

## 基于模糊同质性映射的文本检测方法

黄剑华 承恒达 吴锐 刘家锋

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

**摘要:** 视频图像中的文本是从语义层次对视频图像内容进行描述的非常有效信息, 文本检测为基于语义的图像检索提供了条件。该文提出了一种基于模糊逻辑和同质映射相结合的文本检测方法, 首先利用最大信息熵准则将原始图像模糊化; 然后构造基于边缘信息和纹理信息的图像同质性, 并利用它将图像映射到模糊同质性空间; 最后在模糊同质性空间通过纹理分析检测文本区域。与直接在图像空间域中提取特征的文本检测方法相比, 该方法对复杂背景视频图像的文本检测取得了更好的效果, 并且适用于多种类型的视频图像中文本的检测。

**关键词:** 文本检测; 模糊逻辑; 同质性; 纹理分析

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1009-5896(2008)06-1376-05

## A New Approach for Text Detection Using Fuzzy Homogeneity

Huang Jian-hua Cheng Heng-da Wu Rui Liu Jia-feng

(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 semantic-based image retrieval system. In this paper, a method is proposed for text detection based on fuzzy logic and homogeneity. First, the original images is fuzzified based on the maximum entropy principle. Then edge and textural information is extracted to evaluate the homogeneity of image which is transformed to fuzzy homogeneity domain. Finally text region is confirmed with texture analysis in fuzzy homogeneity domain. Experimental results confirm that the proposed method achieves better performance in complex background and is applicable to various kinds of video and images.

**Key words:** Text detection; Fuzzy logic; Homogeneity; Texture analysis

### 1 引言

由于互联网上多媒体信息的飞速增长, 视频资料的存储、管理和检索的效率越来越成为迫切需要解决的问题。如果计算机能够自动识别视频资料中的文本信息, 将会提供极大的帮助。因此视频图像中的文本信息检测与识别技术具有非常重大的理论意义和实用价值。

文本检测的主要方法有基于连通域的文本检测方法<sup>[1]</sup>、基于边缘的文本检测方法<sup>[2, 3]</sup>和基于纹理的文本检测方法<sup>[4, 5]</sup>。Hua Xian-sheng 等<sup>[6]</sup>提出了一种基于纹理特征和边缘检测相结合的文本检测方法。由于汉字中含有比较丰富的角点, 纹理特征比较明显, 所以这种方法对视频图像中中文文本检测更有效。张引、潘云鹤<sup>[7]</sup>提出了一种结合彩色边缘检测与区域生长的图像分割方法, 对于颜色和背景比较复杂的图像取得了较好的效果。

上述这些方法虽然取得了一定效果, 但是这些方法都是基于图像的空间域特性, 没有充分反映图像本身的特点, 具有一定的局限性和图像类型处理的选择性, 即对特定类型的

图像具有很好的效果, 而对其他类型的图像则效果较差。同时由于视频图像背景复杂及噪声干扰, 图像中文本的颜色、亮度、大小、排列和对齐方式不确定, 文本有扭曲、变形、残缺、模糊等现象, 这些因素都给视频图像中文本的检测与识别造成了极大的困难<sup>[8]</sup>。因此迫切需要找到一种适用于各种类型视频图像的通用的文本检测方法。

由于图像本身具有一定的模糊属性, 如边缘、边界、纹理和对比度的定义, 都是模糊的<sup>[9]</sup>, 模糊集理论具有处理不确定性、含糊性和非精确性的能力<sup>[10]</sup>。因而可以将模糊理论引入到图像增强和分割中, 突出图像中具有某种特殊属性的区域, 使图像增强与分割算法的性能有很大程度的提高。本文提出一种基于模糊同质性映射的文本检测方法, 首先利用隶属度函数将被检测的图像由灰度空间映射到模糊空间; 然后构造基于边缘信息和纹理信息的图像同质性, 并利用它将图像由模糊空间映射到模糊同质性空间; 最后在模糊同质性空间中检测文本区域。

