

基于灰度直方图和谱聚类的文本图像二值化方法

吴锐 黄剑华 唐降龙 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: 在自动文本提取中, 经定位获得的字符区域需二值化后方能有效识别, 由于背景的复杂, 常用的阈值化方法不能有效分割自然环境下的字符图像。该文提出了一种基于谱聚类的图像二值化方法, 该方法利用规范化切痕(Normalized cut, Ncut)作为谱聚类测度, 结合灰度直方图计算相似性矩阵, 并通过实验确定最佳的直方图等级数, 与通常基于像素级相似矩阵相比, 算法的空间复杂度和计算复杂性都大为降低。实验结果表明, 针对自然场景下的字符图像, 该文方法的二值化结果优于常用的阈值分割结果。

关键词: 图像处理; 文本图像; 二值化; 图分割; 谱聚类

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2009)10-2460-05

Method of Text Image Binarization Processing Using Histogram and Spectral Clustering

Wu Rui Huang Jian-hua Tang Xiang-long Liu Jia-feng

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: The located text regions need to be binarized for accurate recognition in automatic textural extraction. Due to the complex backgrounds, traditional thresholding methods can not segment the character image effectively from natural scenes. A novel approach of binarization is proposed for gray images. The proposed algorithm uses the Normalized graph cut(Ncut) as the measure for spectral clustering, and the weighted matrices used in evaluating the graph cuts are based on the gray levels of an image, rather than the commonly used image pixels. Thus, the proposed algorithm requires much smaller spatial costs and much lower computation complexity. Experiments on text images in natural scene show the superior performance of the proposed method compared to the typical thresholding algorithms.

Key words: Image processing; Text image; Binarization processing; Graph partition; Spectral clustering

1 引言

图像中常常包含有丰富的文本信息, 如街道名称、楼牌号、汽车站牌、各种提示板等等, 这些文字信息为理解图像中的内容提供了帮助, 如果计算机能够自动识别图像中的这些文本字符信息, 对于在更高语义层次上进行图像、视频检索和认知极具意义^[1]。例如可以通过摄像机拍摄环境中的路标和建筑物的门牌号码, 利用计算机自动检测、提取、识别这些信息, 并通过语音合成朗读出来, 为盲人指引行进路线。研究计算机在自然环境下的自动文本提取, 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, 因此, 该课题正成为近年来图像处理领域中的一个新的研究热点。

自动文本提取一般包括文本检测、文本定位、

字符二值化和字符识别等过程。在自然场景下, 阴影或者复杂背景会造成传统OCR(Optical Character Recognition)引擎不能有效识别定位后的文字, 通过二值化处理将字符从已定位的文字区域中准确分离出来是十分必要的。目前, 众多的研究人员已在文本检测方法方面做了大量的工作, 提出了很多文本检测与定位的方法^[2-6], 但是专门针对文本提取中的字符分割研究工作还不多, 已有的方法多是利用阈值化实现简单字符图像的分割, 包括全局阈值方法^[7,8]和局部阈值方法^[1,9]。Tsai^[7]采用种子填充与阈值分割相结合的方法提取视频中的文字, Pan^[8]利用Otsu^[10]方法获得全局阈值; Lienhart^[1]在文字亮度与背景亮度之间寻找一个合适的阈值实现二值化, Wu^[9]等采用基于直方图的自动阈值二值化, 分别对各个文本区域实现背景分割。由于阈值化分割的方法一般是利用一定的度量准则进行单调划分, 分割后同一类对象一定大于或者小于某个阈值, 自然场景下文本字符的复杂使得文本区域或背

