

# 模糊逻辑和谱聚类的字符图像分割

吴锐<sup>1</sup>, 尹芳<sup>2</sup>, 唐降龙<sup>1</sup>, 黄剑华<sup>1</sup>

(1. 哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001, simple@hit.edu.cn;

2. 哈尔滨理工大学 计算机科学与技术学院, 哈尔滨 150080)

**摘要:** 为了从复杂背景中有效分离出字符图像, 提出了一种基于模糊逻辑的谱聚类字符图像分割方法. 利用最大信息熵准则获得模糊函数的参数, 将原始图像模糊化; 在模糊后的图像上建立像素间的相似矩阵, 文本图像的纹理、灰度及像素间的距离是定义相似函数的依据, 计算相似矩阵最小特征值对应的特征向量, 并对其聚类划分; 利用分类后的特征向量对相似矩阵进行划分, 进而实现原图像的分割. 实验结果表明: 本文方法优于一般的阈值化分割方法, 能够有效处理背景复杂的自然场景文本图像.

**关键词:** 字符图像分割; 谱聚类; 模糊逻辑; 纹理分析

中图分类号: TP391

文献标志码: A

文章编号: 0367-6234(2010)02-0268-04

## Spectral clustering based text image segmentation using fuzzy logic

WU Rui<sup>1</sup>, YIN Fang<sup>2</sup>, TANG Xiang-long<sup>1</sup>, HUANG Jian-hua<sup>1</sup>

(1. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, simple@hit.edu.cn;

2. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)

**Abstract:** A novel approach of text images segmentation based on fuzzy logic and spectral clustering is proposed in this paper in order to extract the text from complex backgrounds. Parameters of the function are obtained according to the maximum entropy principle and the original image is fuzzified. Then gray, distance and textural information among pixels are extracted from the fuzzified image to construct the affinity matrix. The original image is segmented using the clustered eigenvector corresponding to the minimum eigenvalue of the matrix. Experimental results show that the proposed method is superior to the usual thresholding methods and can handle the natural scene text image with complex backgrounds.

**Key words:** text image segmentation; spectral clustering; fuzzy logic; texture analysis

研究计算机在自然场景下自动识别文本, 具有广泛的应用领域, 如基于内容的图像检索, 汽车自动驾驶, 车牌牌照识别, 工业自动化等等<sup>[1]</sup>. 文本图像的二值化分割是字符识别前的一个关键步骤, 字符分割的质量对字符识别的性能有重要影响. 自然场景下, 文本图像的背景是复杂的, 文字可能存在于任何背景之上, 字符笔划可能有多个亮度值, 这使得从背景中分离出字符图像十分困难. 已有方法多是利用阈值化实现字符图像与背

景的分割, 包括全局阈值方法<sup>[2-3]</sup>和局部阈值方法<sup>[4-5]</sup>.

对复杂的文本字符图像进行分析注意到, 为了突出文本, 图像中文本的颜色、亮度往往不同于背景; 同时, 汉字中含有比较丰富的角点, 纹理特征比较明显, 这些信息都有助于字符图像与背景的分割. 另一方面, 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. 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>[6-7]</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得到提高. 因此, 本文提出一种基于模糊逻辑的谱聚类字符图像分割方法.

收稿日期: 2009-03-10.

基金项目: 国家自然科学基金资助项目(60672090); 黑龙江省自然科学基金资助项目(QC2009C35).

作者简介: 吴锐(1976—), 男, 博士研究生, 讲师;  
唐降龙(1960—), 男, 教授, 博士生导师.

## 1 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  $S(g_{ij}, x, y, z)$  [8] 作为隶属度函数, 其中,  $g_{ij}$  为原始图像的灰度值,  $\mu(g_{ij}) = S(g_{ij}, x, y, z)$  为相应的模糊灰度,  $x, y, z$  分别为  $S$  函数的参数.

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  $y$  利用信息熵准则来确定其值的大小, 具体方法为: 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  $H_g(t)$ :

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

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

$$f(\mu(g_{ij})) = \sqrt{(f_{L5^T \times E5}(\mu(g_{ij})))^2 + (f_{L5^T \times S5}(\mu(g_{ij})))^2 + (f_{E5^T \times L5}(\mu(g_{ij})))^2 + (f_{S5^T \times L5}(\mu(g_{ij})))^2}.$$

并把  $f(\mu(g_{ij}))$  归一化得到  $F(\mu(g_{ij})) = \frac{f(\mu(g_{ij}))}{f_{\max}}$ . 其中,  $f_{\max} = \max\{f(\mu(g_{ij}))\}, 0 \leq i \leq M-1, 0 \leq j \leq N-1$ , 图像大小为  $M \times N$ .

