doi:10.3772/j.issn.1002-0470.2013.10.005

基于谱聚类的改进的文本图像分割方法①

尹 芳②[*,**] 吴 锐③[***] 陈德运[*] 于晓洋[**]

([*]哈尔滨理工大学计算机科学与技术学院 哈尔滨 150080)
([**]哈尔滨理工大学测量技术与通信工程学院 哈尔滨 150080)
([***]哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

摘 要 针对谱聚类方法在图像分割时的高复杂性,提出了一种基于归一化割(Ncut)的改进的谱聚类文本图像分割方法。该方法以经过量化后的颜色集合作为图分割中的顶点以简化加权图模型,从而显著降低谱聚类时的计算复杂性。首先根据文本图像特点建立相似性权值函数,然后根据场景文本颜色分布特性按照颜色直方图对色彩空间进行量化,并以量化后的颜色等级为单位构造相似矩阵,最后在 Ncut 准则下利用谱聚类方法实现图像分割。在包括 ICDAR 2009、2003 竞赛测试集以及其他大量文本图像上的实验表明,该方法具有良好的文本分割性能。

关键词 图像分割,文本图像,谱聚类,归一化割,相似矩阵

0 引 言

文本分割是指将图像中的文本信息从背景中分离出来,自然场景图像中的文本信息对于图像检索和认知具有重要意义[1]。文本分割是文本识别的必要且重要的步骤,它直接影响文本识别系统的性能。近年来,随着自然场景文本提取技术研究的深入[2],文本分割问题越来越受到关注,文本分割正成为文档分析领域中的研究热点。场景图像中的文本区域存在许多不利于分割的因素,如分辨率低、光照不均匀、有噪声干扰、文本区域格式和方向不确定、字符颜色不一致等,使得文本分割的难度很大。已有方法多无法综合考虑各种干扰因素[3-5],即使考虑了,算法性能也不理想[6]。近年来,使用聚类分割色彩提取彩色文本的方法逐渐获得青睐。Thillou 等[7]为提取颜色、亮度和空间信息,提出了基于度量的聚类选择方法,以同时满足文本提取和字符识别的分割要求。目前源于谱图划分理论的谱聚类方法越来越多地应用于图像分割,而且随着划分图的准则的不同出现了多种谱聚类方法。Shi 等[8]提出的归一化割(normalized cut, Ncut)准则可以有效避免划分出图中孤立点的问题,取得了令人满意的图像分割效果。但由于需要求解高维特征系统,复杂度高,应用范围受限。Tao[9]提出的基于 Ncut 准则的阈值分割方法按照灰度等级简化带权图,大大减少了构造相似矩阵的计算量,提高了算法的实时性,当图像背景与目标的灰度对比明显时可得到很好的分割效果,但无法处理彩色图像。受此启发,本文提出了一种基于 Ncut 准则的改进谱聚类文本图像分割方法,该方法首先建立相似性权值函数,然后根据场景文本颜色分布特征在 HSV 颜色空间对像素色彩进行量化,利用颜色直方图在量化后的颜色等级上重新构造加权图,最后利用谱聚类方法实现场景文本图像分割。该方法能够针对彩色图像,并在较小的颜色级数上进行谱图分割,进一步降低谱聚类时的计算复杂度。

1 谱图分割方法

谱图分割的基本思想是将图像看作一个带权图 $G=(V,E)$,每个顶点对应图像的一个像素或区域,

① 国家自然科学基金(61073128),中央高校基本科研业务费专项资金(HIT.NSRIF.2012048),黑龙江省自然科学基金(QC2009C35)和黑龙江省教育厅科学技术研究(12511098)资助项目。
② 女,1978 年生,博士后,副教授;研究方向:模式识别,图像处理,计算机视觉;E-mail: yf812@sohu.com
③ 通讯作者,E-mail: simple@hit.edu.cn
(收稿日期:2012-05-30)