景区往往不具有某种单调性, 比如同一个字符上可能具有明暗两种笔画, 此时阈值化分割就会失效。

谱聚类算法是基于谱图理论的一类新的聚类算法, 能对任意形状的数据进行划分, 近年来在图像分割^[11,12]、信息检索^[13]以及生物信息学^[14]等领域的应用越来越引起研究者的重视。不足的是, 谱图分割在特征值求解时往往需要较大的空间复杂度和计算复杂度, 这限制了该方法的应用。在文献[15]中, 陶文兵利用图谱测度作为阈值分割的准则实现了一种图像阈值分割方法, 通过计算一个 256×256 的灰度矩阵获得权值矩阵, 简化了计算, 在图像背景与目标有明显对比时可获得较好的分割效果。但是该方法在求取最小 Ncut 值时, 是根据阈值化条件, 求取的局部最优值, 这局限了该方法的应用对象。本文提出一种基于灰度直方图的谱聚类分割方法来实现字符提取, 利用像素灰度计算相似性, 在灰度直方图上构造相似矩阵, 并通过实验确定直方图的理想等级数及相关参数。通过将像素级上的划分转化为灰度等级上的划分, 从而大大减少特征值求解时的计算量和开销, 提高算法的实用性及有效性。

2 基于灰度直方图的谱聚类图像分割算法

2.1 图像分割中的谱图理论

图像中进行图分割的基本思想是将图像看作一个带权图 $G = (V, E)$, 其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 属于同一区域的可能性, 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。分割的目标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较高, 集合间的相似度较低。两个集合 A 和 B 之间的不相似程度, 可用去掉 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)^[11]:

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

寻找图中的最小切(minimum cut)^[11], 即是对图的一个最优划分。Shi和Malik^[12]在此基础上提出了规范化切(Normalized cut, Ncut)实现对图像的分割, 规范化切定义如下:

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

其中 $\text{assoc}(A, V) = \sum_{p \in A, v \in V} w(p, v)$, $\text{assoc}(B, V) =$

$\sum_{q \in B, v \in V} w(q, v)$ 分别表示 A 或 B 到整个顶点集合 V 的关

联度, 此时最小的 Ncut 值对应的划分即为图 G 的最优划分。为求得式(2)中最小的 Ncut 值, 将上述最优

化问题转化为求解式(3)中的特征系统, 其中 D 为 $N \times N$ 对角矩阵, 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 $w(p, q) = w(q, p)$, λ 和 z 分别为相应的特征值和特征向量。

$$D^{-(1/2)}(D - W)D^{-(1/2)}z = \lambda z \quad (3)$$

已经知道, 矩阵 $D^{-(1/2)}(D - W)D^{-(1/2)}$ 第二小的特征值 λ_2 所对应的特征向量 z_2 对应着图 G 的最优划分^[12]。当图像尺寸较大时, 上述特征方程的求解将变得困难, 若采用基于像素级的相似矩阵, 需要求解一个 $N \times N$ 的矩阵特征值 (N 为图像中所有像素之和)。

2.2 算法设计

令 $V = \{(i, j) : i=0, 1, \dots, n_h - 1; j=0, 1, \dots, n_w - 1\}$, $H = \{H_0, H_1, \dots, H_L\}$, $LL = \{0, 1, \dots, L\}$, 其中 n_h 和 n_w 分别为图像的高度和宽度, H 代表灰度直方图, 令 $f(i, j)$ 为图像在像素 (i, j) 处的灰度值, 则 V, H 和 $f(i, j)$ 满足如下条件: $(i, j) \in H_l, l \in \{0, 1, \dots, L\}, \forall (i, j) \in V; H_l = \{(i, j) : f(i, j) = l, (i, j) \in V\}, l \in LL; \bigcup_{l=0}^L H_l$

$= V, H_u \cap H_v = \emptyset, u \neq v, \forall u, v \in LL$, 利用像素的灰度值和它们之间的空间位置, 定义顶点 p 和 q 连接边的权值为

$$w(p, q) = \begin{cases} \exp\left\{\frac{-\|F(p) - F(q)\|_2^2}{\sigma_I} + \frac{-\|X(p) - X(q)\|_2^2}{\sigma_X}\right\}, \\ \quad \|X(p) - X(q)\|_2 < r \\ 0, \quad \text{其他} \end{cases} \quad (4)$$

其中, $F(p)$ 为顶点 p 的灰度级, $X(p)$ 是点 p 的空间位置。 σ_I 和 σ_X 为尺度因子, 分别用来调节顶点间的灰度差异和空间位置差异, r 控制 p 和 q 之间的顶点个数。