## 3 基于模糊逻辑的谱聚类方法

### 3.1 谱聚类理论

谱聚类是一种点对聚类算法, 由于能够对任意形状的数据进行最优划分, 因而其近年在图像分割、Web 网页检索以及生物信息学等 [9-12] 领域的应用越来越引起研究者的重视. 谱聚类算法的思想源于谱图划分理论 [13], 它将图像看作一个无向加权图  $G = (V, E)$ ,  $V$  为顶点集合,  $E$  为连接边的集合. 每个顶点对应图像的一个像素或区域, 连接任意两个顶点的边的权值  $w(p, q)$  表示顶点  $p$  和  $q$  属于同一类别的可能性, 权值的大小与两顶点的相似性、邻近程度以及连续性等相关. 基于图论的最优划分就是将顶点划分成不相交的集合  $V_1, V_2, \dots, V_m$ , 使得集合内的相似度高, 集合间的相似度较低. 根据组合理论, 计算图划分准则的最优解是一个 NP 难问题, 一种可行的方法是采用连续松散处理, 将该问题转换成求解相似矩阵或拉普拉斯矩阵的谱分解, 因而, 这类方法统称为谱聚类.

式中:  $P_t = \sum_{i=1}^t p_i$ . 根据最大信息熵确定最佳阈值  $t^*$ , 作为  $S$  函数的参数  $y$  为

$$y = t^* = \underset{t=1}{\text{Argmax}}^K \{H_l(t) + H_g(t)\}. \quad (1)$$

$x, y, z$  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.

## 2 模糊纹理信息

定义 3 个向量:

$$L_5 = (1, 4, 6, 4, 1),$$

$$E_5 = (-1, -2, 0, 2, 1),$$

$$S_5 = (-1, 0, 2, 0, -1).$$

式中:  $L_5$  为“线”特征算子,  $E_5$  为“边”特征算子,  $S_5$  为“点”特征算子, 这 3 个算子计算可得 4 个模板:  $L5^T \times E5, L5^T \times S5, E5^T \times L5, S5^T \times L5$ .

设: 4 个模板在图  $J$  中某点  $g(i, j)$  处线性滤波的响应 (即卷积) 分别为  $f_{L5^T \times E5}(\mu(g_{ij}))$ ,  $f_{L5^T \times S5}(\mu(g_{ij}))$ ,  $f_{E5^T \times L5}(\mu(g_{ij}))$ ,  $f_{S5^T \times L5}(\mu(g_{ij}))$ , 则该点纹理信息为

两个集合  $A$  和  $B$  之间的不相似程度, 可用  $A$  与  $B$  间所有连接边的权值之和来度量, 这个总权值称为切 (cut):

$$\text{cut}(A, B) = \sum_{p \in A, q \in B} w(p, q). \quad (3)$$

Wu 和 Leahy 提出最小化上述切值来划分图  $G$ , 这一划分准则被称为最小切准则 (minimum cut). 利用这个准则对图像进行分割取得了较好的效果, 但该准则易倾向于较小的分割 (歪斜分割). Shi 和 Malik 在此基础上提出了规范化切 (normalized cut, Ncut) 划分准则:

$$\text{Ncut}(A, B) = \frac{\text{cut}(A, B)}{\text{assoc}(A, V)} + \frac{\text{cut}(A, B)}{\text{assoc}(B, V)}. \quad (4)$$

式中:  $\text{assoc}(A, V) = \sum_{p \in A, v \in V} w(p, v)$ ,  $\text{assoc}(B, V) = \sum_{q \in B, v \in V} w(q, v)$  分别为  $A$  或  $B$  到整个顶点集合  $V$  的关联度, 最小化 Ncut 函数即可实现图  $G$  的最优划分. 规范化切准则既可衡量类内数据间的相似程度, 也能体现类间数据间的相异程度, 可有效避免歪斜分割.

