一种基于 Homogeneity 的文本检测新方法

黄剑华,唐降龙,刘家锋,徐莉莉
(哈尔滨工业大学 计算机科学与技术学院,黑龙江 哈尔滨 150001)

摘 要:视频图像中的文本包含了丰富的语义层次上的内容描述信息,为基于语义的图像检索提供重要的索引信息资源.提出了一种基于 Homogeneity 和支持向量机(support vector machine)的视频图像中文本检测方法,首先将图像由空间域映射到 Homogeneity 域中,然后对映射到 Homogeneity 空间中的图像进行特征提取,利用 SVM 判别文本区域.实验表明此文本检测方法优于用基于边缘特征的文本检测方法.

关键词:文本检测;特征提取;Homogeneity;支持向量机
中图分类号:TP391.2 **文献标识码**:A **文章编号**:1673-4785(2007)01-0069-05

A new method for text detection based on Homogeneity

HUANG Jian-hua, TANG Xiang-long, LIU Jia-feng, XU Li-li
(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Text data presented in images and video contains useful and important semantic information for automatic indexing. In this paper, a method for text detection based on homogeneity and SVM is proposed. First an original image is mapped from space domain to homogeneity domain, and then text region property is confirmed by SVM trained to extract property feature in homogeneity domain. Comparison with the text detection method based on edge features shows that the proposed method has a better accuracy.

Keywords: text detection; feature extraction; Homogeneity; SVM

随着数字化存储技术的发展和计算机性能的不断提高,数字视频在各个领域的应用越来越广泛,能够从大量的视频资料中找到需要的信息成为人们迫切的要求.图像和视频中的文本包含许多非常重要的信息,如街道名称、商店名称、路标、交通标示、字幕等,这些信息是图像和视频资料自动注释、索引、压缩等方面重要的依据.

从视频图像处理和文档分析的研究角度出发,目前已经提出了一些文本提取算法,这些算法主要是从感性的特征出发,利用颜色、亮度、形状、纹理等属性来提取文本信息,总结起来可以归纳为 3 类:1)基于连通区域的文本检测方法[1],这一方法假定字符的颜色相近且与背景可分;2)基于纹理特征的文本检测方法[2-3],通过识别图像的纹理特征,如角点特征,来区分文本区域与背景;3)基于边缘特征的文本检测方法[4-5],使用边缘和边缘密度来找出文本区域的位置.在视频片断处理中除使用单帧图像外还可以利用多幅图像来检测、提取和增强文本区域[6].也有部分研究者将颜色信息和边缘或者纹理特征结合在一起使用,不在灰度图像上而是在彩色图像上提取边缘特征[7].还有算法直接在压缩的图像中进行文本检测[8].

上述的方法在检测文本时只考虑了图像区域的全局信息,没有考虑局部信息,一定程度上造成文本检测的错误.文中提出了一种基于 Homogeneity 的文本检测方法,这种方法充分考虑了图像区域的局部信息,反映了一个区域的一致性强度,能够更好地反映文本区域的特征,可以更好地突出文本区域,适用于背景比较复杂的视频图像中文本检测.实验表明在 Homogeneity 空间进行特征提取优于用边缘算子进行特征提取.

收稿日期:2006-03-07.
基金项目:国家自然科学基金资助项目(60573071).

1 图像中文本的特征

图像中的文本可以分为2类,即场景文本(scene text)和图形文本(graph text).场景文本出现在场景中,它能够被视频设备捕捉,是一幅图像的组成部分,可作为真实世界中的一个目标,例如:路牌、告示牌、车牌等.而在另一方面,图形文本是为了补充视频图像内容而人工添加的,例如:标题、关键字、摘要、时间、地点等标志,它是为了观看者阅读而产生的.图像中文本一般具备以下特征:

纹理:丰富的边缘、角点;周期出现的高强度和高频率.

