新闻视频中口播帧检测方法的研究*

马宇飞, 白雪生, 徐光祐, 史元春

(清华大学 计算机科学与技术系,北京 100084)

E-mail: mayufei@263.net

http://www.tsinghua.edu.cn

摘要: 新闻视频分析是视频分析领域的重要课题.提出了一种基于知识的新闻视频分析方法——二阶段模板匹配法,用于检测新闻节目中主持人口播镜头,从而为新闻单元的定位提供基本依据.该方法具有通用性和实时性的特点,可以在新闻视频的自动分析或自动索引系统中得到实际应用.

关键词: 视频分析;模板匹配;新闻视频索引;广义 Hough 变换;彩色直方图交运算

中图法分类号: TP391 **文献标识码**: A

随着计算机技术和网络技术的发展,利用计算机自动分析和管理视频数据成为研究人员关注的热点.视频分析的目的是恢复视频信息中的语义结构,并根据这个结构进行管理或建立索引.新闻视频的结构特征比较明显,其主体内容是一系列新闻单元.准确地定位每个新闻单元的开始和结束位置,是新闻视频自动索引的重要依据,同时也是新闻视频分析的基本问题之一.通过对新闻单元的结构进行分析可以知道,新闻主持人口播镜头的开始通常就是一个新闻单元的开始,因此,检测新闻主持人的口播帧是定位新闻单元的有效途径.

目前,研究人员在新闻视频的分析方面已经提出了很多方法[1~6],并试图建立一个完全自动的新闻视频分析和管理系统,用于浏览和检索视频片段.在新闻视频分析中,检测主持人口播镜头始终是一个重要方面[1,2,4~6].文献[5]中提出一种基于模板匹配的方法,该方法需要多次扫描视频序列,而且相似度测量的算法复杂度较高,因此实时性并不理想.文献[6]提出了一种综合视频、语音和同步文本的方法来分析新闻视频,对于主持人镜头的检测主要采用语音信息进行定位,最后再利用同步文本信息辅助定位,但该方法的局限性较大.文献[4]通过分析人脸的肤色特征,构造标准模板来进行模板匹配,但这种基于皮肤颜色的方法对人的肤色和演播室的光照比较敏感,而且该方法只确定了 5 种标准模板,很难适应所有新闻节目,其通用性受到很大限制.

本文在研究数字视频管理和检索技术的基础上对新闻视频作了深入的研究,提出一种基于知识的新闻视频分析方法——二阶段模板匹配法,以检测新闻节目中主持人的口播镜头.本文第 1 节详细描述二阶段模板匹配法的原理.第 2 节给出所采用的两种模板匹配算法.第 3 节是应用二阶段模板匹配法检测的实验结果.第 4 节为结论.

* **收稿日期**: 1999-10-19; **修改日期**: 1999-12-29

基金项目: 国家 863 高技术研究发展计划资助项目(863-306-ZT04-02-01);国防科技预研基金资助项目

作者简介: **马宇飞**(1972—),男,山西人,硕士生,主要研究领域为基于内容检索,计算机视觉;**白雪生**(1972),男,辽宁人,博士,讲师,主要研究领域为基于内容检索,计算机视觉;**徐光祐**(1940),男,上海人,教授,博士生导师,主要研究领域为计算机视觉,多媒体技术;**史元春**(1967—),女,河南人,博士,副教授,主要研究领域为计算机支持的协同工作,多媒体技术.

1 二阶段模板匹配法

1.1 基本原理

由于不同新闻节目的口播帧各不相同,为了使检测算法具有广泛的通用性,需要构造一个通用模板.这个通用模板不应包含演播室背景因素以及主持人的个性因素,如:服饰、肤色、位置及性别等.在所有主持人口播帧中惟一的共性是标准半身像,因此,可以提取半身像的边缘曲线作为通用模板.对于曲线的匹配可以采用第2.1节中描述的推广的广义Hough变换(general Hough transform,简称GHT)的方法.

尽管利用通用模板可以检测出所有的主持人口播帧,但基于GHT的曲线匹配通常较为耗时.为了提高检测速度,我们提出一种二阶段模板匹配的方法.如图1所示,检测过程可分为两个匹配阶段,在第1阶段中采用通用模板进行推广的GHT匹配,当检测到第1个口播帧时,将其作为当前视频序列的专用模板,并进入第2阶段.在第2阶段中采用专用模板进行基于彩色直方图交运算的匹配(color histogram intersection,简称CHI),由于CHI的算法复杂性远低于GHT,从而大大减少了匹配时间,提高了算法的实时性.

