图文信息技术

自然场景图像的字符识别方法

李颖，刘菊华，易尧华
（武汉大学，武汉 430072）

摘要：目的 基于大津算法（Otsu 算法）对图像进行分割，利用光学字符识别方法对自然场景图像中的英文字符进行识别。**方法** 首先用分块 Otsu 算法对图像进行初步的二值化，然后通过对二值化结果的分析，把原始的输入图片分割成单个字符的子图，再对各子图重新用 Otsu 算法进行二值化，最后对最终得到的二值化结果进行识别，再结合之前得到的每幅图的字符数量信息和词典信息，对识别结果进行修正，得到最终的识别结果。**结果** 在 ICDAR2013 数据集上测试文中算法，单词正确识别率为 46.03%，总编辑距离为 474.5。**结论** 文中提出的以 Otsu 为基础的分块识别算法，能够更好地分割复杂背景图像的背景和文本，同时结合词典信息对识别结果进行了修正，改善了识别效果。

关键词：场景图像；字符识别；Otsu 算法；词典
中图分类号：TP391.4 **文献标识码：**A **文章编号：**1001-3563(2018)05-0168-05
DOI：10.19554/j.cnki.1001-3563.2018.05.032

Character Recognition Method in Natural Scene Images

LI Ying, *LIU Ju-hua*, *YI Yao-hua*
(Wuhan University, Wuhan 430072, China)

ABSTRACT: The work aims to segment the image based on the Otsu algorithm and then recognize the English characters in the natural scene images with the method of optical character recognition. First, preliminary binarization of the image was carried with the block Ostu method. Then, the original input image was segmented into sub-graphs of single character after analyzing the binarization results, and all the sub-graphs were binarized again with the Otsu algorithm. Last, the finally obtained binarization results were recognized. Then, the recognition results were modified in combination with the previously obtained information on the number of characters and the dictionary in each image, so as to obtain the final recognition results. The proposed algorithm was tested on the ICDAR2013 dataset. The correct recognition rate of words was 46.03% and the total editing distance was 474.5. The proposed block recognition algorithm based on the Otsu method can better segment the background and the text in complex background images and improve the recognition effect combined with the dictionary information used to modify the recognition results.

KEY WORDS: scene image; character recognition; Otsu algorithm; dictionary

在观察自然场景图像时，人眼可以快速地定位并识别出文本信息，但计算机要做到同人眼一样很困难。研究人员一直致力于让计算机也能快速地定位并识别出自然场景图像中的文本信息，在这一课题中，字符识别是一个十分重要的步骤。在获取自然场景图像时，由于背景物体、光线、阴影、拍摄角度引起的图片背景千变万化，被拍摄的图片中包含的文字大小、颜色、书写风格各不相同等因素都为字符识别的实现增加了相当的难度[1—7]。场景文本识别包括 3 个步骤：文本检测、文本分割、字符识别。文中主要研究如何从已经成功定位的图像中提取出字符信息。

现有的字符识别方法根据原理的不同可以分成 3

收稿日期：2017-06-26
基金项目：国家自然科学基金青年基金（61601335）；湖北省自然科学基金（2016CFB157）
作者简介：李颖（1994—），女，武汉大学硕士生，主攻图像处理、模式识别。
通信作者：刘菊华（1983—），男，博士后，武汉大学讲师，主要研究方向为图像色彩管理、彩色数字成像与模式识别。