于是, 图像对应的图 $G = (V, E)$ 的一个二划分为 $V = \{A, B\}$, 其中 $A = \bigcup_{k \in L_A} H_k, B = \bigcup_{k \in L_B} H_k$, 同时有 $L_A \cap L_B = \emptyset, L_A \cup L_B = LL$ 。令 $\text{cut}(H_u, H_v) = \sum_{p \in H_u, q \in H_v} w(p, q)$ 为 H_u 中所有点(灰度级为 u)与 H_v 中所有点(灰度级为 v)间总的连接权值之和, 于是上面各式可转化为

$$\text{cut}(A, B) = \sum_{u \in L_A} \sum_{v \in L_B} \text{cut}(H_u, H_v)$$

$$\text{asso}(A, A) = \sum_{u \in L_A} \sum_{v \in L_A} \text{cut}(H_u, H_v)$$

$$\text{asso}(B, B) = \sum_{u \in L_B} \sum_{v \in L_B} \text{cut}(H_u, H_v)$$

又因为有下列式成立:

$$\text{asso}(A, V) = \text{asso}(A, A) + \text{cut}(A, B)$$

$$\text{asso}(B, V) = \text{asso}(B, B) + \text{cut}(A, B)$$

因而, 式(2)可写成如下形式

$$\text{Ncut}(A, B) = \frac{\text{cut}(A, B)}{\text{asso}(A, A) + \text{cut}(A, B)} + \frac{\text{cut}(A, B)}{\text{asso}(B, B) + \text{cut}(A, B)} \quad (5)$$

对给定的一幅图像, 通过计算其对应图中所有顶点间的权值可以构建一个基于灰度直方图的权值矩阵 M , 其中矩阵 $M = [m_{u,v}]$ 为 $L \times L$ 的对称矩阵, L 为直方图的灰度等级数, 且 $m_{u,v} = \text{cut}(H_u, H_v)$, $m_{u,v} = m_{v,u}$ 。于是, 将 M 作为相似度矩阵, 运用谱聚类方法进行图像分割的算法流程如图 1 所示。

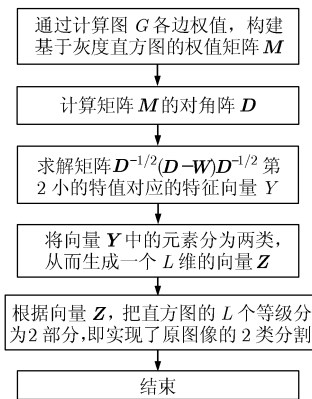


图 1 基于灰度直方图的谱聚类图像分割算法流程

在上述方法中, 相似度矩阵 M 的大小 ($L \times L$) 取决于直方图的等级 L , 而不是图像中像素的个数 $N (N \times N)$; 同时, 算法中只需求解一个 $L \times L$ 维矩阵的特征系统, 而不再是 $N \times N$ 维的, 通常 L 比 N 小得多, 且大小固定。因而算法的计算复杂度和空间复杂度都大为降低。

2.3 相关参数的选择

在本文方法中, 直方图等级数 L 的大小决定了特征系统的复杂程度。另一方面, 式(4)中的连接边权值定义是构造相似矩阵的基础, 尺度因子 σ_I 为灰度高斯函数的标准方差, σ_X 为空间距离高斯函数的标准方差, 典型取值范围在图像灰度函数或空间距离函数最大值的 10%–20%^[12], 可根据实验确定。其中参数 r 的选取对计算相似矩阵有直接的影响, 一般来说, r 值越大, 越能全面地反映像素间的相似关系, 但计算量也越大, 取较小的 r 值, 可以很好控制相似矩阵的稀疏程度, 因此通常是在能够接受

的分割结果的前提下, 使 r 值取得尽量小, 通过实验结果对比, 取 $\sigma_I=50$, $\sigma_X=5$, $r=5$ 较为合适。

为了测试不同灰度等级 L 对文本图像的分割效果, 我们选取了 50 幅经定位分割后包含有字符文本的灰度图像进行实验。同时, 采用正确率 p 和召回率 r 对分割结果进行评价, 正确率表示正确分割到的字符区域与实际的字符区域的匹配程度, 召回率表示对实际字符区域的检出率。 p 和 r 定义如下:

$$p = \frac{\text{com_Im age}_O \cap \text{base_Im age}_O}{\text{com_Im age}_O},$$

$$r = \frac{\text{com_Im age}_O \cap \text{base_Im age}_O}{\text{base_Im age}_O} \quad (6)$$