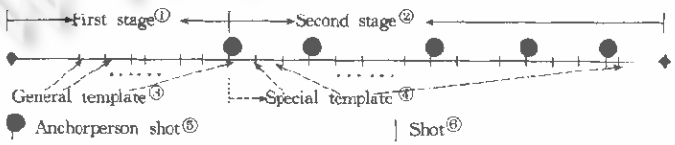


Fig.1 Two-Stage template matching method
图1 二阶段模板匹配法

以上描述的二阶段模板匹配法,既可以在镜头分割的基础上单独使用,也可以将其集成到镜头分割的算法中,即在一次扫描视频序列的过程中完成镜头的分割和口播帧的定位,而系统的处理时间并不会增加很多.

1.2 通用模板和专用模板

通用模板作为先验知识集成在新闻视频分析系统中,与具体被检测视频序列无关,因而在每次检测时不需要更改.专用模板与当前被检测视频序列相关,每次检测时需要通过通用模板的匹配找到当前有效的专用模板.构造通用模板需要以下步骤:首先,任意截取两个标准的主持人口播帧灰度图像,一个是单人正面半身像,另一个是双人正面半身像.然后,对图像进行边缘检测,并滤除噪声和小边缘得到比较干净的半身像边缘,如图2中(a)和(b)所示.接下来,按照第2.1节描述的GHT方法,在半身像曲线边缘的采样点上提取一组参数构成R数组,输出到文件中作为通用模板文件.由于第2.2节中定义的专用模板匹配方法具有区域选择性,在具体应用时,可对口播帧上半部分中演播室背景和主持人脸所在区域设置平均权重,而令其他部分区域的权重为零.因此专用模板可以不区分单人和双人,在进行匹配时,两个通用模板中任意一个得到匹配,即可将匹配的帧作为当前视频序列的专用模板.

2 模板匹配算法

2.1 基于推广的广义Hough变换的匹配方法

本文利用广义Hough变换的思想,设计了一种匹配主持人半身像边缘曲线的方法.首先,利用

DOG(difference of Gaussian)算子对主持人口播帧的灰度图像进行边缘检测,由于DOG算子不能直接得到边缘点的方向信息,所以在Hough变换的过程中避免了使用方向信息.如图3所示,设曲线L是待检测的曲线,任意选择的一个参考点为(x_p,y_p),曲线上任意一点为(x,y),则该点到参考点之间的连线R的长度为r,方向为α,于是,参考点和曲线上任意一点之间的关系如式(1)和式(2)所示.

$$x_p=x+r\cos(\alpha), \tag{1}$$

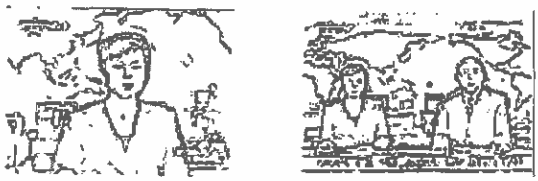
$$y_p=y+r\sin(\alpha), \tag{2}$$

$$r=\sqrt{(x-x_p)^2+(y-y_p)^2}, \tag{3}$$

$$\sin(\alpha)=\frac{y_p-y}{r}, \tag{4}$$

$$\cos(\alpha)=\frac{x_p-x}{r}. \tag{5}$$

在通用模板的边缘图中,对于指定曲线边缘上的每个采样点都可以通过式(3)~(5)得到一组参数$R:\{r,\sin(\alpha),\cos(\alpha)\}$,构成$R$数组.当进行匹配时,对被匹配图像边缘图上的任意一个边缘点(x,y),利用式(1)、式(2)以及R数组参数恢复出参考点的位置,通过参考点位置累加器的累计结果来判断被匹配图像中存在的指定曲线,如果累计结果存在局部极大值,表明图像中存在指定的曲线,否则表明图像中不存在指定的曲线.



(a) General template-1 (a) General template-1
(a) 通用模板 1 (a) 通用模板 1

Fig. 2 General template
图2 通用模板

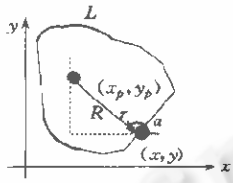


Fig. 3 Generalized GHT
图3 推广的GHT

可以看出,以上推广的GHT匹配方法,只有当被匹配图像中出现与指定曲线具有相同方向、相同尺度以及相同形状的曲线时才能得到匹配,而指定曲线在平面中的位置可以是任意的.这些特点完全满足对标准半身人像边缘曲线的匹配,而且主持人在镜头中的位置不会影响匹配结果.在实际应用中,由于主持人所处的垂直位置变化很小,可以通过适当限制垂直方向的偏移来提高匹配速度.

