

一种基于形态学操作的新闻标题条检测算法¹

姜帆 章毓晋

(清华大学电子工程系 北京 100084)

摘要: 该文提出一种基于形态学操作的新闻标题条检测算法。该方法利用了新闻标题条出现的时空特征,在镜头分割的基础上,采取了分类处理的方式和形态学的滤波操作,实现了对渐变标题的检测和在全局变化剧烈情况下的标题检测,并具有定位准确、规范化的特点。利用 MPEG-7 测试数据集进行的实验验证了该算法的优越性。

关键词: 新闻视频检索,新闻标题条,形态学操作,渐变

中图分类号: TP391 **文献标识码:** A **文章编号:** 1009-5896(2003)12-1647-06

A Caption Detection Algorithm Based on Morphological Operation

Jiang Fan Zhang Yu-jin

(Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract An algorithm of caption detection in news video is proposed in this paper. It makes use of space and time characters of the captions appearing in the video. On the basis of shot boundary detection, this method applies morphological operation and different filter strategies according to different cases. It achieves good results by experiments using MPEG-7 video testing data, especially caption detection in the case of gradual transition and heavy noises on the background.

Key words News video retrieval, News caption, Morphological operation, Gradual transition

1 引言

随着多媒体技术的发展和视频数据库的普及,如何进行视频流的分析和检索,从而在海量的视频数据中查找用户感兴趣的内容,已成为研究人员关注的热点。与一般的视频数据相比,新闻视频节目具有比较层次化的结构模式,这提供了多种有助于理解视频内容、建立索引结构和进行基于内容查询的信息线索。

在新闻视频分析中,字符识别是一项重要的技术,它指的是自动分析和提取节目中出现的文字,这有助于提高对新闻视频内容的整体理解程度。视频中出现的文字可分为两种类型:一种是和背景融合在一起的字符,是原始画面的一部分,如商店招牌、汽车牌号等;另一种是在节目制作中加入的标题、字幕等覆盖于画面的字符。第二种是对内容的注释,含有重要的语义信息^[1,2]。

已经有一些论文讨论了基于文字特征的标题检测方法,文献 [3] 中考虑一个典型的标题条区域与背景之间有很陡峭的边缘,利用这一显著差异来检测标题;文献 [4] 是从视频中抽取一些图像帧,分成 16×16 像素的子区域,用小波神经网络的方法来判别字符区域;文献 [5] 首先利用一个 3×3 的水平差分过滤器检测垂直边界,然后用平滑滤波去掉一些非文字的多余碎片;文献 [6] 还提出了直接利用压缩域的数据信息来进行识别的方法。

¹ 2002-07-11 收到, 2003-03-12 改回

国家自然科学基金项目 (60172025) 资助课题

一种基于形态学操作的新闻标题条检测算法¹

姜 帆 章毓晋

(清华大学电子工程系 北京 100084)

摘 要: 该文提出一种基于形态学操作的新闻标题检测算法。该方法利用了新闻标题条出现的时空特征,在镜头分割的基础上,采取了分类处理的方式和形态学的滤波操作,实现了对渐变标题的检测和在全局变化剧烈情况下的标题检测,并具有定位准确、规范化的特点。利用 MPEG-7 测试数据集进行的实验验证了该算法的优越性。

关键词: 新闻视频检索,新闻标题条,形态学操作,渐变

中图分类号: TP391 **文献标识码:** A **文章编号:** 1009-5896(2003)12-1647-06

A Caption Detection Algorithm Based on Morphological Operation

Jiang Fan Zhang Yu-jin

(Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract An algorithm of caption detection in news video is proposed in this paper. It makes use of space and time characters of the captions appearing in the video. On the basis of shot boundary detection, this method applies morphological operation and different filter strategies according to different cases. It achieves good results by experiments using MPEG-7 video testing data, especially caption detection in the case of gradual transition and heavy noises on the background.