其中, com_Im age_O 是分割得到的文本区域的像素集合, base_Im age_O 是人工标定出的文本区域。同时, 参照文献[16], 定义综合评定指标 f :

$$f = \frac{1}{a/p + (1-a)/r} \quad (7)$$

式(7)中 a 表示准确率和召回率之间的一个相对权重, 设 $a = 0.5$ 。图 2 是 50 幅图像在不同灰度等级下的实验结果, 图 3 是相应的处理一幅图片的平均时间, 实验数据表明, 在 $L=100$, $r=5$ 时, 能获得较好的文本分割效果。

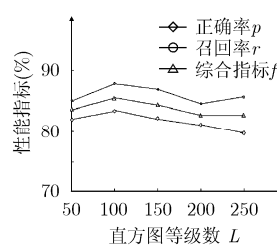


图 2 不同参数下的实验结果

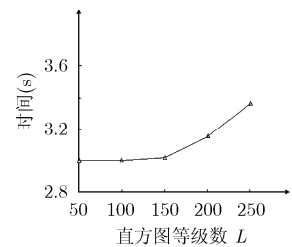


图 3 算法执行时间随灰度等级 L 变化的曲线

3 实验结果

我们通过一系列的实验来测试本文方法的性能, 在下述实验中, 式(4)中各参数设置为 $\sigma_I=50$, $\sigma_X=5$, $r=5$, 直方图的灰度等级 $L=100$, 运行环境为 PIV1.73 G, 内存 512 M。为了对比本文方法的分割性能, 我们选取了基于最大类间方差的 Otsu 方法^[10]与基于 Ncut 的阈值分割方法^[15]与本文方法进行二值化比较。前者是一种经典的阈值化方法, 已广泛应用于文本处理中; 后者则是一种基于谱图准则的阈值化方法, 二者均具有很好的代表性。

图 4 中字符前景与非字符的背景对比明显, 且字符区域灰度值基本一致, 3 种方法都取得较好效果。图 5 中的汉字图像尽管有阴影, 但前景区域的一致性较好, 基于 Ncut 的阈值方法和本文方法均能获得很好的分割效果, 优于传统的 Otsu 方法, 这也



图 4 前景与背景边界清晰图像的分割结果



图 5 有阴影但前景清晰图像的分割结果

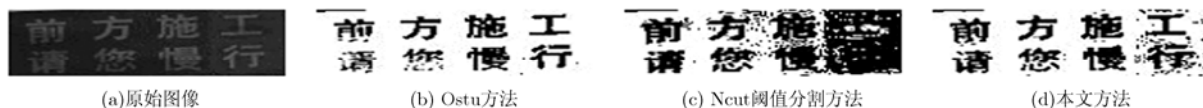


图 6 有阴影图像的分割结果



图 7 前景与背景边界模糊图像的分割结果

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前景与背景边缘较模糊时, 本文方法能获得更为理想的效果, 如图 6、图 7 所示, 阈值化分割的结果则会使得此类图像字符笔画断裂(如图 6(b))或者字符与背景不可分(如图 6(c))。

在自然环境下, 图像经常会由于光照致使同一区域的像素灰度差异较大, 一般的阈值分割方法难以准确提取出字符区域, 如图 8 中的字母 ‘B’, 本文方法能够将字母 ‘B’ 上的部分黑点由于反光变成白点的点正确分割, 这是一般阈值化方法难以达到的。

为了验证本文方法对不同环境下图像的分割性能, 我们选取了 300 幅来自数码相机、手机、因特网上的自然场景下的字符图像对 3 种方法进行综合测试, 图像中既有中文字符也有英文字符, 图像大小不低于 20000 像素, 测试结果如表 1 所示。召回率高表明本文方法能够尽量多地获得字符区域, 但由于容易产生离散噪声点, 因而准确率略低, 从综合评定指标来看, 本文方法优于另两种方法。从处