### 2 图像的模糊化处理

对图像进行模糊化处理, 需要根据图像本身的特性选择适当的隶属度函数及其参数。本文选择最常用的标准  $S$  函数<sup>[11]</sup>  $S(g_{ij}, x, y, z)$  作为隶属度函数, 这里  $g_{ij}$  为原始图像的灰

2006-11-09 收到, 2007-04-27 改回

国家自然科学基金(60573071)和国家 863 计划(2003AA5015 22)资助课题

度值,  $x, y, z$  是  $S$  函数的参数。恰当地选择隶属度函数的参数可以使得图像由灰度空间到模糊空间映射的过程中, 有效地抑制噪声, 减少图像信息的丢失。参数  $x$  和  $z$  可以根据图像灰度直方图来确定, 本文选取图像灰度直方图的第一个峰值和最后一个峰值作为参数  $x$  和  $z$  的值。

参数  $y$  利用信息熵准则<sup>[12]</sup>来确定其值的大小, 具体方法如下: 设  $p_i$  是图像灰度  $i, i = 1, 2, \dots, K$  的概率分布,  $K$  为图像灰度级个数, 选取一个阈值  $t$ , 定义小于或等于阈值和大于阈值的灰度分布信息熵分别为  $H_l(t)$  和  $H_g(t)$ :

$$H_l(t) = -\sum_{i=1}^t \frac{p_i}{P_t} \ln \frac{p_i}{P_t} \quad (1)$$

$$H_g(t) = -\sum_{i=t+1}^N \frac{p_i}{1-P_t} \ln \frac{p_i}{1-P_t} \quad (2)$$

其中  $P_t = \sum_{i=1}^t p_i$ 。

根据最大信息熵确定最佳阈值  $t^*$ , 作为  $S$  函数的参数  $y$ ,

$$y = t^* = \text{Arg max}_{t=1}^K \{H_l(t) + H_g(t)\} \quad (3)$$

参数  $x, y, z$  确定之后, 利用  $S$  函数将图像由灰度空间映射到模糊空间, 得到模糊化处理后的图像。

### 3 模糊同质性映射(fuzzy homogeneity mapping)

同质性与图像的局部信息有密切的关系, 实质上反映图像区域的均匀性, 它在图像分割中起到很重要的作用。文本区域的检测实质上是图像分割问题, 而图像分割的目的是将图像划分为若干某种特定属性意义下均匀的区域。本文采用模糊图像中的模糊纹理信息和模糊边缘信息构成图像的模糊同质性。模糊同质性更好地体现了图像的局部信息, 反映图像区域的均匀程度。相对于非文本区域, 文本区域的模糊同质性较小, 可以利用模糊同质性检测文本区域。

设  $g_{ij}$  是  $M \times N$  图像  $G$  在点  $(i, j)$  处的灰度值,  $\mu(g_{ij}) = S(g_{ij}, x, t^*, z)$  是衡量图像模糊信息的隶属度函数, 采用以下方法分别计算模糊纹理信息和模糊边缘信息。

#### 3.1 模糊纹理信息的计算

定义 3 个向量  $L5 = (1, 4, 6, 4, 1)$ ,  $E5 = (-1, -2, 0, 2, 1)$ ,  $S5 = (-1, 0, 2, 0, -1)$ 。 $L5$  表示“线”特征算子,  $E5$  表示“边”特征算子,  $S5$  表示“点”特征算子, 这 3 个算子按如下方式计算可得 4 个掩模:  $L5^T \times E5$ 、 $L5^T \times S5$ 、 $E5^T \times L5$  和  $S5^T \times L5$ 。

设 4 个掩模在图像  $G$  中某像点  $g_{ij}$  处线性滤波的响应(即卷积)分别为  $f_{L5^T \times E5}(\mu(g_{ij}))$ ,  $f_{L5^T \times S5}(\mu(g_{ij}))$ ,  $f_{E5^T \times L5}(\mu(g_{ij}))$ ,  $f_{S5^T \times L5}(\mu(g_{ij}))$ , 则该点模糊纹理信息为

$$f(\mu(g_{ij})) = \left[ \left( f_{L5^T \times E5}(\mu(g_{ij})) \right)^2 + \left( f_{L5^T \times S5}(\mu(g_{ij})) \right)^2 + \left( f_{E5^T \times L5}(\mu(g_{ij})) \right)^2 + \left( f_{S5^T \times L5}(\mu(g_{ij})) \right)^2 \right]^{1/2} \quad (4)$$