Key words News video retrieval, News caption, Morphological operation, Gradual transition

1 引言

随着多媒体技术的发展和视频数据库的普及,如何进行视频流的分析和检索,从而在海量的视频数据中查找用户感兴趣的内容,已成为研究人员关注的热点。与一般的视频数据相比,新闻视频节目具有比较层次化的结构模式,这提供了多种有助于理解视频内容、建立索引结构和进行基于内容查询的信息线索。

在新闻视频分析中,字符识别是一项重要的技术,它指的是自动分析和提取节目中出现的文字,这有助于提高对新闻视频内容的整体理解程度。视频中出现的文字可分为两种类型:一种是和背景融合在一起的字符,是原始画面的一部分,如商店招牌、汽车牌号等;另一种是在节目制作中加入的标题、字幕等覆盖于画面的字符。第二种是对内容的注释,含有重要的语义信息^[1,2]。

已经有一些论文讨论了基于文字特征的标题检测方法,文献 [3] 中考虑一个典型的标题条区域与背景之间有很陡峭的边缘,利用这一显著差异来检测标题;文献 [4] 是从视频中抽取一些图像帧,分成 16×16 像素的子区域,用小波神经网络的方法来判别字符区域;文献 [5] 首先利用一个 3×3 的水平差分过滤器检测垂直边界,然后用平滑滤波去掉一些非文字的多余碎片;文献 [6] 还提出了直接利用压缩域的数据信息来进行识别的方法。

¹ 2002-07-11 收到, 2003-03-12 改回

国家自然科学基金项目 (60172025) 资助课题

2.2.3 分类处理 首先, 引入一个描述帧间全局变化剧烈程度的算子 δ , 定义如下:

$$\delta(Q_i) = \text{帧 } Q_i \text{ 中值为非零的像素点的个数} / \text{单帧像素点总数}$$

其表达的意义是: 帧间差图中颜色和亮度变化比较大的点与单帧总像素点的个数的比例。设定一个临界阈值 P , 分别考虑下面两种情况:

(1) $\delta(Q_i) < P$, 对应全局变化较小的情况, 先对画面下 $1/4$ 区域进行中值滤波, 令该子区域为预检区域 A_i 。这样做的目的是去掉一些非标题条的噪声碎片, 并且填补由于阈值处理而在标题区域形成的黑色空洞, 使得待检测的标题区域成为一个完整的块状;

(2) $\delta(Q_i) > P$, 对应全局变化较大的情况, 由于画面剧烈运动造成的噪声已使标题条成为一个整体, 可直接令画面下 $1/4$ 区域为预检区域 A_i 。

2.2.4 标题预检 对于从每个镜头中获得的预检区域 A_i , 设定一个临界阈值 P_a , 分别考虑下面两种情况:

(1) $\delta(A_i) < P_a$, 即帧 Q_i 中没有标题条出现, 继续检查下一帧;

(2) $\delta(A_i) > P_a$, 则对 A_i 进行腐蚀操作, 选取矩形形状的腐蚀模板, 用以滤出其他不规则碎片。如果操作后 δ 值为 0, 即不存在符合特征的标题条, 直接跳至下一帧; 如果操作后 δ 值不为 0, 则表明可能有标题条出现, 此时记录该子区域 A_i 。

经过上述 4 个步骤, 可在一个镜头中找到若干个可能为标题条出现的时间位置。

2.2.5 标题条出现时间定位 对于搜索到可能有标题条出现的若干子区域 A_i , 再用相同模板进行一次膨胀操作, 结果记为 B_i (图 1(d))。比较同一镜头各 B_i 中矩形区域面积的大小, 将面积最大的矩形块 (对应标题条的最明显变化) 所处的差图 Q_i 所对应的原始视频帧 S_i 作为标题条出现的帧 I (图 1(e))。

2.2.6 标题条的空间定位 将确定为标题条出现的 B_i 区域中各像素点分别沿 x, y 轴方向投影, 获得标题条所在的矩形区域位置。并将该位置坐标映射到相应原始帧 I 中。

2.2.7 对标题文字进行识别 (识别结果见图 1(f))