连接任意两个顶点的边的权值 $w(p,q)$ 表示顶点 p 和 q 属于同一区域的可能性,其大小与两顶点的相似性、邻近程度以及连续性等相关。分割的目标是将顶点划分成不相交的集合,使得集合内的相似度较高,集合间的相似度较低。两个集合 A 和 B 之间的相似程度,可以用 A 与 B 间所有连接边的权值之和即割(cut)来度量:

$$cut(A,B) = \sum_{p\in A,q\in B} w(p,q) \tag{1}$$

寻找图中的最小割,就是对集合 V 的最优划分,也是图 G 的一个最优划分。

Shi 和 Malik[8] 在此基础上提出的 Ncut 的定义是

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \tag{2}$$

其中 $assoc(X,V) = \sum_{p\in X,v\in V} w(p,v)$,表示 X 到整个顶点集合 V 的关联度。求解式(2)的最小值问题等价于求解式

$$\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{D}-\boldsymbol{W})\boldsymbol{D}^{-\frac{1}{2}}z = \lambda z \tag{3}$$

的标准特征系统。其中 $\boldsymbol{W}$ 为相似矩阵,其元素为 $w(p,q)$,$\boldsymbol{D}$ 为 $N\times N$ 对角矩阵,$D_{pp} = \sum_{q\in N} w(p,q)$,$\lambda$ 和 z 分别为相应的特征值和特征向量。最小非零特征值 λ_2 对应的特征向量 z_2 对应图 G 的一个最优划分[8]。

2　基于颜色空间量化的谱聚类算法

实现谱聚类分割方法的关键步骤是构建相似矩阵。为了降低相似矩阵的维数,本文采用在量化后的颜色空间上建立相似矩阵,即利用在颜色空间中对图像预分割的结果,将基于像素的权值模型转化为基于集合的权值模型。

2.1　定义相似性权值函数

相似矩阵通过权值函数 $w(p,q)$ 描述像素之间的相似度。根据自然场景中的文本图像颜色差别显著、纹理特殊、文本区域相对集中的特点,定义图像中连接两点 p、q 的边的权值函数 $w(p,q)$ 为以下的形式:

$$w(p,q) = \begin{cases} \exp\left(\dfrac{-\|P(p)-P(q)\|_2^2}{\sigma_P^2} + \dfrac{-\|F(p)-F(q)\|_2^2}{\sigma_F^2} + \dfrac{-\|X(p)-X(q)\|_2^2}{\sigma_X^2}\right), & \|X(p)-X(q)\|_2 < r \\ 0, & \text{其他} \end{cases} \tag{4}$$

其中:$P(p)$、$F(p)$、$X(p)$ 分别表示像素点 p 的颜色或灰度值、纹理特征和空间位置;σ_P、σ_F、σ_X 为颜色、纹理、空间距离的高斯函数标准方差,用来调节像素点间的差异;r 为两像素之间的有效距离,控制矩阵 $\boldsymbol{W}$ 的稀疏程度,超过该距离则认为两像素之间的相似度为零。σ_P、σ_F、σ_X 及 r 值由实验确定。

2.1.1　颜色特征计算

与 RGB 颜色空间相比,HSV 颜色空间更接近于人们对颜色的主观认识,因此本文采用 HSV 颜色空间计算颜色特征,颜色距离计算公式如下式所示:

$$\begin{aligned} d(p,q) = 1 - 1/(\sqrt{5}[&(v_p - v_q)^2 + (s_p\cos(h_p) - s_q\cos(h_q))^2 \\ &+ (s_p\sin(h_p) - s_q\sin(h_q))^2]^{1/2}) \end{aligned} \tag{5}$$

其中 (h_p,s_p,v_p) 和 (h_q,s_q,v_q) 分别代表两种 HSV 空间中的颜色。该度量相当于一个圆柱形颜色空间中的欧拉距离,此时空间中的颜色值表示为 $(s\cosh, s\sinh, v)$。