色彩:文字多为单色,并与背景有明显的对比,特殊情况下有特定的色彩.

形状:字符的尺寸有一定范围,字符之间的距离不会过大,一段文字一般在同一水平或垂直线上.

其他:前面少有遮挡;文字多为正向;同一段文字会在连续的多帧图像中出现.

可以利用以上这些特征进行文本区域的检测,然而视频图像中的文本,尤其是场景文本往往镶嵌在复杂的背景图像中;文本的颜色、亮度和对比度经常发生变化;文本的大小、排列和对齐方式不确定;受拍摄角度等因素的影响,文本会产生扭曲、变形、残缺、模糊断裂等现象.这些因素都给视频图像中文本的检测造成了极大的困难,因此迫切需要找到一种适用于各种类型视频图像中文本检测的方法.

2 Homogeneity 的定义

Homogeneity 与从图像中提取出的局部信息有关,它的数值反映了一个区域的一致性的强度,因此它可以被用来进行图像和视频中的文本检测,文中把 Homogeneity 定义为2个部分的组合:标准差和强度的不连贯性.

设 $I(i,j)$ 是一幅 $M\times N$ 的图像在 (i,j) 位置上的像素值,$w_n(i,j)$ 是一个以 (i,j) 为中心的 $n\times n$ 大小的窗口,用来计算标准差;$w_m(i,j)$ 是一个以 (i,j) 为中心 $m\times m$ 的大小的窗口,用来计算不连贯性,其中 m,n 为奇数,$m>1$,$n>1$.

像素 $I(i,j)$ 的标准差定义为

$$\sigma(i,j)=\sqrt{\frac{1}{n^2}\sum_{p=i-\frac{n-1}{2}}^{i+\frac{n-1}{2}}\sum_{q=j-\frac{n-1}{2}}^{j+\frac{n-1}{2}}(I(p,q)-\mu(i,j)^2},$$

$$0\leq i\leq M-1,0\leq j\leq N-1. \quad (1)$$

式中:$\mu(i,j)$ 是在 $w_n(i,j)$ 窗口中所有像素的均值,计算方法如下:

$$\mu(i,j)=\frac{1}{n^2}\sum_{p=i-\frac{n-1}{2}}^{i+\frac{n-1}{2}}\sum_{q=j-\frac{n-1}{2}}^{j+\frac{n-1}{2}}I(p,q). \quad (2)$$

将 $\sigma(i,j)$ 归一化,得到归一化的标准差:

$$V_n(i,j)=\frac{\sigma(i,j)}{\sigma_{\max}}. \quad (3)$$

式中:$\sigma_{\max}=\max\{\sigma(i,j)\}$,$0\leq i\leq M-1$,$0\leq j\leq N-1$.

一个像素的不连贯性可以用边缘信息来描述,文中应用 Sobel 算子来计算像素 $I(i,j)$ 的不连贯性 $e(i,j)$:

$$e(i,j)=\sqrt{G_x^2+G_y^2}. \quad (4)$$

将 $e(i,j)$ 归一化,得到归一化的不连贯性:

$$E_m(i,j)=\frac{e(i,j)}{e_{\max}}. \quad (5)$$

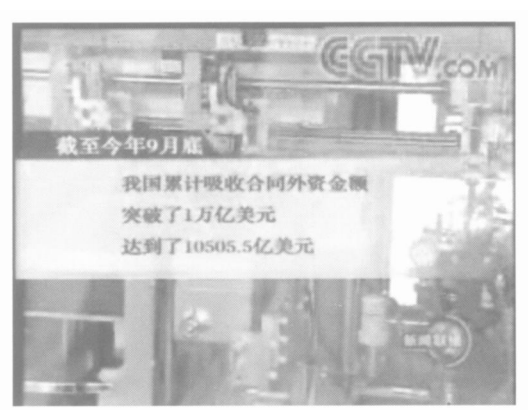
式中:$e_{\max}=\max\{e(i,j)\}$,$0\leq i\leq M-1$,$0\leq j\leq N-1$.