类：神经网络方法、模板匹配方法[8]、传统 OCR 方法[9]。神经网络方法，主要是通过海量的字符样本来训练字符分类器，然后将待识别图像分割成至多包含一个字符的图像，再把这些图像输入到训练好的分类器中判别属于哪一个字符即可。例如，Bissacco 等构建了一个基于深度神经网络的 PhotoOCR 系统，该系统使用了数以百万计的训练样本和分布式语言模型，并采用机器学习方法提高分类器性能[10]。模板匹配法，首先需要构建标准模板库，然后从待识别的字符图像区域中提取出特征量与模板相应的特征量进行比较，再将待识别的字符区域归于相应的字符类中，这种方法适用于字体规范的字符，对于字符图像的缺损、污迹有较强的抗干扰能力，但是对于字符的旋转、扭曲和变形抵抗能力不强，一般应用在车牌识别中，在场景文本识别这一课题中很少被使用。传统 OCR 方法进行字符识别，则需要处理待识别的自然场景图像，使其图像质量接近扫描文档图像，再使用 OCR 方法对已处理的图像进行字符识别就可以得到较好的结果。比如，Neumann 提出一种基于 OCR 技术的端对端的文本定位和识别方法。首先，挑选出最有可能是字符的极值区域，然后把极值区域聚集成文本行，最后当一个文本行的所有字符都都确定之后，挑选出一个可能性最大的字符序列[11]。

为了将传统 OCR 技术应用在识别算法中，文中对大津算法做了一些改进，改善了场景图像的二值化结果，进而改善了字符识别的效果。文中使用大津算法来进行二值化处理，是因为根据大津算法的原理可以把自然场景图像分成文本像素和背景像素 2 类，得到较好的二值化效果，有助于后续的字符识别步骤。

1　理论知识

1.1　大津算法

大津算法是由日本学者大津博士提出的一种基于概率统计学原理的自适应阈值分割算法，又称为最大类间算法，它是以图像中每个可能的灰度值作为阈值将图像的像素分成 2 类，并计算其类间方差，选择使类间方差最大的那个灰度值作为将图像二值化的阈值[12—13]。

假设给定图像的灰度值在[1,2...L]范围内变化，用阈值 k 把图像像素分成 2 类 C_0 和 C_1（分别代表背景和前景，或者相反），C_0 是灰度值在[1...k]范围内的像素，C_1 是灰度值在[k+1...L]范围内的像素。那么背景类 C_0 和前景类 C_1 这 2 类出现的概率和平均灰度值分别为 $\omega_0, \omega_1, \mu_0, \mu_1$，计算公式为：

$$\omega_0 = Pr(C_0) = \sum_{i=1}^{k} p_i = \omega(k) \tag{1}$$

$$\omega_1 = Pr(C_1) = \sum_{i=k+1}^{L} p_i = 1-\omega(k) \tag{2}$$

$$\mu_0 = \sum_{i=1}^{k} i\,Pr(i|C_0) = \sum_{i=1}^{k} ip_i/\omega_0 = \mu(k)/\omega(k) \tag{3}$$

$$\mu_1 = \sum_{i=k+1}^{L} i\,Pr(i|C_1) = \sum_{i=k+1}^{L} ip_i/\omega_1 = \frac{\mu_T - \mu(k)}{1-\omega(k)} \tag{4}$$

计算灰度值从 0 到 k 的像素点出现的平均灰度值见式（5），整幅图像的平均灰度值见式（6），类间方差见式（7）。使类间方差最大的最优阈值 k*可以用来进行图像的二值化[14]。

$$\mu(\mathrm{k}) = \sum_{i=1}^{k} ip_i \tag{5}$$

$$\mu_{\mathrm{T}} = \mu(L) = \sum_{i=1}^{L} ip_i \tag{6}$$

$$\sigma_{\mathrm{B}}^2 = \omega_0(\mu_0 - \mu_{\mathrm{T}})^2 + \omega_1(\mu_1 - \mu_{\mathrm{T}})^2 = \omega_0\omega_1(\mu_1 - \mu_0)^2 \tag{7}$$

1.2　分块大津算法

目标或背景灰度值分布不均是影响大津算法二值化效果的最重要的因素。在整幅图像里目标与背景各自内部的灰度变化范围很大，如果将图像分成更小块，在每个小块内目标或背景的灰度变化范围会更小，更容易将所有像素分成目标或背景 2 类，因此，文方法考虑将图像划分成合适的小块，使每一小块里都包含目标和背景像素，且小块内目标和背景 2 类像素里灰度都变化不大。在每一个小块里，使用大津算法分别进行二值化。对图像使用分块大津算法得到的二值化效果会优于对整幅图像直接使用大津算法得到的效果。

