一种在MPEG压缩流上检测播音员镜头的快速算法*)

A fast Anchor Shot Detection Algorithm in Mpeg Compression Domain

王伟强[1] 高 文[1,2]

(中科院计算所数字化技术实验室 北京100080)[1]

(哈尔滨工业大学计算机科学与工程系 哈尔滨150001)[2]

Abstract Detecting anchor shots accurately is very improtant for automatically parsing news video and extracting meaningful news items. The paper proposes a fast anchor shot detection algorithm, which is based on backgroud choronminance and skin tone models. The attractive characteristics of the algorithm include only simple computation involved. At the same time, it operates in MPEG compression domain, which makes the detection speed very fast. The algorithm was evaluated on a big test set containing more than 480000 frames and news video from two different TV stations. More than 98% accuracy and 100% recall have been gained. The experiment results also show the system has an average detection speed of 77.55 f/s. The experiments demonstrate the algorithm is a fast and effective one.

Keywords Anchor shot, News video, Video parsing

1.引言

为了在视频或多媒体数据库中对视频信息进行有效的索引、浏览、检索,需要建立各种自动化工具对视频节目源进行结构、语义的分析,提取出刻画视频节目源内容的可供索引的特征。限于目前计算机视觉及音频信号分析技术的现状,从一般的视频节目中自动抽取语义信息还无法实现,但我们可以利用一定的先验知识模型建造特定类型视频节目的自动解析工具。电视新闻节目便是一类具有很强先验时间结构模型的视频节目。

一些研究者针对不同电视广播电台的新闻节目进行了视频分析技术的研究,如文[1~3]。它们对新闻项的分割均涉及对播音员镜头的检测,因为播音员镜头的检测对于实现新闻项的分割起到非常重要的作用,通常它标志着一段新闻条目的开始及(或)结束。文[3]首先利用工具生成新闻节目记录文字形式,然后寻找一些固定的语言模式,如"我是〈播音员姓名〉"等,并结合有关新闻节目的一些先验结构模型知识来确定播音员镜头。而文[1,2]通过图像分析技术来确定播音员镜头。首先,利用镜头分割技术将新闻节目分成一个镜头集,然后通过一个多阶段判定过程确定出其中的播音员镜头。该过程包括利用镜头中相邻帧间直方图与对应像素值两种度量量差值的均值与方差确定候选播音员镜头,利用区域模型对候选播音员镜头进行测试筛选,测试成功后将该镜头与已存在的模型帧图像匹配,若匹配失败,则基于该镜头构造添加一个新的模型帧图像,最后与那些不存在匹配成功的模型帧图像相关联的镜头被滤掉,余下的镜头判定为播音员镜头。文[1]利用两天的新加坡广播公司(SBC)的新闻做测试数据进行了评估实验,在镜头分割正确率100%的假设下报告了95%以上的播音员镜头检测正确率。文[4]结合音视频线索来检测播音员镜头。他们采用话者识别技术将音频流分成由不同话者为特征的片断,同时将表示各个镜头的关键帧进行聚类,并利用如下的启发性知识来选择播音员镜头:播音员的语音与影像在整个新闻节目中会具有较高的比例,且在时间轴上的分布也较为分散。因此,利用该知识模型对音频、视频两个通道的结果通过与关系来发现最具可能性的镜头。据报告,该系统具有98%的检测正确率。

本文提出了一种更为简单有效的基于模型的播音员镜头检测算法。同文[1~4]中算法相比,本算法具有如下特色:①在检测前不需要先进行镜头分割计算,因而检测算法性能不受镜头分割正确率的影响;②在压

*)本文的研究工作得到国家自然科学基金重点项目(69789301)、国家863计划项目(863-306-ZT03-01-2)和中科院百人计划的资助。王伟强 博士生,主要研究领域:多媒体技术、人工智能。高 文 教授、博士生导师,主要研究领域:多媒体数据压缩、图像处理、计算机视觉,多模式接口,人工智能、虚拟现实等。