2.1.2　纹理特征计算

定义 $\boldsymbol{L5} = (1,4,6,4,1)$,$\boldsymbol{E5} = (-1,-2,0,2,1)$ 和 $\boldsymbol{S5} = (-1,0,2,0,-1)$ 三个向量[10],分别表示线、边、点特征算子,计算 $\boldsymbol{L5}^{\mathrm{T}}\times\boldsymbol{E5}$、$\boldsymbol{L5}^{\mathrm{T}}\times\boldsymbol{S5}$、$\boldsymbol{E5}^{\mathrm{T}}\times\boldsymbol{L5}$、$\boldsymbol{S5}^{\mathrm{T}}\times\boldsymbol{L5}$,可得 4 个模板。

设 4 个模板在图像 $I(i,j)$ 中某点处线性滤波的响应(即卷积)分别为 $f_{L5^T\times E5}(i,j)$,$f_{L5^T\times S5}(i,j)$,$f_{E5^T\times L5}(i,j)$,$f_{S5^T\times L5}(i,j)$,则该点纹理信息如下式所示:

$$\begin{aligned} f(i,j) = [&(f_{L5^T\times E5}(i,j))^2 + (f_{L5^T\times S5}(i,j))^2 \\ &+ (f_{E5^T\times L5}(i,j))^2 + (f_{S5^T\times L5}(i,j))^2]^{\frac{1}{2}} \end{aligned} \tag{6}$$

为方便计算,把 $f(i,j)$ 归一化为

$$F(i,j) = \frac{f(i,j)}{f_{\max}} \tag{7}$$

其中 $f_{\max} = \max\{f(i,j)\}$,$0\leqslant i\leqslant M-1$,$0\leqslant j\leqslant N-1$,$M$、$N$ 分别为图像高度和宽度。

2.2　彩色图像的量化处理

一幅图像内所包含的实际颜色数只是全部颜色的一个很小的子集,进一步观察会发现,图像中主要色彩覆盖了绝大多数像素,而对于文本图像来说,文本像素通常属于主要颜色。HSV 空间中,H 分量和 S 分量共同决定图像的色彩,H 分量对色彩的区分能力更强,本文采用查色表颜色量化技术,并将量化后的 H、S、V 三个颜色分量组合成一维特征颜色分量 L[11]:

$$L = \begin{cases} 0, & V < 0.2 \\ \left[\dfrac{7(V-0.2)}{0.8}\right], & S < 0.2, V \geqslant 0.2 \\ 4H + 2S + V + 8, & \text{其他} \end{cases} \tag{8}$$

其中，$[x]$ 表示不小于 x 的最小整数。在 L 个颜色等级上构造直方图，并按直方图子集大小排序，通过实验，取满足式

$$\sum_{k}^{L} H_k \geqslant 0.9V, H_k \geqslant H_{k+1} \tag{9}$$

时的最小 L。文本图像是灰度图像则依据灰度等级量化。

2.3 相似矩阵构造及分割算法

对给定的一幅图像，通过计算图中所有顶点间的权值可以构建基于灰度级的相似矩阵[9]，类似地，借助图像的颜色直方图，可以构造基于颜色等级的相似矩阵。

设 $I(i,j)$ 是大小为 $M \times N$ 的图像，P_{ij} 为图像像素 (i,j) 在 HSV 空间中的颜色值，该颜色是一个三维矢量。设 $V = \{(i,j), 0 \leqslant i \leqslant M-1, 0 \leqslant i \leqslant N-1\}$，$B_1, B_2, \cdots, B_L$ 为量化后的颜色值，令 $H_k = \{(i,j) \mid (i,j) \in V, P_{ij} \in B_q, 1 \leqslant k \leqslant L\}$，那么 $V = \cup_{k=1}^{L} H_k$ 为 V 的一个分割，且有 $H_u \cap H_v = \varnothing$，$\forall u \neq v, u, v \in LL$，这里每一个 H_k，称为 B_k 的直方图子集。于是，图像对应的图 $G = (V,E)$ 的一个二划分为 $V = \{A,B\}$，$A = \cup_{k \in L_A} H_k$，$B = \cup_{k \in L_B} H_k$，同时有 $L_A \cap L_B = \varnothing$，$L_A \cup L_B = LL$，$L_R = \{k \mid P_{ij} \in H_k, \forall (i,j) \in R\}$。令