2.2 基于彩色直方图交运算的匹配方法

彩色直方图交运算(color histogram intersection)如式(6)所示,首先取两个直方图中相应bin中像素数的最小值的和,然后求它与总像素数的比值.这个比值在0和1之间,越是接近1表明两副图像越相似.

$$\mathrm{Sim}(A,B)=\frac{\sum_{i=1}^{n}\min(A_i(Y,U,V),B_i(Y,U,V))}{\sum_{i=1}^{n}A_i(Y,U,V)}, \tag{6}$$

其中$\mathrm{Sim}(A,B)\in[0,1]$表示相似度,$A$和$B$分别为两幅图像的彩色直方图,$n$为彩色直方图bin的

个数,而 $A_i(Y,U,V)$和 $B_i(Y,U,V)$则分别是直方图中第 i 个 bin 中的像素数.

当采用式(6)进行相似度的比较时,对图像中物体的运动以及分布不敏感,而对光照以及物体的颜色变化较为敏感,但这些特点不适于进行口播帧的模板匹配.于是将式(6)改写成式(7),即可将图像任意划分成多个矩形区域,然后给每个区域赋予一个[0,1]的权重,对于不关心的区域,赋予其权重为 0,对于关心的区域,可根据重要性赋予(0,1]之间的值.式(7)中,r 表示将图像划分成的区域数,w_j 则是第 j 区域的权重且 $\sum_{j=1}^{r} w_j=1$,其他参数和式(6)相同.实验表明,式(7)作为专用模板的匹配方法,可以得到很好的性能.

$$\mathrm{Sim}(A,B)=\sum_{j=1}^{r} w_j \frac{\sum_{i=1}^{b}\min(A_i(Y,U,V),B_i(Y,U,V))}{\sum_{i=1}^{b} A_i(Y,U,V)}. \tag{7}$$

3 实验结果

根据本文提出的算法,我们在 Windows NT 平台上实现了一个新闻视频的分析系统,并在 Pentium II 266 PC 机上对电视台的新闻节目进行测试,视频序列中包括新闻片头、新闻摘要、新闻单元、天气预报和广告等内容,其中新闻主持人的口播镜头共有 23 个.两个通用模板如图 2 中(a)和(b)所示,作为先验知识集成在分析系统中.实验分为两部分,实验 1 是将二阶段模板匹配法集成到镜头分割算法中,分别检测 4 段新闻的镜头和口播镜头,测试该方法的通用性和准确率.记主持人的口播镜头总数为 T_{ap},误检的镜头数为 F_{ap},漏检的镜头数为 M_{ap},则准确率 P 由式(8)得到.实验 2,在比较镜头分割算法中加入二阶段模板匹配法后,通过系统处理时间增加的比率,验证了二阶段模板匹配法的实时性和实用性.

$$P=\frac{T_{ap}-(F_{ap}+M_{ap})}{T_{ap}}\times 100\%. \tag{8}$$

实验 1 的结果见表 1,尽管单人半身像的通用模板是从某半身像中提取出来的,但对于其他电视台的男女主持人同样适用.在大约 35 分钟的新闻视频中,二阶段模板匹配法得到总的准确度为 91.3%,没有误检的情况,漏检的两个主持人口播镜头都是由于主持人镜头的切换采用 *dissolve* 方式渐变.如果将算法稍加修改,选择每个镜头中的中间帧而不是第 1 帧作为匹配对象,将可避免这种漏检.实验 2 的结果见表 2,表中分别列出每个视频序列的原始长度,只采用镜头分割算法时的系统处理时间、集成了二阶段模板匹配法后的系统处理时间以及时间增加的百分比.可以看出,时间的增加通常在 3.0%左右,个别序列增加的时间较大,这是因为新闻主持人口播帧出现较晚,造成第 1 阶段的匹配时间较长.这种情况可能由于以下两个原因造成:(1) 视频序列是从一个较长的新闻单元的中间开始;(2) 新闻开始前插播较长的广告.第 1 种情况在实际应用中不应出现,而第 2

Table 1

表 1

News video file①	Double②	Single③	Total④	False⑤	Miss⑥	Precision⑦(%)
news1.mlv	1	6	7	0	0	100
news2.mlv	1	4	5	0	0	100
news3.mlv	0	7	7	0	1	85.7
news4.mlv	0	4	4	0	1	75
Total	2	21	23	0	2	91.3

①新闻视频文件,②双人,③单人,④总数,⑤误检,⑥漏检,⑦准确率.