缩域上完成对播音员镜头的检测，避免了全解码过程中 IDCT 带来的高昂计算费用；③检测过程的计算简单，运算的时间复杂度明显低于文[1]中的算法；④算法的检测过程可以在线一遍完成。

2. 压缩域播音员镜头快速检测算法

播音员镜头是新闻节目中的重要组成部分。在不同广播公司制作的新闻节目中，播音员镜头中的帧图像普遍具有两个共同元素，播音员与背景，如图1。在不同的日期，可能会发生播音员的改变，或者同一播音员会改变服饰，但通常整个背景的内容或其中局部区域的内容具有相对的稳定性，长期保持不变，且与一般的非播音员镜头的对应位置处的内容具有明显的视觉差异。我们假设在该内容具有长期不变性的区域中存在具有较一致色彩纹理的子区域 T。基于播音员镜头中区域 T 的色彩纹理特点，我们可以利用统计学习的方法建立起播音员镜头中该背景特征区域的色彩模型，并利用该模型来检测播音员镜头的出现及其起始终止位置。

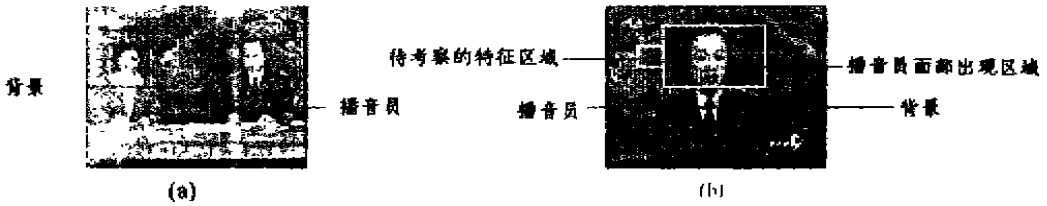

图1 来自 CCTV 的典型播音员镜头

2.1 算法描述

在 MPEG 编码的视频流中，每帧图像被分割成称作宏块的16 * 16像素块，每个宏块含有若干个8 * 8块。利用文[5,6]中的方法无需全部解码便可以从压缩视频流中提取出每帧的 DC 图像。每个块的 DC 系数代表了该块所有像素相应分量的均值，我们的算法是通过对 Cb、Cr 分量中的特定块的 DC 系数进行基于模型的检测来实现的。通过交互式地选取播音员镜头背景区域中某个内容长期不变且纹理较一致的子区域作为特征区 T，如图1(b)所示，便可以针对区域 T 建立其色彩模型，包括的重要信息有：在区域 T 中 Cb、Cr 分量块的 DC 系数值分布范围，以及区域 T 中所有块 DC 系数均值及方差的分布范围。由于播音员镜头的持续时间一般要5秒钟以上，为了提高检测速度，检测算法采用粗细两种粒度进行分阶段检测。首先，对码流中的 I 帧进行检测，在图片组 GOP 分辨率上确定出播音员镜头的出现位置，然后再在帧分辨率上进一步细化，确定出播音员镜头的起始、结束位置。整个算法的细节描述如下：

①初始化。打开视频流文件 fp，提取出第一个 I 帧的帧序号 CurFrmNum，以及码流中一个图片组 GOP(Group of Picture)的长度 gl。

②提取帧序号为 CurFrmNum 的帧的两个色度分量的 DC 图像 $X_{cb}=\{x_{ij}^{cb}\}$ 与 $x_{cr}=\{x_{ij}^{cr}\}$。