那么,像素 $I(i,j)$ 的 Homogeneity 的定义如下:

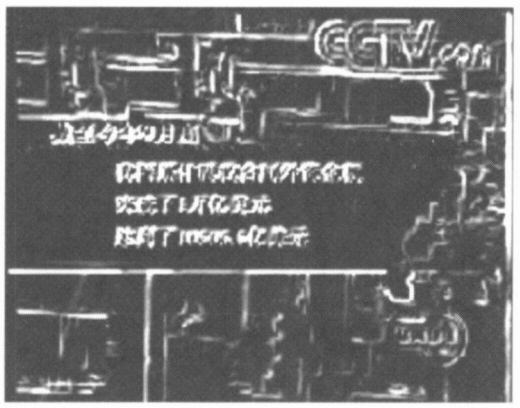
$$H_m(i,j)=(1-E_m(i,j))\times(1-V_n(i,j)). \quad (6)$$

Homogeneity 的取值在0和1之间.由以上的定义可以看出,如果一个区域越均匀,每一点处的归一化标准差 $V_n(i,j)$ 和归一化不连贯性 $E_m(i,j)$ 就越小,计算出每一个点处的 Homogeneity 值 $H_{mn}(i,j)$ 就越大.

图像中文本的主要特性可概括为:不连惯性和高频性,因为文本的区域含有丰富的边角纹理信息.文本区域像素点的 Homogeneity 值比较小,能很好地与背景区域区分开,这个特点为在 Homogeneity 空间来进行文本检测提供了条件.通过计算图像的 Homogeneity 将图像映射到 Homogeneity 空间中得到特征图像,如图1所示.



(a)原始图像



(b) Homogeneity 域中特征图像

图 1　Homogeneity 映射

Fig. 1　Homogeneity mapping

3　基于 Homogeneity 的文本检测

文本获取是指在输入图像中确定文本区域的位置,并标识出来的过程. 文本获取可分为以下几个步骤:文本检测、文本定位、文本提取和字符识别 4 个阶段. 文中主要研究的是文本检测的方法. 首先应用 Homogeneity 映射来对图像进行处理,突出其中文本区域的特征,得到特征图像;然后在得到的特征图像上,使用一个大小为 16 ×16 的滑动窗口得到图像的局部数据,在窗口内提取特征,送入分类器来判别此窗口所对应的图像区域是否为文本区域,从而确定图像中的文本区域.

3.1　预处理

预处理主要是从视频片段中提取视频帧,对图像进行去噪处理,将彩色图像转化为灰度图像等操作.

3.2　基于 Homogeneity 的特征提取

经过预处理后的图像通过 Homogeneity 映射,把图像转换到 Homogeneity 空间,然后使用一个大小为 16 ×16 的滑动窗口来扫描 Homogeneity 空间中的图像,对于窗口覆盖的图像区域,文中使用了如下的 6 个统计量作为特征,这里 G 为对特征图像使用滑动窗口得到的矩阵,$\bar{G}$ 为此矩阵的均值:

密度:

$$D = \sum_{i}^{m}\sum_{j}^{n} G(i,j). \tag{7}$$

均值:

$$M = \frac{1}{m \times n}\sum_{i}^{m}\sum_{j}^{n} G(i,j). \tag{8}$$

二阶矩:

$$M_2 = \frac{1}{m \times n}\sum_{i}^{m}\sum_{j}^{n} (G(i,j) - \bar{G})^2. \tag{9}$$

三阶矩:

$$M_3 = \frac{1}{m \times n}\sum_{i}^{m}\sum_{j}^{n} (G(i,j) - \bar{G})^3. \tag{10}$$

标准差:

$$v = \left(\frac{1}{m \times n - 1}\sum_{i}^{m}\sum_{j}^{n} (G(i,j) - \bar{G})^2\right)^{1/2}. \tag{11}$$

能量:

$$E = \sum_{i}^{m}\sum_{j}^{n} \frac{G^2(i,j)}{m \times n}. \tag{12}$$

式中:$m = n = 16$.