Table 2
表 2

News video file①	Length②(s)	Time of segment③(s)	Time of segment and TTM④(s)	Increase of time⑤(%)
news1.mlv	900	248	256	3.2
news2.mlv	600	165	176	6.7
news3.mlv	270	78	80	2.7
news4.mlv	300	83	85	2.4
Total⑥	2070	574	597	4.0

①新闻视频文件,②视频长度,③镜头分割时间,④镜头分割+二阶段法,⑤时间增加,⑥合计

种情况可以通过准确定位广告或新闻的开始片头,使问题得到解决.

4 结 论

本文通过分析不同电视台新闻节目的特点,提出一种基于知识的二阶段模板匹配法,用于新闻视频分析中主持人口播帧的检测.该方法通过提取口播帧中抽象的共同特征作为通用模板,使通用性得到很大改善.实验表明,在镜头检测算法中集成本文提出的二阶段模板匹配法,尽管增加了大约3.0%的镜头检测时间,但在一次扫描视频序列的过程中不仅可以完成镜头分割,而且能够准确定位主持人的口播帧,准确率可以达到90%以上,为新闻单元的定位提供基本的依据.因此,本文提出的检测方法在通用性、实时性和准确率方面均达到了实用化水平.

有效地定位主持人的口播镜头是分割新闻单元的重要依据,但有时并不能解决所有问题,例如,有的新闻单元开始并不出现主持人,而在有的新闻单元中出现多次,这时就需要借助其他特征来分割新闻单元.此外,在新闻单元之间有时插有新闻片头或广告,为了使用户在检索新闻时避免得到这些冗余信息,还需要精确定位新闻片头和广告等片段,这将是我们下一步的工作.

References:

[1] Swanberg, D., Shu, C.F., Jian, R. Knowledge guided parsing in video databases. In: Nibblack, W., ed. IS & T/SPIE. San Jose, CA: SPIE, 1993. 13~24.

[2] Ariki, Y., Saito, Y. Extraction of TV news articles based on scene cut detection using DCT clustering. In: Delogne, P., ed. Proceedings of the International Conference in Image Processing. Lausanne, Switzerland: IEEE Computer Society Press, 1996. 3,847~850.

[3] Chen, L., Faudemay, P. Multi-Criteria video segmentation for TV news. In: Wang, Y., Reibman, A.R., Juang, B.H., *et al.*, eds. Proceddings of the 1st IEEE Workshop on Multimedia Signal Processing. New York: IEEE, 1997. 319~324.

[4] Gunsel, B., Ferman, A.M., Tekalp, A.M. Video indexing through integration of syntactic and semantic features. In: IEEE Computer Society ed. Proceedings of the IEEE Workshop on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 1996. 90~95.

[5] Hanjalic, A., Lagendijk, R.L., Biemond, J. Semi-Automatic news analysis, indexing and classification system based on topics preselection. In: Yeung, M.M., Yeo, B.L., Bouman, C.A. eds. SPIE. San Jose, CA: SPIE, 1999. 3656:86~97.

[6] Huang, Q., Liu, Z., Rosenberg, A. Automated semantic structure reconstruction and representation generation for broadcast news. In: Yeung, M.M., Yeo, B.L., Bouman, C.A. eds. SPIE. San Jose, CA: SPIE, 1999. 3656:50~62.

Research on Anchorperson Detection Method in News Video*

MA Yu-fei, BAI Xue-sheng, XU Guang-you, SHI Yuan-chun

(*Department of Computer Science and Technology, Tsinghua University, Beijing* 100084, *China*)

E-mail: mayufei@263.net

http://www.tsinghua.edu.cn

Abstract: News video analysis is an important field of video analysis. In this paper a knowledge-based analysis method for news video——a two-stage template matching method-to detect the anchorperson shot of news video is presented, which is a fundamental step for segmenting news video units. The method is both generic and real-time, and can be used in automatic news video analyzing and indexing systems.

Key words: video analysis; template matching; news video indexing; general Hough transform; color histogram intersection

* Received October 19, 1999; accepted December 29, 1999

Supported by the National High Technology Development Program of China under Grant No. 863-306-ZT04-02-01; the National Defence Research Foundation