③设待考察的特征区域为 T，分量 X_{cb}，X_{cr} 中索引为 (i,j) 的元素对应的区域为 b_{ij}，令 $T=\{b_{ij}|b_{ij}\subseteq R\}$，$C_{cb}=\{c_{ij}^{cb}|c_{ij}^{cb}=x_{ij}^{cb},b_{ij}\in T\}$，$C_{cr}=\{c_{ij}^{cr}|c_{ij}^{cr}=x_{ij}^{cr},b_{ij}\in T\}$。计算用于检测标题文字出现的特征向量 $Fearture=(r_s^{cb},r_e^{cb},avg^{cb},sd^{cb},r_s^{cr},r_e^{cr},avg^{cr},sd^{cr})$，其中 $r_s^{cb}=\min_{d\in C_{cb}} d$，$r_e^{cb}=\max_{d\in C_{cb}} d$，$avg^{cb}=\frac{1}{|C_{cb}|}\sum_{d\in C_{cb}} d$，$sd^{cb}=\sqrt{\frac{1}{|C_{cb}|}\sum_{d\in C_{cb}}(d-avg^{cb})^2}$，$r_s^{cr}=\min_{d\in C_{cr}} d$，$r_e^{cr}=\max_{d\in C_{cr}} d$，$avg^{cr}=\frac{1}{|C_{cr}|}\sum_{d\in C_{cr}} d$，$sd^{cr}=\sqrt{\frac{1}{|C_{cr}|}\sum_{d\in C_{cr}}(d-avg^{cr})^2}$，其中 $|C|$ 表示集合 C 的基数。

④若 $[r_s^{cb},r_e^{cb}]\subseteq[v_s^{cb},v_e^{cb}]$，$avg^{cb}\in[avg_s^{cb},avg_e^{cb}]$，$sd^{cb}\in[sd_s^{cb},sd_e^{cb}]$，$[r_s^{cr},r_e^{cr}]\subseteq[v_s^{cr},v_e^{cr}]$，$avg^{cr}\in[avg_s^{cr},avg_e^{cr}]$，$sd^{cr}\in[sd_s^{cr},sd_e^{cr}]$ 均成立，则序号为 CurFrmNum 的帧属于某个播音员镜头的播音员帧，其中 $[v_s^{cb},v_e^{cb}]$，$[avg_s^{cb},avg_e^{cb}]$，$[sd_s^{cb},sd_e^{cb}]$，$[v_s^{cr},v_e^{cr}]$，$[avg_s^{cr},avg_e^{cr}]$，$[sd_s^{cr},sd_e^{cr}]$ 为系统利用统计学习的方法获得的有关特征向量 Fearture 各元素值的动态分布的特征范围。除了上面的检测条件外，还可以利用系统通过学习发现的两个色度分量间存在的一些关系统计特征来进一步提高检测的正确率，如对于 CCTV 新闻联播中播音员镜头我们选取的待考察的特征区域 R 存在如下关系特征：$avg^{cb}\geqslant avg^{cr}$。

⑤为了提高算法的抗噪性。在检测播音员镜头出现事件时，当且仅当系统在连续 W_s 个 I 帧中均检测到播音员帧的存在，系统才确认播音员镜头出现事件的发生，记下该播音员镜头在 GOP 分辨率下的起始帧号 SAnchorFrmNum。类似地，在检测播音员镜头结束事件时，仅当系统在连续 W_e 个 I 帧中均检测不到播

音员帧的存在，系统才确认播音员镜头结束事件的发生，记下该播音员镜头在GOP分辨率下的结束帧号EAnchorFrmNum。其中W_1、W_2为可由系统设定的参数，在我们对CCTV新闻联播的播音员镜头检测系统中$W_1=W_2=3$。

⑥搜索下一个I帧，令CurFrmNum=CurFrmNum+gl，若没有到流文件文件尾，则转②；否则转⑦。

⑦设经过上面的计算，产生了在GOP分辨率下播音员镜头集$\{S_i=(sf_i,ef_i)\}$，其中sf_i、ef_i分别表示播音员镜头S_i的起始I帧与结束I帧。对于每个S_i，分别在帧sf_i-gl与sf_i之间，以及帧ef_i与ef_i+gl之间，按照前面③④描述的计算方法，确定出在帧分辨率下镜头S_i的起始帧与结束帧。