3.3　用 SVM 作为分类器进行分类

文中使用的分类器是支持向量机(SVM),因为 SVM 是从线性可分情况下的最优分类面提出的,它不仅能将 2 类样本无错误的分开,而且使得分类距离最大. 它在很大程度上解决了传统方法(如神经网络)存在的问题,如模型选择、过学习、非线性、多维问题、局部极小点等问题.

文中 SVM 分类器的核函数选择多项式核函数:

$$K(x,y) = (\gamma x^t y + C)^d. \tag{13}$$

实验中选择 $d = 3$ 多项式核函数,参数 $\gamma = 0.1$, $C = 0.1$. 在训练 SVM 时,使用标记为文本属性或非文本属性的图像块作为训练样本,训练样本中 2 种属性图像块的比例对文本检测器的训练结果有直接影响. 训练集中每一幅图像都包含文本区域,但通常文本区域都远少于非文本区域,因此从这些图像直接得到的文本块远远少于非文本块. 为了保证文本检测器对文本和非文本块识别率的均衡,训练样本中文本块和非文本块的比例要适当. 文中比较了不同文本和非文本训练样本比例情况下训练得到的分类器的分类正确率的变化情况,表 1 为实验结果,其中 c 表示分类的正确率,c_T 表示文本区域的正确率,c_B 表示非文本区域的正确率,定义如下:

$$c = R/T_{\text{total}}, \tag{14}$$

$$c_T = R_T/T_{\text{text}}, \tag{15}$$

$$c_B = R_B/B. \tag{16}$$

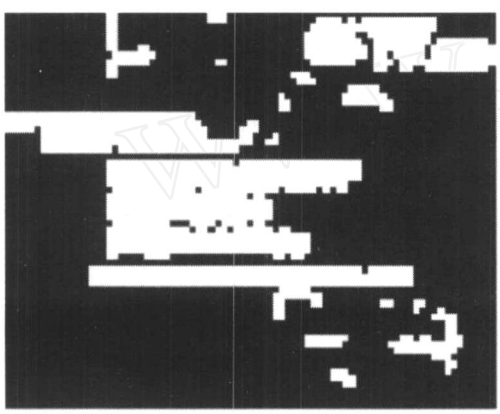
表 1　不同的训练样本比例下的分类结果

Table 1　Classification result of difference training sample rate　　%

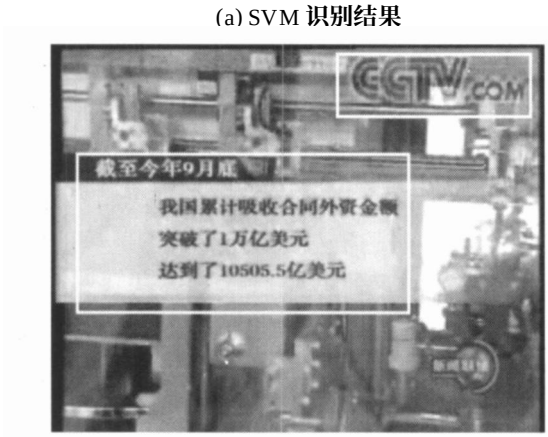
文本:非文本	20:10	15:10	10:10	10:15	10:20	5:15	5:20
c	76.48	78.41	86.79	86.86	88.63	90.98	91.34
c_T	91.95	91.14	88.70	75.93	67.52	53.45	42.23
c_B	74.38	76.68	84.62	88.34	91.50	96.08	98.02

这里 T_{total} 表示送入分类器总的样本数目，R 表示分类正确的样本数目，T_{text} 为文本区域的样本数目，B 为非文本区域的样本数目. 为了保证识别率的均衡，避免文本区域过多的误识，文中选择使用10:10作为训练样本比例(文本:非文本).