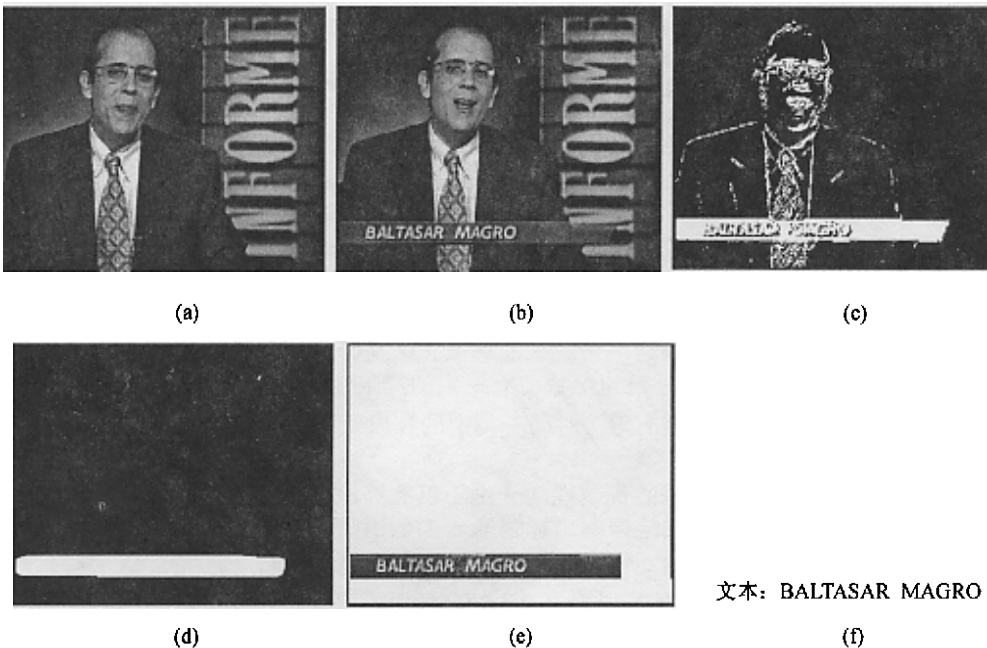


图 1 标题条检测算法示意图

可以看出,该标题检测算法仍然利用了新闻标题条出现对帧间画面变化的影响,采用求帧间差的方法进行判断。但是由于适当增加了扫描帧间的间隔 L ,使该方法不仅能检测标题突变,也能有效检测标题的渐变。通过分析可知,对于渐变越缓慢的标题条, L 应取值越大,但相应的实验时间定位准确度降低,而对于突变的标题条, L 应尽量小以提高定位精度。参数 L 将在后面的实验中采用样本训练的方法来产生。同时,算法中引入了一个描述视频变化剧烈程度的算子 δ ,对不同的镜头全局变化情况进行分类处理,能够减少因内容变化造成的标题误检。另外,在标题条的定位过程中,没有采用传统的 Sobel 算子进行边缘提取,而是利用形态学的腐蚀、膨胀等操作,更加准确地搜索判断标题条位置。所以,该方法的抗干扰和自适应的能力较强,具有一定的鲁棒性。这一点可以在检测实验结果中看出。

3 实验结果和讨论

为检验本文提出的基于形态学操作的标题条检测算法的检测效果,用 MPEG-7 测试数据集第 15 张和第 17 张光盘的 3 段有代表性的新闻片段(片段 1 和片段 2 选自文件 News1.mpg,片段 3 选自文件 Jornaldanoite2.mpg)进行了实验,其中片段 1 作为训练样本,其余两个片段作为测试数据。

下面先列出标题检测实验的整体统计数据结果,然后对该方法相对原有标题检测算法的两点改进作了具体实验和分析讨论。

3.1 新闻标题条检测统计结果

为了反映算法效率,我们使用通常的两个衡量标准:

(1) 查全率 (Recall) = 正确检测数 / (正确检测数 + 漏检数)

(2) 准确率 (Precision) = 正确检测数 / (正确检测数 + 误检数)

显然,算法中参数的选取对标题条检测效果至关重要,主要包括扫描帧间的间隔 L 、确定分类处理策略的阈值 P 和控制标题出现的最小临界比例值 P_a 。文献 [6] 探讨了参数选取问题,思路是变动一个参数,固定其它参数,记录漏检、误检的变化,最终选取一组折衷的参数。实验中采用类似的方法先利用训练样本确定参数的范围,得出 $L=3$, $P=1/3$, $P_a=1/20$ 。

表 1 是对测试片段(片段 2 和片段 3)的新闻标题条检测的结果。根据表中统计的数据,所得到的平均查全率约为 0.95,而平均准确率约为 0.87。可见该算法能够比较有效地检测新闻视频中标题条的出现。

表 1 新闻标题条检测结果

新闻视频段	总帧数	镜头数	正确检测数	漏检数	误检数	查全率	准确率
片段 2(News1.mpg 节选)	16,625	61	21	1	3	0.954	0.875
片段 3(Jornaldanoite2.mpg 节选)	23,450	131	32	2	5	0.941	0.865