⑧关闭流文件fp，整个检测过程结束。

经过上面基于背景特征区色彩模型的检测过程后，理想的情况会使所有的播音员镜头出现在结果的片断集中。但可能在结果的片断集中也会出现一些在待考察的特征区域中具有与播音员镜头具有相似纹理特征的新闻镜头集。为了滤掉这些错误的检测，我们可以通过人脸的肤色来进一步确定某个片断是否为播音员镜头。

在MPEG视频编码方案中，颜色信息被分解成亮度与色度不同的分量，由于人眼对色度变化不敏感，分解后便于压缩。心理实验指出人对颜色的感知有三个属性：色调，饱合度，强度。强度值对应于亮度值Y，而色调与饱合度则保存在色度分量Cb与Cr中，人的皮肤颜色形成与大多数自然物体明显不同的颜色范围。人的皮肤颜色虽然不同人，不同种族之间存在一定的差异，但它在色度平面上却分布在非常小的一个特定区域内。该事实已被许多研究者注意到，并作为一个重要组成部分应用到人脸的监测、定位、追踪系统中[2~9]，在专门面向人脸检测的应用中，通常利用该事实作为初步检测，主要是用来获得后面进一步确认的候选人脸对象。高度准确地检测定位人脸还需使用其它的证据线索，如人脸的长宽比、形状、边缘分布，甚至特征器官的检测等。这里，人脸检测仅作为一种辅助手段来优化基于背景特征区色彩模型的检测结果。因此，为了不至由于该技术的引入使系统的检测速度影响过大，仅利用肤色模型来对人脸进行最初步的简单检测，从而对虚假检测到的播音员镜头进行过滤。后面的实验将说明经过背景特征区色彩模型的检测后，下面描述的处理检测方法已经充分有效了。

如图1(b)所示，设播音员面部出现的区域为F，在MPEG流中每一个属于区域F的色度块的两个色度分量的直流值分别为dc_{mn}^{Cb}、dc_{mn}^{Cr}，其中m、n表示该色度块在矩形区域F中的位置索引。我们定义一个肤色检测函数：

$$Skin(m,n)=\begin{cases}1 & \text{若 } dc_{mn}^{Cb}\in[s\min Cb, s\max Cb] \\ & \quad \text{且 } dc_{mn}^{Cr}\in[s\min Cr, s\max Cr] \\ 0 & \text{否则}\end{cases} \tag{1}$$

其中$[s\min Cb, s\max Cb]$、$[s\min Cr, s\max Cr]$为我们通过对80个不同人脸的肤色进行采样建立的人脸肤色在Cb-Cr空间上的变化范围。令

$$Pface=\sum_{块_{(m,n)}\in 区域F} Skin(m,n)/BlockNum \tag{2}$$

其中$BlockNum$为区域F中包含的总色度块数。若

$$Pface>\mu \tag{3}$$

则判定在区域F中存在播音员的人脸，其中μ为预定义的门限参数，介于0与1之间。

2.2 播音员镜头检测中所用模型的建立

利用2.1节中描述的算法进行播音员镜头检测前，需要针对特定的新闻节目建立各类播音员镜头的特征模型。模型的建立可以通过半自动的交互方式来建立。下面给出播音员镜头特征模型的形式化描述。

定义1 每一类播音员镜头的特征模型是一个三元组$G=(L,D,F)$，其中L是块对齐的背景特征区，D是背景特征区中色度分量块DC系数值动态分布的统计模型，F是块对齐的播音员面部出现区域。

定义2 背景特征区中色度分量块DC系数值动态分布的统计模型是一个六元组$D=(rv^{Cb}, ravg^{Cb}, rsd^{Cb}, rv^{Cr}, ravg^{Cr}, rsd^{Cr})$，其中$rv^{Cb}$是Cb分量块DC系数的动态分布区间，$ravg^{Cb}$是背景特征区中Cb分量所有DC系数均值的分布区间，rsd^{Cb}是背景特征区中Cb分量所有DC系数标准方差的分布区间，rv^{Cr}是Cr分量块DC系数的动态分布区间，$ravg^{Cr}$是背景特征区中Cr分量所有DC系数均值的分布区间，rsd^{Cr}是背景特征区中Cr分量所有DC系数标准方差的分布区间。