$$cut(H_u, H_v) = \sum_{p \in H_u, q \in H_v} w(p,q) \tag{10}$$

式(10)表示 H_u 中所有点(颜色区间为 B_u)与 H_v 中所有点(颜色区间为 B_v)之间总的连接权值之和。

构造相似矩阵的过程如下：

步骤 1 构造给定图像的颜色直方图。对于一幅 RGB 图像，将其变换到 HSV 颜色空间，按照式(8)的量化方法，找到每个像素对应的量化区间，统计每个区间中像素个数，即得到以颜色(区间)为横坐标，颜色出现的次数为纵坐标的颜色直方图。若为灰度图像，则相应地构造灰度直方图。

步骤 2 建立基于颜色直方图等级的权值矩阵 $\boldsymbol{W}$。$\boldsymbol{W} = (w_{u,v})_{L \times L}$，$L$ 为直方图的量化等级，L 由式(9)确定，按照式(4)和式(10)计算 $w_{u,v} = cut(H_u, H_v)$，必然 $w_{u,v} = w_{v,u}$。

基于改进谱聚类的文本图像分割算法概括为三个步骤：构建基于颜色量化等级的相似矩阵 $\boldsymbol{W}$；计算 $\boldsymbol{W}$ 对应拉普拉斯矩阵的前 k 个特征值与特征向量，构建特征向量空间；利用 k-means 或其他经典聚类算法对特征向量空间中的特征向量进行聚类，实现矩阵 $\boldsymbol{W}$ 的划分。

3 算法复杂性分析

本文算法的相似度矩阵 $\boldsymbol{W}$ 的维数取决于颜色量化等级 L，与图像大小无关，算法需求解一个 $L \times L$ 维而非 $N \times N$ 维(N 是图像的像素个数)的特征系统，通常 L 远小于 N，因而算法的计算复杂度和空间复杂度都大为降低。下面分析算法的计算量。

本文方法的计算量主要包括两部分：一是构造基于颜色空间量化等级的相似矩阵 $\boldsymbol{W}$ 所需的时间 T_1；二是求解特征系统(3)式所需要的时间 T_2。建立相似矩阵 $\boldsymbol{W}$ 的计算量与参数 r 有关，r 越大，式(4)中需要计算的顶点数越多，计算量也相应增加。当 $r = 1$ 时，式(4)无实际意义；当 $r > 1$ 时，除图像边界上的像素外，每个像素均有

$$t_1 = \frac{[2(r-1)+1]^2 - 1}{2} \times N = 2r(r-1)N \tag{11}$$

式中除以 2 是因为对无向图 G 中每个顶点只需计算一次权值。设每次计算(4)式所需时间为 $s1$，另设每次统计量化区间里像素数的累加运算时间为 $s2$，则 $T_1 = t_1 \times (s1 + s2)$，$T_1$ 与图像大小近似正比。

计算 T_2 需要求解一个 $L \times L$ 维矩阵的特征系统(L 为颜色空间量化级数)，标准计算时间为 $O(L^3)$。这里 T_2 与图像大小无关。

比较本文方法与基于 Ncut 的谱聚类分割方法[8]和基于 Ncut 的阈值分割方法[9]的计算复杂度，表 1 是三种方法的计算时间对比。与另两种方法相比，本文在 T_1 部分多出 $t_1 \times s2$，这是用于颜色量化的计算时间，由于 $s2$ 以求和运算为主，用时很少，相对于 $s1$ 可以忽略。在 T_2 部分，因为通常 L 远小于 N，本文方法的计算复杂度明显低于基于 Ncut 的谱聚类方法，基于Ncut的阈值方法由于是通过多次的