并将  $f(\mu(g_{ij}))$  归一化, 得到归一化的模糊纹理信息:

$$F(\mu(g_{ij})) = \frac{f(\mu(g_{ij}))}{f_{\max}} \quad (5)$$

其中  $f_{\max} = \max\{f(\mu(g_{ij}))\}$ ,  $0 \leq i \leq M-1, 0 \leq j \leq N-1$ 。

#### 3.2 模糊边缘信息的计算

模糊边缘信息用来描述灰度等级的不连续性, 是构成同质性的另一因素。本文采用 Sobel 算子计算模糊边缘值  $e(\mu(g_{ij}))$ , 并归一化得到模糊边缘信息:

$$E(\mu(g_{ij})) = \frac{e(\mu(g_{ij})) - e_{\min}}{e_{\max} - e_{\min}} \quad (6)$$

其中  $e_{\max} = \max(e(\mu(g_{ij})))$ ,  $e_{\min} = \min(e(\mu(g_{ij})))$ ,  $0 \leq i \leq M-1, 0 \leq j \leq N-1$ 。

在得到了模糊纹理信息和模糊边缘信息后, 在此基础上构成图像的模糊同质性:

$$HO(\mu(g_{ij})) = (1 - F(\mu(g_{ij}))) \times (1 - E(\mu(g_{ij}))) \quad (7)$$

通过求出原始图像中每一点对应的 HO 值, 就可以把原始图像映射到模糊同质性空间, 从而得到特征图像, 图像经模糊同质映射后的效果如图 1(b)所示。可以看到, 比较均匀的背景图像区域内像素点的模糊同质性数值较大, 而文本区域含有丰富的边角纹理信息, 均匀性较差, 文本区域内像素点的模糊同质性数值较小。本文利用这个特点在模糊同质性空间进行文本检测。

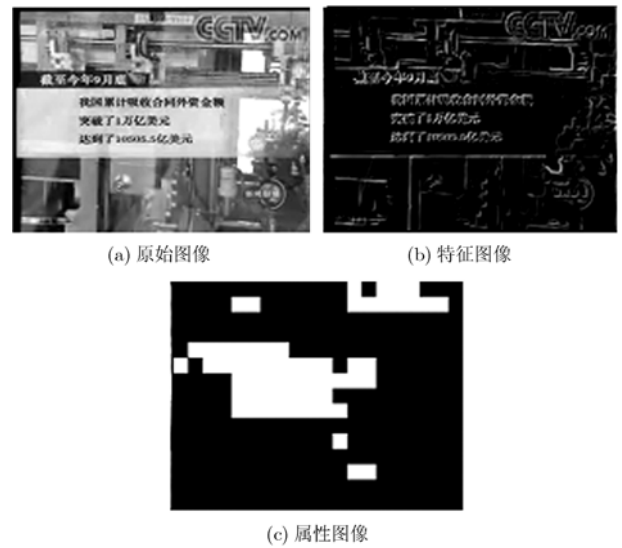


图1 模糊同质性映射

## 4 基于模糊同质性映射的文本检测

与非文本区域相比, 文本区域具有特殊的纹理特征和较高空间频率。本文首先通过模糊同质性映射生成原始图像的特征图像, 突出了这些特征。然后利用  $m \times n$  的窗口扫描特征图像, 提取各个窗口区域的纹理特征, 通过支持向量机(SVM)分类器将此窗口区域区分为文本或非文本生成属性图像, 最后对属性图像进行区域分析处理确定文本区域的最终位置, 并给出文本区域的外接矩形。