根据瑞利商理论, 求解式 (4) 中 Ncut 的最小值可通过计算特征值获得:

$$D^{-\frac{1}{2}} L D^{-\frac{1}{2}} z = \lambda z. \quad (5)$$

式中:  $L = D - W$  称为图  $G$  的拉普拉斯矩阵,  $D$  为  $N \times N$  对角矩阵( $N$  为顶点个数), 其对角线元素  $D_{pp} = \sum_{q \in N} w(p, q)$ ,  $W$  为相似矩阵, 其元素为  $w(p, q)$ , 显然有  $w(p, q) = w(q, p)$ . 对应第二小特征值的特征向量  $z_2$  包含了图  $G$  的最优划分信息. 通过聚类若将向量  $z_2$  中的元素分为两类, 即可实现原图像的二值分割.

### 3.2 相似性权值的确定

相似矩阵  $W$  的构造依赖于权值函数  $w(p, q)$  的定义, 权值函数的选取直接影响分类的效果. 目前, 关于相似矩阵的构造还没有一个通用的规则<sup>[14]</sup>, 通常依赖于领域知识的指导. 在本文中, 连接边的权值函数  $w(p, q)$  定义为

$$w(p, q) = \begin{cases} \exp\left(\frac{-\|G(p) - G(q)\|_2^2}{\sigma_I^2} + \frac{-\|F(p) - F(q)\|_2^2}{\sigma_F^2}\right) \cdot \exp\left(\frac{-\|X(p) - X(q)\|_2^2}{\sigma_X^2}\right), & \|X(p) - X(q)\|_2 < r \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

式中:  $G(p)$  为图像模糊化处理后点  $p$  模糊灰度值, 其值在  $[0, 1]$  之间,  $F(p)$  为相应的模糊纹理特征,  $X(p)$  为点  $p$  的空间位置.  $\sigma_I$  为灰度高斯函数的标准方差,  $\sigma_F$  为纹理高斯函数的标准方差,  $\sigma_X$  为空间距离高斯函数的标准方差, 分别用来调节节点间的灰度差异、纹理差异和空间位置差异,  $r$  为两像素之间的有效距离, 超过这一距离则认为两像素之间的相似度为 0,  $r$  同时控制矩阵  $W$  的稀疏程度. 根据权值函数, 两像素间的灰度值越接近, 纹理值越接近, 距离越近, 则两像素之间的相似度越大.  $\sigma_I, \sigma_F, \sigma_X$  及  $r$  的值可根据实验确定.

### 3.3 算法描述

模糊逻辑的谱聚类方法包括 3 个步骤:

1) 构造相似矩阵  $W$ .

① 对图像模糊化处理;

② 计算图像每点模糊纹理信息;

③ 计算图像中任意两点间的  $w(p, q)$  值, 构造相似矩阵  $W$ ;

④ 定义向量  $d$ : 设  $D$  是  $N \times N$  对角矩阵, 其对角元素为  $d, N$  为图像像素个数,

$$d(p) = \sum_q w(p, q);$$

2) 计算特征系统:  $D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}z = \lambda z$ .

3) 对第二小的特征值对应的特征向量  $z_2$  的元素二聚类, 利用分类后的  $z_2$  将图像像素分割成两类.

在运用谱聚类方法进行图像分割时, 计算相似矩阵的特征值和特征向量的复杂度通常很高, 求解一个  $N \times N$  大小矩阵的特征系统其计算复杂度为  $O(N^3)$ ,  $N$  为图像大小. 本文中的字符图像是经过定位分割后包含文字的矩形区域, 通常尺寸较小, 因此适合应用谱聚类方法.

## 4 实验结果及分析

通过一系列的实验来测试本文所设计方法的性能, 在实验中, 运行环境为 PIV 1.73 G, 内存 512 M. 图 1 (b) 是本文方法对图 1 (a) 中图像模糊化提取纹理后的效果图, 图 1 (c) 为分割后的结果. 可以看出, 利用模糊灰度和纹理信息作为权值函数的谱聚类方法, 能够准确分割出颜色复杂的文本图像.



(a)原始图像

(b)模糊纹理图像

(c)谱聚类分割图像

图 1 基于模糊逻辑的谱聚类分割

为了对比本文方法的分割性能, 选取了基于最大类间方差的 otsu 方法与基于 Ncut 的阈值分割方法<sup>[15]</sup>与本文方法进行二值化比较. 在实验中, 对于中文文本, 式(6)中的参数设置均为  $\sigma_I = 0.1, \sigma_F = 0.1, \sigma_X = 10, r = 10$ ; 对英文字符, 式(6)中参数取  $\sigma_I = 0.1, \sigma_F = 0.01, \sigma_X = 10, r = 10$ .

