_n}\\ \frac{\partial^2 L}{\partial\omega_2\partial\omega_1} & \frac{\partial^2 L}{\partial\omega_2^2} & \cdots & \frac{\partial^2 L}{\partial\omega_2\partial\omega_n}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial^2 L}{\partial\omega_n\partial\omega_1} & \frac{\partial^2 L}{\partial\omega_n\partial\omega_2} & \cdots & \frac{\partial^2 L}{\partial\omega_n^2}\end{pmatrix}=\begin{pmatrix}\frac{\sum_{l=1}^{n}\omega_l f_l}{\omega_1^3} & -\frac{f_2}{\omega_1^2} & \cdots & -\frac{f_n}{\omega_1^2}\\ -\frac{f_1}{\omega_2^2} & \frac{\sum_{l=1}^{n}\omega_l f_l}{\omega_2^3} & \cdots & -\frac{f_n}{\omega_2^2}\\ \cdots & \cdots & \cdots & \cdots\\ -\frac{f_1}{\omega_n^2} & -\frac{f_2}{\omega_n^2} & \cdots & \frac{\sum_{l=1}^{n}\omega_l f_l}{\omega_n^3}\end{pmatrix},$$

由于权矩阵的求解对于各个 ω_l 是顺序无关的,因此可以设定 $f_1\ge f_2\ge\ldots\ge f_n>0$.

公式(8)$\Rightarrow f_l\omega_l^2=\left(\sum_{k=1}^{n}\sqrt{f_k}\right)^2=c,(1\le l\le n)\Rightarrow f_i\omega_i^2=f_j\omega_j^2,1\le i,j\le n,$

所以 $$\frac{f_i}{\omega_j^2}=\frac{f_j}{\omega_i^2},1\le i,j\le n.$$

因此,矩阵 M 为实对称阵.M 的顺序主子式均大于 0 是二次型 ΔL 正定的充要条件.

M 的顺序主子式均大于 0 的证明如下:

$$|M|=\begin{vmatrix}\frac{\sum_{l=1}^{n}\omega_l f_l}{\omega_1^3} & -\frac{f_2}{\omega_1^2} & \cdots & -\frac{f_n}{\omega_1^2}\\ -\frac{f_1}{\omega_2^2} & \frac{\sum_{l=1}^{n}\omega_l f_l}{\omega_2^3} & \cdots & -\frac{f_n}{\omega_2^2}\\ \cdots & \cdots & \cdots & \cdots\\ -\frac{f_1}{\omega_n^2} & -\frac{f_2}{\omega_n^2} & \cdots & \frac{\sum_{l=1}^{n}\omega_l f_l}{\omega_n^3}\end{vmatrix}=\prod_{i=1}^{n}\frac{1}{\omega_i^3}\begin{vmatrix}\sum_{l=1}^{n}\omega_l f_l & -\omega_2 f_2 & \cdots & -\omega_n f_n\\ -\omega_1 f_1 & \sum_{l=1}^{n}\omega_l f_l & \cdots & -\omega_n f_n\\ \cdots & \cdots & \cdots & \cdots\\ -\omega_1 f_1 & -\omega_2 f_2 & \cdots & \sum_{l=1}^{n}\omega_l f_l\end{vmatrix}$$

$$=\prod_{i=1}^{n}\frac{1}{\omega_i^3}\begin{vmatrix}\sum_{l=1}^{n}\sqrt{c}\sqrt{f_l} & -\sqrt{c}\sqrt{f_2} & \cdots & -\sqrt{c}\sqrt{f_n}\\ -\sqrt{c}\sqrt{f_1} & \sum_{l=1}^{n}\sqrt{c}\sqrt{f_l} & \cdots & -\sqrt{c}\sqrt{f_n}\\ \cdots & \cdots & \cdots & \cdots\\ -\sqrt{c}\sqrt{f_1} & -\sqrt{c}\sqrt{f_2} & \cdots & \sum_{l=1}^{n}\sqrt{c}\sqrt{f_l}\end{vmatrix}=\left(\sqrt{c}\right)^n\prod_{i=1}^{n}\frac{1}{\omega_i^3}\begin{vmatrix}\sum_{l=1}^{n}\sqrt{f_l} & -\sqrt{f_2} & \cdots & -\sqrt{f_n}\\ -\sqrt{f_1} & \sum_{l=1}^{n}\sqrt{f_l} & \cdots & -\sqrt{f_n}\\ \cdots & \cdots & \cdots & \cdots\\ -\sqrt{f_1} & -\sqrt{f_2} & \cdots & \sum_{l=1}^{n}\sqrt{f_l}\end{vmatrix}.$$