块对齐的播音员面部出现区域可以通过人机交互与机器分析计算来建立。通过工具可以从码流中选取出属于播音员镜头的帧，并用鼠标可视化地框出播音员面部的位置，设它们的区域分别为矩形A_1，A_2，…，A_K，令

$$FA'=\bigcup_{i=1}^{K} A_i \tag{4}$$

设完全覆盖FA'的最小矩形为(fl,ft,fr,fb)，令FA''为矩形$(fl-\tau,ft-\upsilon,fr+\tau,fb+\upsilon)$，其中$\tau$、$\upsilon$为松弛因子。最后对区域$FA''$求取其内部的边界为MPEG编码的块边界的最大矩形，从而获得块对齐的播音员面部出现区域F。

块对齐的背景特征区首先通过人的主观评判确定出候选的特征区。主观评判的标准是特征区中颜色非

常相似，我们希望背景特征区中包含的色度分量的所有DC系数分布在一个尽可能狭窄的区间上。

针对Cb分量，设训练例中帧 $F_i(i=1,2,\cdots,M)$ 的候选背景特征区 CB 中包含的所有块的DC系数值分别为 $D_{ij}(j=1,2,\cdots,N)$，令 rv^{cb} 表示的区间为 $[v_l^{cb}, v_r^{cb}]$，则

$$v_l^{cb}=\min_i\{\min_j D_{ij}\}-\tau_v^{cb} \tag{5}$$

$$v_r^{cb}=\max_i\{\max_j D_{ij}\}+\tau_v^{cb} \tag{6}$$

其中 τ_v^{cb} 为对于区间 rv^{cb} 的松弛因子，用于在一定程度上防止训练例的不充分导致Cb分量块DC系数的动态分布区间不够宽。建议 τ_v^{cb} 的取值为3～7，在为CCTV新闻建立模型所取值为4。

令 $ravg^{cb}$ 表示的区间为 $[avg_l^{cb}, avg_r^{cb}]$，则

$$E_i=\frac{1}{N}\sum_{j=1}^{N}D_{ij} \tag{7}$$

$$avg_l^{cb}=\min_i\{E_i\}-\tau_{avg}^{cb} \tag{8}$$

$$avg_r^{cb}=\max_i\{E_i\}+\tau_{avg}^{cb} \tag{9}$$

其中 τ_{avg}^{cb} 为对于区间 $ravg^{cb}$ 的松弛因子。建议 τ_{avg}^{cb} 的取值为3～6，在为CCTV新闻建立模型所取值为3。

令 rsd^{cb} 表示的区间为 $[sd_l^{cb}, sd_r^{cb}]$，则

$$V_i=\sqrt{\frac{1}{|N|}\sum_{j=1}^{N}(D_{ij}-E_i)^2} \tag{10}$$

$$sd_l^{cb}=\min_i\{V_i\}-\tau_{sd}^{cb} \tag{11}$$

$$sd_r^{cb}=\max_i\{V_i\}+\tau_{sd}^{cb} \tag{12}$$

其中 τ_{sd}^{cb} 为对于区间 rsd^{cb} 的松弛因子，建议 τ_{sd}^{cb} 的取值为4～7，在为CCTV新闻建立模型所取值为4。

同理，可利用统计方法建立对于Cr分量的 rv^{cr}，$ravg^{cr}$，rsd^{cr} 的特征区间。最后按照 $[v_l^{cb}, v_r^{cb}]$ 与 $[v^{cr}, v_r^{cr}]$ 区间的大小，排列所有候选背景特征区，并基于该模型利用评估实验来选择出一个合适的背景特征区供以后长期使用。