表 1 算法复杂性比较

	T_1	T_2
本文方法	$t_1 \times (s1 + s2)$	$O(L^3)$
文献[9]方法	$t_1 \times s1$	$O(N^3)$
文献[10]方法	$t_1 \times s1$	约 2^{23} 次加法

迭代运算计算阈值,其在 T_2 部分所需时间是个常数[9]。因此,本文方法的计算复杂度低于基于 Ncut 的谱聚类分割方法但高于基于 Ncut 的阈值分割方法。

4　实验结果与分析

为了验证本文方法对文本图像尤其是场景文本图像的分割性能,下面通过一组实验进行测试。实验运行环境为迅驰 CPU 1.73G,内存 1G。实验中,对中文字符图像取 $\sigma_P = 0.2$, $\sigma_F = 0.1$, $\sigma_X = 10$, $r = 10$;对英文字符图像取 $\sigma_P = 0.2$, $\sigma_F = 0.05$, $\sigma_X = 10$, $r = 10$,不同的参数对算法性能有较大影响。实验分别在三个不同的数据集上进行。

4.1　自建数据集实验

该数据集共 300 幅场景图片,包括中文图片和英文图片各 150 幅。中文图片来源于数码相机、因特网、手机以及视频中的图像,英文图片选自文档分析与识别国际会议(ICDAR) 2003 开放图片集,所有样本都是经过定位提取后的图片。同时,采用准确率(precision, p)和召回率(recall, r)对分割结果进行评价, p 和 r 定义如下:

$$p = \frac{com\_Image_0 \cap base\_Image_0}{com\_Image_o} \tag{12}$$

$$r = \frac{com\_Image_0 \cap base\_Image_0}{base\_Image_0} \tag{13}$$

其中, com_Image_0 是分割后目标点即文本区域的像素集合; $base\_Image_0$ 是人工标定的文本区域。同时,参照文献[12],定义综合评定指标(F-measure, f):

$$f = \frac{1}{a/p + (1-a)/r} \tag{14}$$

a 表示准确率和召回率之间的一个相对权重,设 $a = 0.5$。f 越大分割效果越好。

表 2 是本文方法与基于 NCut 阈值方法[9]在上述数据集上的测试结果,数据表明本文方法能获得较高的综合指标,在准确率、召回率和综合指标上均优于文献[9]的方法,反映出本文方法对复杂背景场景文本图像分割的有效性,但是基于 Ncut 的阈值方法在计算时间上优于本文方法。图 1 是实际图像的分割结果,第一行是原始样本,第二行是基于 Ncut 阈值方法的结果,第三行是本文方法的结果,实验结果表明当图像前景与背景对比度不强或者图像有阴影时,阈值方法难以准确区分出前景与背景。

表 2　字符图像的分割结果

所用方法	p(%)	r(%)	f(%)	平均时间(s)
文献[10]方法	83.64	80.11	81.84	1.12
本文方法	84.00	89.2	86.6	2.8

(a) 背景复杂的文本图像的分割结果

(b) 背景复杂且有部分光照图像的分割结果

图 1　复杂背景文本图像分割结果

4.2　DIBCO2009 数据集实验

ICDAR2009 公开测试集 DIBCO2009 数据集包含 10 个文本图像样本(手写体文本和印刷体文本各 5 个)如图 2(a)行所示。作为竞赛的一部分,DIBCO 同时提供四种评价标准:F-Measure, PSNR, negative rate metric, misclassification penalty metric。

(1) F-Measure

$$F\_Measure = \frac{2 \times recall \times precision}{recall + precision} \tag{15}$$

其中 $recall = \frac{TP}{TP + FN}$, $precision = \frac{TP}{TP + FP}$, 这里 TP 表示正确检测出的前景像素,FP 表示错误检测的前景像素,FN 表示漏检的前景像素,TN 表示正确检测出的背景像素。F-measure 越大表示分割效果越好。