由该实验结果可以看出,本文算法有很高的查全率,但也有较明显的误检率。通过进一步考察,大部分误检发生在同一镜头中内容变化较大、运动较剧烈的情况。由于在本文的算法中增加了扫描帧间的间隔,帧间颜色、亮度之差受到全局画面变化的影响也增大。在这种情况下,算法分类处理以及形态学操作难以确定一个统一的标准,这是影响该算法检测效果的主要原因。需要指出的是,采用文献 [6] 的方法确定参数值有一定的局限性,首先对不同类型的片段参数范围会略有不同,其次由于观测数据有限,对同一类型的片段该范围的合理性也值得商榷。

3.2 新闻标题条渐变的检测

本文介绍的新闻标题条检测算法的优越性之一是不仅能够检测在相邻帧间的区域亮度及颜色有明显跳变的标题条,还能较准确地检测出标题渐变(淡出渐隐)的近似位置。

这里限于篇幅,仅给出标题渐变检测的一个例子予以说明。如图 2 所示,第 1 行是截取自某新闻片段中连续的 4 帧,从第 k 帧到第 $k+3$ 帧,画面下部有一条蓝色新闻标题逐渐缓慢出现。第 2 行的前 3 幅图是两两相邻的各帧分别相减并经阈值处理所得的差图,第 4 幅图是间隔为 3 的两帧之差图。显而易见,在标题淡出的过程中,相邻帧间的标题区域变化较小,通过一般方法只能检测到零散的标题碎片。而如果增大检测帧的间隔($L=3$),便容易得到标题条的完

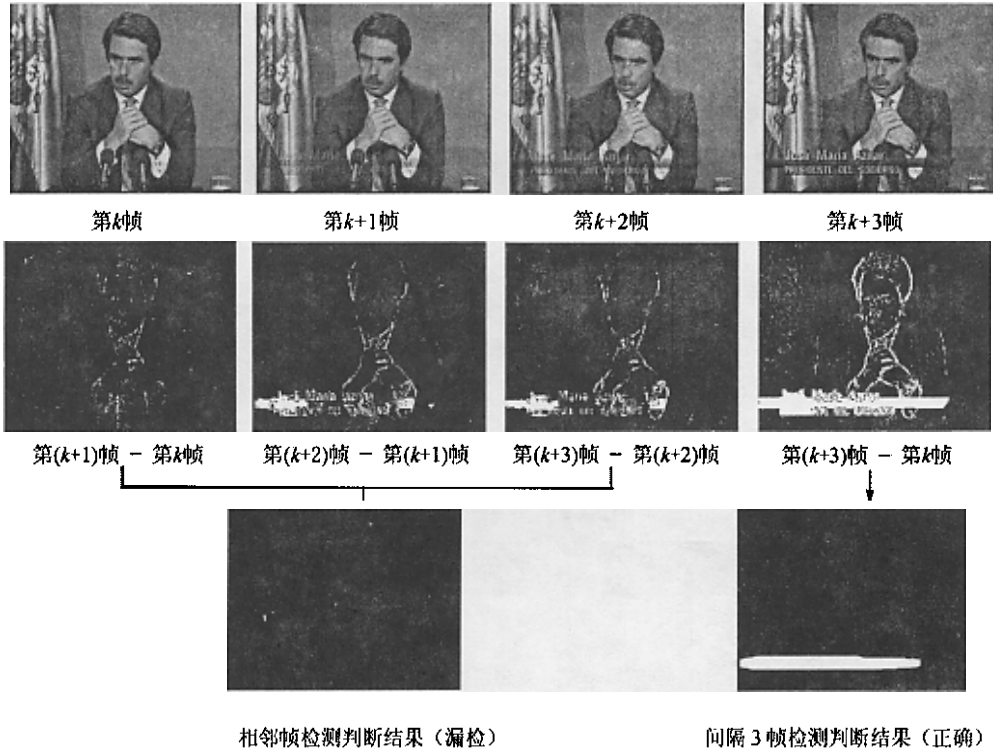


图 2 渐变标题条的检测

整形, 再通过相应的腐蚀和膨胀操作就能进行有效的标题条定位 (见图 2 的第 3 行检测判断结果)。

值得指出, 在所选择的测试视频片段中, 有一半以上的标题条是在播放过程中逐渐显现的。正是由于该算法能够有效检测标题渐变, 避免了类似的漏检现象, 才保证了检测结果有较理想的查全率。