在测试时，把滑动窗口得到的6维特征输入到SVM，SVM的输出为0和1分别代表非文本和文本. 使用SVM的分类结果可以得到一个与原始图像对应的二值图像，如图2(a)所示.



(a) SVM 识别结果



(b) 文本检测结果

图2 文本检测实例

Fig. 2 Examples of text detection

4 实验结果

为了验证文中算法的性能，文中作了以下的实验：分别使用边缘算子与 Homogeneity 映射2类方法得到特征图像，然后在所得到的特征图像中按照上述方法，在相同的条件下进行特征提取和分类器分类.

图片样本集为：453幅图片，这些图片是从视频中截取出来的包括动画片、新闻、体育、电影等方面. 其中训练样本为138幅图片，测试样本为315幅图片. 在训练SVM分类器时，根据4.3节的实验结果，训练样本比例选择10:10(文本:非文本)，选取的特征为4.2节中描述的6维统计量. 测试结果如表2所示.

表2 SVM对从视频中提取出图片的分类结果

Table 2 Result of SVM tested on our dataset %

	Robert	Sobel	Canny	LOG	Color Robert	Our method
c	83.40	77.30	80.78	68.72	65.63	86.79
c_T	86.96	89.03	55.85	84.24	89.40	88.70
c_B	82.88	75.58	84.44	66.44	62.15	84.62

文中同时采用2003年国际自然场景文本阅读比赛(ICDAR '2003 Robust Reading Competition)[9]提供的测试集进行了测试，测试集为507幅图片，这些图片都是场景文本图片，训练样本为258幅图片，测试样本为249幅图片. 测试结果如表3所示.

表3 SVM对ICDAR '2003测试图片的分类结果

Table 3 Result of SVM tested on ICDAR '2003 dataset %

	Robert	Sobel	Canny	LOG	Color Robert	Our method
c	74.79	73.72	47.96	69.83	54.61	76.89
c_T	52.78	55.09	70.14	53.99	74.79	57.09
c_B	77.38	76.78	44.31	72.40	51.33	78.36

从实验结果可以看出，在 Homogeneity 空间进行特征提取比用边缘算子直接提取文本的效果好. 图2(b)是利用文中方法进行文本检测的实际效果.

5 结束语

文中提出了一种基于 Homogeneity 的文本检测的方法，Homogeneity 这种方法已经被成功地应用到图像分割中，文中把它应用到文本检测中，通过实验可以看出这种方法是有效的. 由实验结果也可以看出，该算法中的一些经验参数的选择和特征提取、特征选择等问题上还有待研究. 今后将进一步研究多分辨分析和特征选择等问题，进一步提高文本检测的准确率.

参考文献:

[1] JEONG K Y, JUNG K, KIM E Y, et al: Neural network-based text location for news video indexing [J]. IEEE Transactions on Information Theory, 1998, 44 (5): 319 - 323.

[2] KIM K I, JUNG K, KIM J H. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm

[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2003, 25(12):1631 - 1639.

[3]LI H P, DOERMANN D, KIA O. Automatic text detection and tracking in digital video[J]. IEEE Transaction on Image Processing, 2000, 9(1):147 - 156.

[4]CHEN X R, ZHANG H J. Text area detection from video frames[A]. IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing [C]. [s.l.],2001.

[5]LIENHART R, WERNICKE A. Localizing and segmentation text in images and videos[J]. IEEE Transactions On Circuits and Systems For Video Technology, 2000, 12(4):256 - 268.

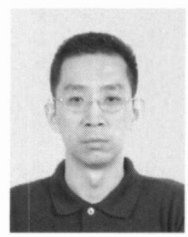
[6]YE Q X, HUANG Q M, GAO W, ZHAO D B. Fast and robust text detection in images and video frames[J]. Image Vision and Computing, 2005(23):565 - 576.