令 $g_i=t_i+\sum_{i=1}^{n}t_i$, $t_i=\sqrt{f_i}$,则:

$$\begin{vmatrix}\sum_{l=1}^{n}\sqrt{f_l} & -\sqrt{f_2} & \cdots & -\sqrt{f_n}\\ -\sqrt{f_1} & \sum_{l=1}^{n}\sqrt{f_l} & \cdots & -\sqrt{f_n}\\ \cdots & \cdots & \cdots & \cdots\\ -\sqrt{f_1} & -\sqrt{f_2} & \cdots & \sum_{l=1}^{n}\sqrt{f_l}\end{vmatrix}=\begin{vmatrix}\sum_{i=1}^{n}t_i & -t_2 & \cdots & -t_n\\ -t_1 & \sum_{i=1}^{n}t_i & \cdots & -t_n\\ \cdots & \cdots & \cdots & \cdots\\ -t_1 & -t_2 & \cdots & \sum_{i=1}^{n}t_i\end{vmatrix}=\begin{vmatrix}g_1 & -g_2 & 0 & \cdots & 0 & 0\\ 0 & g_2 & -g_3 & \cdots & 0 & 0\\ 0 & 0 & g_3 & \cdots & 0 & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & g_{m-1} & -g_m\\ -t_1 & -t_2 & -t_3 & \cdots & -t_{m-1} & \sum_{i=1}^{m}t_i\end{vmatrix}$$

$=\ldots=\left(\prod_{i=1}^{n}g_i\right)\bullet\left(\frac{t_n}{g_n}+\sum_{i=1}^{n-1}\left(\frac{t_i}{g_n}-\frac{t_i}{g_i}\right)\right)$.所以,

$$|M| = \left(\sqrt{c}\right)^n \bullet \left(\prod_{i=1}^{n} \frac{1}{\omega_i^3}\right) \bullet \left(\prod_{i=1}^{n} g_i\right) \bullet \left(\frac{t_n}{g_n} + \sum_{i=1}^{n-1}\left(\frac{t_i}{g_n} - \frac{t_i}{g_i}\right)\right).$$

因为 $f_1 \geq f_2 \geq \ldots \geq f_n > 0$,

所以,

$t_i \geq t_n > 0$, $g_i = \left(t_i + \sum_{i=1}^{m} t_i\right) \geq \left(t_n + \sum_{i=1}^{m} t_i\right) = g_n > 0$, $1 \leq i \leq n$,

$$=> \frac{t_i}{g_i} \leq \frac{t_i}{g_m} => \left(\frac{t_i}{g_n} - \frac{t_i}{g_i}\right) \geq 0 => \sum_{i=1}^{n}\left(\frac{t_i}{g_n} - \frac{t_i}{g_i}\right) \geq 0 => \frac{t_n}{g_n} + \sum_{i=1}^{n}\left(\frac{t_i}{g_n} - \frac{t_i}{g_i}\right) > 0$$

$$=> \left(\prod_{i=1}^{n} g_i\right) \bullet \left(\frac{t_n}{g_n} + \sum_{i=1}^{n-1}\left(\frac{t_i}{g_n} - \frac{t_i}{g_i}\right)\right) > 0$$

$$c = \left(\sum_{k=1}^{n} \sqrt{f_k}\right)^2 => c > 0 => \left(\sqrt{c}\right)^n > 0$$

$$\omega_l \geq 1 => \prod_{l=1}^{n} \frac{1}{\omega_l^3} > 0$$

所以, $|M| > 0$.

同理可证,M 的各阶顺序主子式大于 0.又 M 是实对称阵,所以 M 是正定的.命题得证. □

A Method to Detect Anchorperson Shots for Digital TV News*

YANG Na, LUO Hang-zai, XUE Xiang-yang

(*Department of Computer Science and Engineering, Fudan University, Shanghai* 200433, *China*)

E-mail: xyxue@fudan.edu.cn

http://www.fudan.edu.cn

Abstract: A new method is presented to detect anchorperson shots automatically for digital TV news programs. The algorithms are required for automatically parsing and indexing of news program. The identification process consists of two steps. The first step is that potential anchorperson shots are grouped by clustering based on the fact that anchorperson shots always present repeatedly within a news program. At the second step, a well-trained neural network classifier is used to identify the correct anchorperson shots according to temporal and spatial features of their occurrence law. Experimental results show that the proposed method can detect the anchor shots quickly with very high precision.

Key words: video information retrieval; multimedia database; video content parsing; anchorperson shot detection; video database indexing

* Received September 15, 2001; accepted March 15, 2002

Supported by the National Natural Science Foundation of China under Grant Nos.60003017, 69935010; the National High-Tech. Research and Development Plan of China under Grant No.2001AA114120; the Rising-Star Program of Shanghai under Grand No.01QD14013; the Foundation of Shanghai Science and Technology Development under Grant No.015115044