3.3 视频画面剧烈变化情况下的新闻标题条检测

如何在视频画面剧烈变化的情况下, 检测到新闻标题条的出现是目前标题检测的技术难点。现有算法中, 一部分回避了这种情况, 将全局变化较大的帧视为镜头的切换点, 并不进行相应的标题检测判断; 另一部分只是简单地判断在帧画面下 1/3 或 1/4 的变化超过一定阈值就视为新闻标题条的出现, 那么一旦全局画面剧烈变化就总是认为检测到了标题, 这必然会造成严重的误检现象。这里选用的测试视频片段中, 有许多时候是画面强烈变化的同时又出现了新闻标题条 (例如图 3(a), 3(b))。本文提出的算法采用了形态学操作, 能够有效去除一些非标题条的干扰碎片, 因此取得了相对较好的检测效果。

这里举一例来说明。如图 3 所示, 图 3(a), 3(b) 分别是截取自原始视频文件的间隔为 3 的两帧 (有新闻标题条出现), 3(c) 图为 3(a), 3(b) 两帧的差图。可以看到, 由于画面中人物与背景有剧烈的运动, 所求得的差图内容比较杂乱, 标题的块状区域被淹没而不可见。但经过该算法中矩形模板的腐蚀和膨胀操作后, 却可以得到一个比较明显的标题位置 (图 3(d))。这种方法利用了标题条比较固定的形状特征和空间分布, 有较好的抗干扰能力, 所以对于背景变化较大的标题区域也有一定的识别能力。

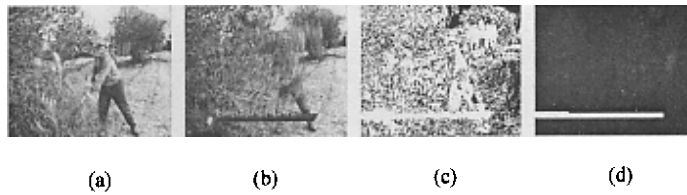


图 3 全局变化剧烈时标题条的检测 (正确)

当然, 该方法成立的一个前提条件是, 在标题可能出现的区域 (每一帧的下部 1/4), 画面非标题部分的噪声碎片比较零散, 不能形成类似新闻标题条的实体矩形块。经过考察, 这个条件在绝大多数情况下是满足的, 但是有少数画面中物体快速移动也可能形成块状区域, 从而导致标题条误检 (图 4), 这也是使检测准确率降低的重要原因 (表 1 中参数 Precision=0.87)。



图 4 全局变化剧烈时标题条的误检

4 总结

本文提出了一种基于形态学操作的新闻标题条检测算法。这种算法充分运用了新闻标题条的时空位置、形状、显隐等特征和一般规律, 采取了分类处理的方式和形态学的滤波操作。实验证明, 该方法能够有效地检测渐变的标题条, 并且对于镜头变化剧烈的情况有一定的抗干扰能力, 所以能够达到较高的查全率和准确率。

由实验和分析还可以看出, 该算法中一些经验参数的选择以及帧画面中物体运动对检测判断准确性的影响仍然存在值得研究的问题。另外, 对于不同的新闻节目, 其标题条的样式也有较大的差异, 如何提高标题条检测算法的鲁棒性还需要进一步的研究。

参 考 文 献

- [1] 熊华, 老松杨, 吴玲琦, 李恒峰, 吴玲达, 李国辉. NewsVideoCAR: 一个基于内容的视频新闻节目浏览检索系统. 计算机工程, 2000, 26(11): 73-75.
- [2] 任金昌, 赵荣椿, 冯大渝. 一个基于内容的数字新闻检索系统的设计与实现. 计算机应用, 2001, 21(10): 49-50.
- [3] Smith M A, Kanade T. Video skimming and characterization through the combination of image and language understanding techniques. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997: 775-781.
- [4] Li H, Doermann D, Kia O. Automatic text extraction and tracking in digital video. *IEEE Trans. on Image Processing*, 2000, 9(1): 147-156.
- [5] Sato T, Kanade T, Huges E K, et al.. Video OCR: Indexing digital news libraries by recognition of superimposed caption. *ACM multimedia system*, 1999, 7(5): 385-395.
- [6] Yeo B L, Liu B. Rapid scene analysis on compressed videos. *IEEE Trans on Circuits and Systems for Video Technology*, 1995, 5(6): 533-544.
- [7] Boreczky J S, Rowe L A. Comparison of video shot boundary detection techniques. In: Proc. of the SPIE Conference Storage and Retrieval for Video Databases IV, San Diego/La Jolla, CA, USA: SPIE, 1996, 2670: 170-179.

姜 帆: 男, 1980 年生, 硕士生, 主要研究方向为基于内容的多媒体检索。

章毓晋: 男, 1954 年生, 教授, 博士生导师, 主要研究方向为图像工程。