3. 实验评价

为了验证2节中描述算法的有效性，我们从建立的视频节目数据库中抽取11天的CCTV新闻电视节目构成实验数据集对算法进行评价。其中3天的节目被用于建立播音员镜头背景区域中所选特征区的色彩模型及人脸的肤色模型，其余8天的新闻节目作测试集。整个实验都是在PⅡ-450、64M内存的机器上进行的。测试数据的帧率为24帧/秒，帧尺寸为720＊576，整个测试数据集的节目长度共约4个半小时，包含369925帧，在测试前我们对8天的新闻节目手工标注出了所有的播音员镜头，作为算法自动检测结果的标准参照。

表1给出了仅基于播音员镜头背景区域中所选特征区的色彩模型对测试集中播音员镜头检测的试验结果。利用表1的实验数据我们可以得到检测的正确率：$P=1-\frac{E}{D}=1-\frac{8}{96}=91.7\%$；查全率 $R=1-\frac{U}{S}=1-\frac{0}{88}=100\%$。检测结果具有100%的查全率是很理想的，因为对于用户来说通过人机交互从很少的若干探测结果镜头中选出几个非播音员镜头是一件容易、轻松的工作。对于该测试集，在实验中共遇到的8处虚假检测镜头如图2所示。由观察不难看出，在新闻视频中的确存在一些片断，它们在背景特征区中含有与播音员镜头相类似的蓝色调。

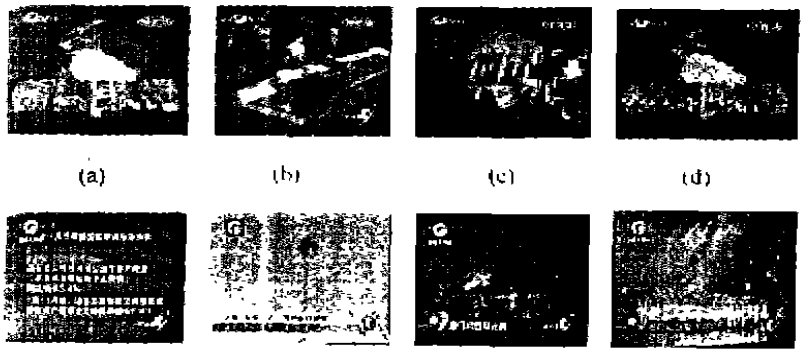

图2 基于背景中所选特征区的色彩模型出现的对播音员镜头的错误检测

在前面检测结果的基础上，我们利用人脸的肤色模型对检测结果作进一步的检测确认后，在保持100%查全率的前提下，有效地滤掉了前面图2除(c)外的7处错误检测，使检测正确率提高到 $P=1-\frac{E}{D}=1-\frac{1}{89}=98.9\%$。实验中，我们取 $\mu=0.17$。所获得的实验结果对算法的有效性提供了有力的证明。

表1 对测试集中各视频流所含播音员镜头进行检测的试验结果

视频流	帧数	实际播音员镜头数目(S)	检测出的播音员镜头数目(D)	检测错误的播音员镜头数目(E)	未检测到的播音员镜头数目(U)
News0	44966	9	9	0	0
News1	44770	11	15	4	0
News2	45106	15	17	2	0
News3	44913	13	15	2	0
News4	36339	13	13	0	0
News5	59610	12	12	0	0
News6	44447	10	10	0	0
News7	49775	5	5	0	0
总计	369925	88	96	8	0

另外，我们将综合两种模型检测播音员镜头的算法对测试集中各视频流的检测时间进行了统计，列于表2中，求反映算法的计算效率。算法的实现采用的是对测试集中的 MPEG-2码流具有8帧/秒全解码能力的解码引擎。由表2可知，我们实现的检测系统具有平均77.55帧/秒的超实时检测速度，这种非常快速的检测速度除了来源于算法本身涉及的运算类型较简单外，还由于计算是在压缩域中进行的，避免了完全解码涉及的运算复杂度很高的反 DCT 计算，并且采用了粗细两种粒度进行分阶段检测，在粗粒度状态下，大量的非 I 帧被跳过。