#### 4.1 特征提取及分类

本文采用视觉比较敏感但是对光照不敏感的纹理特征作为分类的特征。借助灰度共生矩阵来描述图像区域的纹理。灰度共生矩阵是建立在估计图像的二阶组合条件概率密度函数基础上的纹理分析方法。设  $S$  为目标区域  $R$  中具有特定空间联系的像素对的集合, 则共生矩阵  $P_{\Delta x, \Delta y}$  中位于  $(g_1, g_2)$  的元素  $P_{\Delta x, \Delta y}(g_1, g_2)$  表示灰度为  $g_1$  与灰度为  $g_2$  的两个相距为  $(\Delta x, \Delta y)$  的像素对出现次数的归一化值, 即

$$P_{\Delta x, \Delta y}(g_1, g_2) = \# \{ (x_1, y_1), (x_2, y_2) \in S | f(x_1, y_1) = g_1 \} \\ \& f(x_2, y_2) = g_2 \& (x_1 - x_2 = \Delta x) \\ \& (y_1 - y_2 = \Delta y) \} / \# \{ S \} \quad (8)$$

其中  $\# \{ S \}$  为计算集合  $S$  中像素对数目的运算, 式(8)的分子为集合  $S$  中具有上述空间关系、灰度值分别为  $g_1$  和  $g_2$  的像素对个数, 分母为集合  $S$  中像素对的总和个数。

在共生矩阵基础上, 进一步提取图像的纹理能量、纹理熵、纹理对比度和纹理均匀性 4 种纹理特征。本文使用尺寸为  $16 \times 16$  的窗口扫描特征图像, 计算每一个窗口区域的灰度共生矩阵。计算共生矩阵时, 分别计算  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  4 个方向上的灰度共生矩阵  $P_H, P_{RD}, P_V, P_{LD}$ ,  $\Delta x$  和  $\Delta y$  分别取 1、3、5 3 种情况, 共得到 12 个共生矩阵。对每一个共生矩阵计算上述 4 个纹理特征, 共得到 48 维特征向量。

文本检测分类器采用支持向量机, 选择径向基函数(radial basis function)作为其核函数:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0 \quad (9)$$

通过交叉对比(cross-validation)确定核函数的参数  $C = 50$ ,  $\gamma = 1$ 。

通过分类器确定窗口区域的属性, 具有文本属性的窗口区域分类器输出值为 1, 否则为 0, 用其输出值充填对应的  $16 \times 16$  窗口区域, 生成与特征图像尺寸相同的属性图像, 如图 1(c)。

#### 4.2 基于密度的文本区域确定

在模糊同质性空间中对特征图像进行分类处理后得到属性图像, 其像素可以按照是否属于文本区域, 分为文本像素和非文本像素。图像中的文本区域可以认为是一些文本像素的聚类, 根据文本区域中文本像素密度较大的特点, 文本像素比较集中的区域可以认为是文本区域, 本文提出了一种基于密度的区域确定法, 算法使用  $m \times n$  矩形模板扫描图像, 计算模板中心具有文本属性的候选像素的像素密度, 根据其像素密度的值确定该候选像素是否为种子像素, 这里像素密度为模板内的候选像素数量与模板面积比值。对于每一个种子像素, 将与之密度相关的像素视为是基于密度连通的, 属于同一个文本区域。

图 2 给出了基于密度的文本区域确定法的结果, 没有进行区域确定前的候选区域如图 1(c)所示, 基于密度的文本区域确定法后的结果如图 2(a)。从两幅图片的对比中可以看到: 区域确定法能够将那些候选像素点不够密集的孤立区域

去除。但由于只考虑了密度而没有考虑文本区域本身的特点, 得到的候选区域往往不规则如图 2(a)。在充分考虑文本区域特点的基础上, 本文对文本候选区域进行基于行密度的规格化和平滑处理, 图 2(b)给出了规格化和平滑处理后的结果。可以看出: 对于那些行密度较大的行被填充为文本; 行密度较小的行被填充为背景。为了更直观地说明基于密度的文本区域确定算法的效果, 将通过上述算法得到的文本区域映射到原始图像中, 得到文本区域图像, 如图 2(c)所示。