（2）PSNR（peak signal to noise ratio）

$$PSNR = 10\log\left(\frac{C^2}{MSE}\right) \tag{16}$$

其中 C 表示前景与背景之间的差异，对灰度图像，C 最大为 255。这里 $MSE = (\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y) - I'(x,y))^2)/MN$，PSNR 描述了两幅图像的相似程度，该值越大表明分割效果越好。

（3）NRM（negative rate metric）

$$NRM = (NR_{FN} + NR_{FP})/2 \tag{17}$$

其中 $NR_{FN} = N_{FN}/(N_{FN} + N_{TP})$ 表示前景漏检率，$NR_{FP} = N_{FP}/(N_{FP} + N_{TN})$ 表示背景漏检率，NRM 表示了前景与背景漏检率的均值，该值越小表示分割效果越好。

（4）MPM（misclassification penalty metric）

$$MPM = (MP_{FN} + MP_{FP})/2 \tag{18}$$

这里 $MP_{FN} = (\sum_{i=1}^{N_{FN}} d_{FN}^{i})/D$，$MP_{FP} = (\sum_{i=1}^{N_{FN}} d_{FP}^{j})/D$，其中 d_{FN}^{i} 表示第 i 个漏检像素到 Ground Truth 轮廓的距离，d_{FP}^{j} 表示第 j 个误检像素到 Ground Truth 轮廓的距离，归一化因子 D 表示 Ground Truth 所有内部像素到轮廓的距离总和。MPM 值越低，说明算法从场景中分割物理目标及识别目标边界效果越好。

DIBCO 2009 共有 43 种方法参加比赛，采用上述四种评价标准对每一种方法进行评价，表 3 列举了部分比赛结果。

表 3 本文方法与 DIBCO 2009 竞赛结果比较

Rank	1	2	3	4
Method	26	14	本文方法	24
F-Measure	91.24	90.06	89.43	89.34
PSNR	18.66	18.23	17.91	17.79
NRM ($\times 10^{-2}$)	4.31	4.75	7.28	5.32
MPM($\times 10^{-3}$)	0.55	0.89	2.15	1.90

本文方法位列第三，获得最好比赛结果的是由新加坡信息通信研究院的 Lu 等人提出的方法 26，该方法通过在局部邻域内检测出的边缘计算阈值，利用局部化边缘信息进行分割。与其相比，本文方法在分割前未做任何预处理，对样本图像直接分割后仅使用均值滤波做简单的去噪处理，显然其还有性能提升的空间，但容易形成不锋利的边界，在背景与前景分离的情况下使得前景区域增大，造成以像素为单位的评价指标下降，NRM 和 MPM 指标偏大也反映了这一点。图 2 中的(b)为本文方法分割结果。

(a) 原始图像

(b) 本文方法的处理结果

(c) Ground truth

图 2 DIBCO 2009 测试数据库样本分割结果

4.3 ICDAR2003 数据集实验

该数据集为 ICDAR 2003 的 Robust Reading Competition[12] 提供的公开样本库，包含 171 个单词，具备光照不均、背景复杂、文字大小、字体不同、低分辨率等场景图像的通常特征。

实验采用与文献[7]相同的条件：选取文献[7]和[3]均使用的 18 幅图像，首先使用 Chen 的方法[13] 进行文本检测，该方法是 ICDAR2003 Robust Reading 竞赛中性能最好的文本定位方法，已开放使用；然后在定位后的文本区域使用本文方法进行分割，将分割后的字符串送入商用 OCR 系统 ABBYY OCR SDK 11 识别，最后使用相对于 Ground Truth 的 Levenshtein 距离评价识别结果，见表 4。表中数据表明若直接使用 OCR 进行识别，有 201 个字符未准确识别，采用本文方法，错误率相对于 Gatos 方法下降约 54%，相对 SMC 方法下降 19%，反映出本文方法良好的字符分割性能。