图 2 是 3 种方法的分割结果. 在图 2 中, 自上

到下分别是原始文本图像、otsu 方法分割结果、基于 Ncut 的阈值方法分割结果以及本文方法的分割结果. 可以看出, 本文方法明显优于另两种阈值分割方法. 在图像前景(字符区域)的亮度单一、一致的情况下, 基于 Ncut 的阈值分割方法也有很好的分割效果, 显示出图谱准则作为分割量度的良好性能. 当字符前景亮度不一致或者背景复杂时, 本文方法能获得更为理想的效果.



图2 3种方法对复杂文本图像的分割结果

为了验证本文方法对不同环境下图像的分割性能,选取了460幅来自数码相机、手机、因特网上的自然场景下的字符图像,图像大小不超过65 536像素.同时,定义正确率 $p$ 和召回率 $r$ 对分割结果进行评价,并参照文献[16],定义一个综合评价指标 $f$ , $p$ 、 $r$ 和 $f$ 为

$$p = \frac{\text{com\_Image}_o \cap \text{base\_Image}_o}{\text{com\_Image}_o} \quad (7)$$

$$r = \frac{\text{com\_Image}_o \cap \text{base\_Image}_o}{\text{base\_Image}_o} \quad (8)$$

$$f = \frac{1}{a/p + (1-a)/r} \quad (9)$$

式中: $a$ 为准确率和召回率之间的一个相对权重,设 $a=0.5$ .其中, $\text{com\_Image}_o$ 为分割后目标点的像素集合,这里目标点全部为文本区域; $\text{base\_Image}_o$ 为人工标定出的文本区域.测试结果如表1所示.从表1中可以看出,本文方法在正确率 $p$ 、召回率 $r$ 和 $f$ 值上均优于其他两种方法.召回率高表明本文方法能够尽量多地获得有效字符区域,综合评价指标高反映出本文方法的分割效果更好.

表1 3种方法对字符图像的分割结果 %

所用方法	$p$	$r$	$f$
ostu 方法	79.21	70.04	74.34
基于 Neut 的 阈值方法	82.46	79.15	80.77
本文方法	86.12	82.25	84.14

通过对实验结果和测试样本的分析,3种方法在处理背景简单字符清晰的图像时都能得到比较好的分割效果,但对那些前景与背景对比不强烈或者因光照致使字符图像亮度不均匀的复杂图像,ostu方法和基于 Neut 的阈值方法分割性能明显下降,本文所设计的方法则仍能取得较好的分割效果.说明本

文方法适用于自然场景下的复杂字符图像的二值化分割,在整体性能上优于另两种方法.

## 5 结 论

1) 本文充分考虑字符图像的特点,利用文本特有的纹理特征及灰度和距离信息构造相似函数,并使用模糊逻辑表达字符图像的诸多不确定性,提出了一种基于模糊逻辑的谱聚类字符图像分割方法.

2) 实验结果表明,本文方法优于基于最大类间方差的 otsu 阈值分割方法和基于 Neut 准则的阈值分割方法.

3) 本文方法适用于处理背景比较复杂的自然场景字符图像的分割.

## 参考文献:

- [1] JUNG K, KIM K I, JAIN A K. Text information extraction in image and video: A survey [J]. Pattern Recognition, 2004, 37(5): 977-997.
- [2] TSAI Tsung-Han, CHEN Yung-Chien, FANG Chih-Lun. A comprehensive motion videotext detection localization and extraction method [C]//Proc of IEEE 23rd International Conference on Data Engineering Workshop. Washington: IEEE Computer Society, 2007: 113-116.
- [3] PAN W M, BUI T D, SUEN C Y. Text segmentation from complex background using sparse representations [C]//Proc of the Ninth International Conference on Document Analysis Recognition. Washington: IEEE Computer Society, 2007: 412-416.
- [4] LIENHART R, WERNICKE A. Localizing and segmenting text in images and videos [J]. IEEE Transaction on Circuits and Systems for Video Technology, 2002, 12(4): 256-268. (下转第276页)