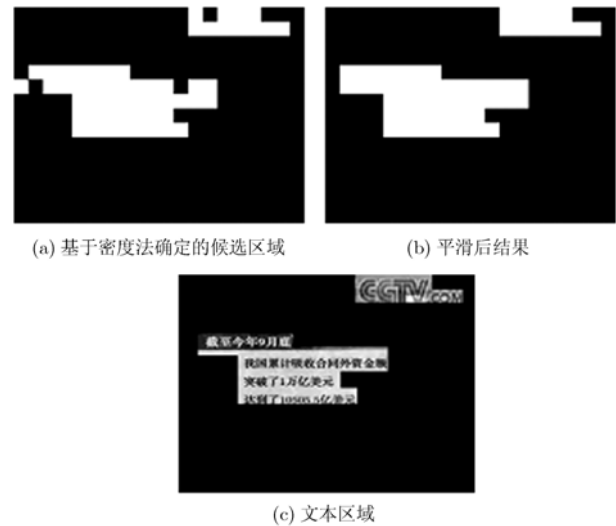


图 2 基于密度的区域确定法

## 5 实验结果及分析

为了验证本文算法对多种视频图像的性能, 我们选择了 5 种常见的视频: 电影、新闻、体育比赛、讲座和动画片, 视频的来源分别为: 中央电视台的新闻联播、体育比赛和讲座, VCD 格式的电影和动画片。从这些视频流中截取了共 717 张图片进行实验, 138 张图片作为训练样本, 抽取图像块 14994 块, 其中 1/3 为文本块, 2/3 为非文本块; 其余 579 张图片做测试样本, 共有图像块 218298 块。本文实验分为两个部分, 模糊同质性映射性能验证和文本检测性能的综合评价。

### 5.1 模糊同质性映射性能验证

本文分别选择基于 Robert, Sobel, Canny, LOG 和 Wavelet 5 种映射算子的文本检测方法, 与基于模糊同质映射文本检测方法的性能进行比较。表 1 给出了使用相同纹理统计特征和 SVM 分类器, 在相同训练样本集和测试样本集条件下不同映射得到的实验结果, 表中的  $F$  由下式定义:

$$F = \frac{2c_t c_b}{c_t + c_b} \quad (10)$$

其中  $c_t$ ,  $c_b$  分别为文本块和非文本块识别的正确率,  $F$  是综合考虑文本和非文本识别的正确率的指标, 越高表示性能越好。由实验结果可以看出, 基于模糊同质映射的方法获得了更高的识别准确率, 表明基于模糊同质映射提取的纹理统

计特征作为分类特征可以得到较好的分类结果。

## 5.2 文本检测方法的综合评价

文本检测方法的综合评价采用 ICDAR 2003 Robust reading competition<sup>[13]</sup>区域匹配方法定义的准确率  $p'$  和召回率  $r'$ , 并设定了一个综合评定指标  $f$ :

$$f = \frac{1}{a/p' + (1-a)/r'} \quad (11)$$

本文实验中, 系数  $a = 0.5$ 。准确率说明检测到的文本区域与实际文本区域的匹配程度, 越高则越准确; 召回率则说明对于实际文本区域的检出率, 越高表示漏掉文本区域越少; 综合评定指标是对两者的综合, 越高表示检测效果越好。

为了说明本文提出的基于模糊同质性映射的文本检测方法的有效性, 本文选择基于边缘密度的文本检测方法和基于小波变换的文本检测方法作比较, 其中边缘密度算子选择效果较好的 LOG 算子, 见表 1。采用相同的测试集分别对上述方法进行文本检测实验, 结果见表 2。从表 2 中可以看出, 本文提出的方法在准确率  $p'$ , 召回率  $r'$  及  $f$  值上均优于其他两种方法。

表 1 模糊同质性映射性能实验结果(%)

	Robert 算子	Sobel 算子	Canny 算子	LOG 算子	Wavelet 映射	模糊 同质性 映射
$F$	83.94	79.58	81.64	85.37	86.39	88.82
$c_t$	85.83	87.30	73.42	70.25	85.30	87.30
$c_b$	83.71	78.62	82.65	87.25	87.51	90.39

表 2 不同文本检测方法对比实验结果(%)

	基于边缘密度 的文本检测 方法	基于小波变换 的文本检测 方法	基于模糊同质性 映射的文本检测 方法
$p'$	79.28	82.21	85.75
$r'$	51.12	54.23	56.56
$F$	62.16	65.35	68.18