[7]张　引,潘云鹤. 面向彩色图像和视频的文本提取新方法[J]. 计算机辅助设计与图形学报,2002, 14(1):36 - 40.
ZHANG Yin, PAN Yunhe. A new approach for text extraction from color image and video[J]. Journal of Computer-aided Design & Computer Graphics,2002, 14(1): 36 - 40.

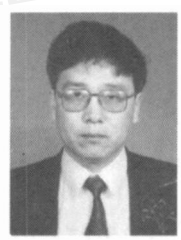
[8]ZHONG Y, ZHANG Hongjiang, JAIN A K. Automatic caption location in compressed video[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2000, 22(4):385 - 392.

[9]LUCAS S M, PANARETOS A, SOSA L. ICDAR 2003 robust reading competition[A]. In: IEEE Proceeding of The 7th International Conference on Document Analysis and Recognition[C]. [s.l.],2003.

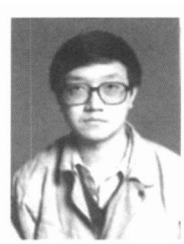
作者简介:



黄剑华,男,1967 年生,副研究员,中国计算机学会会员. 主要研究方向为人工智能、模式识别、图像处理、自然环境下文本认知、人体运动智能分析等.
E-mail:jhhuang@hit.edu.cn



唐降龙,男,1960 年生,教授,博士生导师,主要研究方向为模式识别、人体运动智能分析、人工智能、图象处理医学图象处理、人体生物特征身份鉴别等. 哈尔滨工业大学计算机学院模式识别研究中心主任,中国计算机学会会员,黑龙江省人工智能学会副理事长.



刘家锋,男,1968 年生,副教授,主要研究方向为人工智能、模式识别、中文信息处理等.

第 26 届中国控制会议
The 26th Chinese Contol Conference

由中国自动化学会控制理论专业委员会组织召开的中国控制会议,现已成为有关控制理论与技术的国际性学术年会。大会采用会前讲座、大会报告、分组报告与张贴论文等形式进行学术交流。自 2005 年起会议论文 ISTP(Index to Scientific and Technical Proceedings)收录,自 2006 年起会议论文集进入 IEEE CPP (Conference Publications Program), ISTP 检索。

第 26 届中国控制会议由中国自动化学会控制理论专业委员会主办,中南大学信息科学与工程学院承办,将于 2007 年 7 月在风景秀丽的张家界举行。热忱欢迎海内外广大同仁踊跃投稿参加本届大会,共同交流学术成果。

征文范围如下:系统理论与控制理论;非线性系统及其控制;复杂性与复杂系统理论; 分布参数系统;混杂系统与 DEDS;大系统;随机系统;稳定性与镇定;建模、辨识与信号处理;最优控制与优化;鲁棒控制与 H-nf 控制;自适应控制与学习控制; 变结构控制;神经网络;模糊系统与模糊控制;模式识别;控制设计方法;遗传算法与演化计算;运动控制;智能机器人;分布式控制系统;信息处理系统;故障诊断;通讯网络系统;CIMS 与制造系统;交通系统;生物与生态系统;社会经济系统;工业系统;其他。

征文要求:

1. 论文采用网上投稿,请登陆 http://ccc.amss.ac.cn/pms/ 了解具体事宜并投稿,提交论文截止日期为 2007 年 3 月 1 日。

2. 大会设立关肇直优秀论文奖及张贴论文奖,申请办法和条例请查看控制理论专业委员会网页 http://tcct.amss.ac.cn/ 或会议网页:http://ccc.amss.ac.cn/ 。

3. 拟组织邀请组的组织者,请提供 1000 字的组织建议书及该组全部拟邀请论文的摘要。同一邀请组的论文的主题应鲜明、集中,邀请组一般有 6 篇论文。