表2 对测试集各视频流检测播音员镜头的检测时间

视频流	News0	News1	News2	News3	News4	News5	News6	News7
检测时间(分:秒)	9:40	8:44	9:20	9:39	8:28	12:36	9:22	11:41

表3 对测试集中江西卫视新闻视频流的播音员镜头检测的实验结果

视频流	帧数	实际播音员镜头数目(S)	检测出的播音员镜头数目(D)	检测错误的播音员镜头数目(E)	未检测到的播音员镜头数目(U)
JXTV0210	29690	9	9	0	0
JXTV0212	29464	9	9	0	0
JXTV0213	25930	8	8	0	0
JXTV0214	29500	8	8	0	0
总计	114574	34	34	0	0

为了对算法的有效性、强壮性进行更进一步的评估，系统采用4天的江西卫视新闻对算法进行第二组评估测试，主要考察算法对于不同电视新闻节目的适用性。测试数据的帧率为24帧/秒，帧尺寸为720 * 576，共约80分钟，包含114574帧。表3列出了有关的实验结果。计算所得的探测查全率与正确率为：

$$\text{查全率 } R=\frac{\text{系统输出正确的标题数目}}{\text{实际包含的标题数目}}=1-\frac{U}{S}=1-\frac{0}{34}=100\%$$

$$\text{正确率 } P=\frac{\text{系统输出正确的标题数目}}{\text{系统输出的标题数目}}=1-\frac{E}{D}=1-\frac{0}{34}=100\%$$

可见，播音员镜头检测系统对江西卫视新闻取得了异常高的查全率与正确率，表明算法对于不同的电视新闻节目同样是有效、强壮的。

结束语 本文提出了一种基于背景色彩及人脸肤色模型的播音员镜头检测方法。我们对算法的评估实验不仅给出了有关检测准确性及完整性的实验结果，也报告了有关反映算法计算费用的检测时间信息。实验证明它是一种十分快速有效的播音员镜头检测算法。本算法不仅适用于 CCTV 新闻、江西卫视新闻，同样适用于其他满足2节中假设的新闻节目，只是需要重新选择符合假设条件的背景特征区并通过统计建立新的背景特征区色彩模型，同时重新定义播音员面部出现的区域 F，这些可以通过一个交互式工具来实现。

播音员镜头是新闻视频中一种重要事件，准而全地检测该种事件对实现新闻视频自动新闻条目分割及解析有重要意义。

参 考 文 献

1 Zhang H J, et al. Automatic Parsing and Indexing of News Video. Multimedia Systems, 1995, 2: 256～266

2 Low C Y, Tian Q, Zhang H J. An Automatic News Video Parsing, Indexing and Browsing System. In: Proc. ACM Multimedia 96, Boston, MA, Nov. 1996. 425～426

3 Merlino A, Morey D, Maybury M. Broadcast News Navigation Using Story Segmentation. In: Proc. ACM Multimedia 97, Seattle, USA, Nov. 1997. 381～391

4 Qi W, et al. Integrating Visual, Audio and Text Analysis for News Video. IEEE ICIP-2000. Vancouver, Canada, Sep. 2000

5 Yeo B L, Liu B. Rapid Scene Analysis on Compressed Videos. IEEE Transactions on Circuits and Systems for Video Technology, 1995, 5(6): 533～544

6 Song J, Yeo B L. Spatially Reduced Image Extraction from MPEG-2 Video: Fast Algorithms and Applications. Storage and Retrieval for Image and Video Database VI, 1998, SPIE3321(Jan.)

7 Wang H, Chang S.-F. A Highly Efficient System for Automatic Face Region Detection in MPEG Video. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Multimedia Technology, Systems, and Applications, 1997, 7(4)

8 Sobottka K, Pitas I. A Novel Method for Automatic Face Segmentation, Facial Feature Extraction and Tracking. Signal Processing: Image Communication, 1998, 12(3): 263～281

9 Zhang H M, et al. Combining Skin Color Modal and Neural Network for Rotation Invariant Face Detection. Int. Conf. Multimodal Interface 2000, Beijing, 2000