通过对实验结果和测试样本的分析, 这 3 种方法在处理背景相对简单的视频图像时都能得到比较好的检测结果。但对那些背景复杂的视频图像, 基于边缘密度和基于小波变换的文本检测方法的准确率明显下降, 本文的方法则仍能取得较好的结果, 说明适用于复杂背景的文本检测, 在整体性能上优于上述两种方法。

## 6 结束语

本文提出了一种基于模糊同质映射的视频图像文本检

测方法, 实验表明, 相对于空间域中常用的基于边缘密度的文本检测方法和基于小波变换的文本检测方法, 该方法对于背景比较复杂的视频图像中文本的检测具有更好的性能, 并且适用于多种类型的视频图像, 具有良好的适应性, 提高了文本检测的正确率。下一步工作将考虑引入字符语言特征, 即利用字符识别引擎对文本候选区域进行识别, 将识别引擎的识别结果作为检测文本区域的依据, 进一步提高文本检测的准确率。

## 参考文献

- [1] Jeong Ki-young, Jung Keechul, and Kim Eun-yi, *et al.*. Neural network-based text location for news video indexing. 1999 International Conference on Image Processing, Kobe, Japan, 1999, 3: 319-323.
- [2] Chen Xiang-rong and Zhang Hong-jiang. Text area detection from video frames. IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing, Beijing, China, 2001: 222-228.
- [3] Lienhart Rainer and Wernicke Axel. Localizing and segmentation text in images and videos. *IEEE Trans. on Circuits and Systems for Video Technology*, 2002, 12(4): 256-268.
- [4] Ye Qi-xiang, Huang Qing-ming, and Gao Wen, *et al.*. Fast and robust text detection in image and video frames. *Image and Vision Computing*, 2005, 23(6): 565-576.
- [5] Kim Kwang-in, Jung Kee-chul, and Kim Jin-hyung. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003, 25(12): 1631-1639.
- [6] Hua Xian-sheng, Liu Wen-ying, and Zhang Hong-jiang. Automatic performance evaluation for video text detection. Document Analysis and Recognition 2001, Proceedings, 6<sup>th</sup> International Conference, Seattle, WA, USA, 2001, 9: 545-550.
- [7] 张引, 潘云鹤. 面向彩色图像和视频的文本提取新方法. 计算机辅助设计与图像学报, 2002, 14(1): 36-40.  
Zhang Yin and Pan Yun-he. A new approach for text extraction from color image and video. *Journal of Computer-aided Design & Computer Graphics*, 2002, 14(1): 36-40.
- [8] Jung Keechul, Kim Kwang-in, and Jain Anil K. Text information extraction in image and video: A survey. *Pattern Recognition*, 2004, 37(5): 977-997.
- [9] Cheng H D and Xu H J. A novel fuzzy logic approach to contrast enhancement. *Pattern Recognition*, 2000, 33(5): 809-819.
- [10] Ross T J. Fuzzy Logic with Engineering Applications. New York, McGraw-Hill, 1995, chapter 12.

- [11] Pal S K and Majumder D K D. *Fuzzy Mathematical Approach to Pattern Recognition*, New York, John Wiley, 1986: 34-40.
- [12] Kapur J N, Sahoo P K, and Wong A K C. A new method for grey level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 1985, 29(3): 273-285.
- [13] Lucas S M, Panaretos A, and Sosa L, *et al.*. ICDAR2003 Robust reading competition. 2003, Proceedings, 7th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, 2003: 682-687.
- 黄剑华: 男, 1967年生, 副研究员, 研究方向为人工智能、模式识别、图像处理、自然环境上下文认知、人体运动智能分析等方面的研究.
- 承恒达: 男, 1944年生, 教授, 博士生导师, 研究方向为人工智能、医学图像处理、人体运动智能分析.
- 吴 锐: 男, 1976年生, 讲师, 研究方向为字符识别、文本提取、图像处理.
- 刘家锋: 男, 1968年生, 博士, 副教授, 研究领域为人工智能、模式识别、中文信息处理.