图 8 光照不均图像的分割结果

表 1 3 种方法对字符图像的分割结果

所用方法	p (%)	r (%)	f (%)	平均时间(s)
Ostu 方法	80.89	83.61	82.23	0.266
基于 Neut 的 阈值方法	83.64	80.11	81.84	2.12
本文方法	83.01	88	85.43	3

理一幅图像的平均时间来看, Ostu 方法最为快速, 另两种方法慢得多, 这也表明基于图谱方法的实时性仍有待提高。

4 结论

自然环境下文本图像背景复杂, 常规阈值分割方法往往无法有效分割; 基于谱聚类的图像分割方法利用图分割理论可以有效地实现图像分割, 但过高的计算复杂度和空间复杂度使其在处理大图像时不能满足实际需求。本文提出的基于灰度直方图的谱分割方法将谱聚类对象从像素点转化到灰度级, 使算法复杂度得到有效控制。实验结果表明, 该方法针对字符图像分割具有良好的性能, 优于一般的阈值化分割方法。下一步的工作将针对权函数的参数获取做深入研究, 以拓展本文方法的应用范围。

参考文献

[1] Lienhart R and 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] Mariano V Y and Kasturi R. Locating uniform-colored text in video frames. Proc. of Int'l Conference on Pattern Recognition, Barcelona, Spain, 2000, 4: 539-542.
- [3] Chen D, Odobez J M, and Bourlard H. Text detection and recognition in images and video frames. *Pattern Recognition*, 2004, 37(3): 595-608.
- [4] Zhong Yu, Zhang Hong-jiang, and Jain A K. Automatic caption localization in compressed video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(4): 385-392.
- [5] Chen Xi-lin, Yang Jie, Zhang Jing, and Waibel A. Automatic detection and recognition of signs from natural scenes. *IEEE Transactions on Image Processing*, 2004, 13(1): 87-99.
- [6] Chen Xiang-rong and Yuille A L. Detecting and reading text in natural scenes. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 2004: 366-373.
- [7] Tsai T H and Chen Y C. A comprehensive motion videotext detection localization and extraction method. Proc. of IEEE Int'l Conference on Data Engineering Workshop, Istanbul, Turkey, 2007: 113-116.
- [8] Pan W M, Bui T D, and Suen CY. Text segmentation from complex background using sparse representations. Proc. of Int'l Conference on Document Analysis Recognition, Curitiba, Brazil, 2007: 412-416.
- [9] Wu V, Manmatha R, and Riseman E M. Text finder: An automatic system to detect and recognize text in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(11): 1224-1229.
- [10] Otsu N. A threshold selection method from grey-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 1979, 9(1): 62-66.
- [11] Wu Z Y and Leahy R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, 15(11): 1101-1113.
- [12] Shi Jian-bo and Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905.
- [13] He X, Cai D, and Wen J R, *et al.* Clustering and searching WWW images using link and page layout analysis. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2007, 3(2): Article No. 10.
- [14] Higham D J, Kalna G, and Kibble M. Spectral clustering and its use in bioinformatics. *Journal of Computational and Applied Mathematics*, 2007, 204(1): 25-37.
- [15] 陶文兵, 金海. 一种新的基于图谱理论的图像阈值分割方法. *计算机学报*, 2007, 30(1): 110-118.
Tao Wen-bing and Jin Hai. A new image thresholding method based on graph spectral theory. *Chinese Journal of Computers*, 2007, 30(1): 110-118.
- [16] Lucas S M, Panaretos A, and Sosa L, *et al.* ICDAR 2003 robust reading competition. Proc. of 7th Int'l Conference on Document Analysis and Recognition, Scotland, 2003: 682-687.
- 吴 锐: 男, 1976年生, 讲师, 研究方向为文本提取、图像处理、字符识别.
- 黄剑华: 男, 1967年生, 研究员, 研究方向为模式识别、文本提取、图像处理.
- 唐降龙: 男, 1960年生, 教授, 博士生导师, 研究方向为模式识别、计算机视觉、人工智能.
- 刘家锋: 男, 1968年生, 副教授, 研究方向为人工智能、模式识别、中文信息处理.