表 4 识别结果比较

采用方法	Levenshtein 距离
OCR alone	201
Gatos 方法[4]	83
SMC 方法[8]	47
本文方法	38

5　结　论

本文针对场景文本图像的特点,提出了一种基于 Ncut 的谱聚类图像分割方法。该方法结合文本图像的颜色分布特性,根据颜色直方图将图像在颜色空间中量化,得到不同颜色的像素集合,然后在这些像素集合的基础上构造相似矩阵,利用谱聚类的方法进行图像分割。在构造相似矩阵时利用文本像素的颜色分布、特殊纹理及分布局域性定义权值函数。由于采用颜色集合间的相似关系来构建相似矩阵,显著降低相似矩阵的维数,与普通谱聚类图像分割方法相比,计算复杂度大大降低,从而提高了谱聚类方法在图像分割方面的应用能力。与基于 NCut 阈值分割方法相比,本文方法既可以直接处理彩色图像,也可以有效分割复杂的场景文本,大量对比实验表明本文方法在文本图像分割效果上具有良好性能。

参考文献

[1] Lienhart R, Wernicke A. Localizing and segmenting text in images and videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, 12(4): 256-268

[2] Jung K, Kim K I, Jain A K. Text information extraction in images and video: a survey. *Pattern Recognition*, 2004, 37(5): 977-997

[3] B Gatos, I Pratikakis, et al. Text detection in indoor/outdoor scene images. In: Proceedings of the 1st Workshop of Camera-based Document Analysis and Recognition, 2005. 127-132

[4] Hase H, Shinokawa T, Yoneda M, et al. Character string extraction from color documents. *Pattern Recognition*, 34(7), 2001: 1349-1365

[5] Karatzas D, Antonacopoulos A. Text extraction from web images based on a split-and-merge segmentation method using color perception. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. 634-637

[6] Du Y Z, C L Chang, Thouin P D. Unsupervised approach to color video thresholding. *Optical Imaging*, 43(2), 2004: 282-289

[7] Celine M T, Bernard G. Color text extraction with selective metric-based clustering. *Computer Vision and Image Understanding*, 107(1), 2007: 97-107

[8] Shi J B, Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intellence*, 2000, 22(8): 888-905

[9] Tao W B, Jin H, Zhang Y M, et al. Image thresholding using graph cuts. *IEEE Transactions on System, Man, and Cybernetics- Part A: System and Humans*, 2008, 38(5): 1181-1195

[10] K I Laws. Textured Image Segmentation. University of Southern California, 1980

[11] 王向阳,杨红颖,郑宏亮等. 基于视觉权值的分块颜色直方图图像检索算法. 自动化学报, 2010,36(10): 1489-1492

[12] Lucas S M, Panaretos A, Sosa L. ICDAR 2003 robust reading competitions: entries, results, and future directions. *International Journal on Document Analysis and Recognition*, 2005, 7(2): 105-122

[13] Lucas S, Gonzales C J. Web-based deployment of text locating algorithms. In: Proceedings of Camera-based Document Analysis and Recognition, 2005. 101-107

Improved text image segmentation based on spectral clustering

Yin Fang[*,**], Wu Rui[***], Chen Deyun[*], Yu xiaoyang[**]

([*] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080)

([**] College of Measurement-Control Tech & Communications Engineering, Harbin University of Science and Technology, Harbin 150080)

([***] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

Abstract

This paper proposes an improved spectral clustering method for image segmentation based on normalized cut (Ncut). In order to effectively reduce the computational complexity of spectral clustering, the method uses color sets quantized as vertexes of graphs to simplify the weighted graph model. Firstly, the similarity function is established according to the characteristics of text images. And then, the color space is quantified by using the color histogram according to the color distribution of scene images, and the affinity matrix is constructed under the quantized levels. Finally the method uses the spectral clustering to segment images under the Ncut criterion. The experiments conducted with a large number of scene images including a publicly available database from the contest of ICDAR 2009 and 2003 show that the proposed method has the good performance in text image segmentation.

Key words: Image segmentation, text image, spectral clustering, normalized cut, affinity matrix