2　文中算法

文中方法首先对图像进行分块 Otsu 二值化，然后通过分析得到可以进行切分的位置信息，把原灰度图像切分成仅有单个字符的子图像，对单个字符的子图像分别重新进行二值化，再把二值化成功的子图像拼接起来输入到 Tesseract OCR 中进行识别。文中提出的方法可以分成 3 个步骤：分块 Otsu 二值化、分割和识别。文中提出的算法的整个流程见图 1。

2.1　分块 Otsu 二值化

将输入的彩色图像灰度化，得到灰度图像，然后使用分块大津算法对灰度图像初步进行二值化，得到一个可以看到字符轮廓的二值图像，以便后续步骤找到将二值图像切分成单字符子图像的切分位置。通过大量实验发现，将图像切分成 2 行 3 列的小块时，二值化的效果最好。

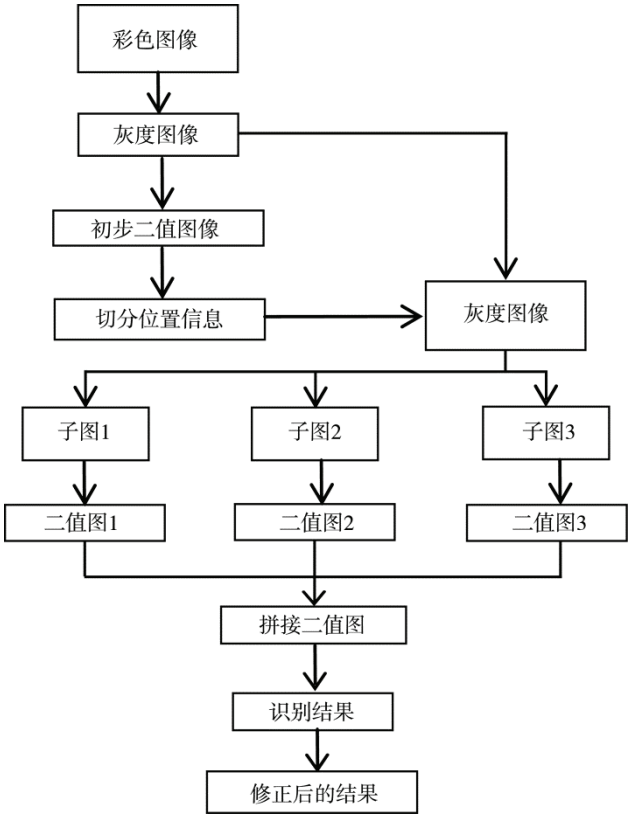


图 1　文本识别方法流程

Fig.1 The process of the text recognition method

2.2　分割

经过分块 Otsu 二值化后会得到初步二值图像，接下来需要在这个二值图像中找到切分位置。首先，检查二值图像边界像素点的灰度值。如果灰度值等于 255，则对图片进行反相处理，使图片呈现出黑底白字。再分别统计每列白色像素点所在的行数，若前后 2 列白色像素点的行数是连续的，则认为这 2 列的笔画是连着的，均属于同一个字符；若前后 2 列白色像素点的行数不连续，则可认为前后 2 列分别属于 2 个字符，即找到了将 2 个字符分离开的切分位置。然后根据找到的切分位置，把原灰度图像切分成几个子图，再对几个子图分别进行二值化，将二值子图按照原先切分的顺序拼接起来。在这整个过程中各子图面积之和等于原灰度图面积，也等于拼接二值图面积。切分好并进行二值化后的子图像以及最终输入 Tesseract OCR 中的拼接好的二值图像见图 2。拼接二值图像就是原彩色图像最终的分割结果，可直接输入 Tesseract OCR 中获得识别结果。

2.3　识别

得到分割结果后，可以通过 Tesseract OCR 获得识别结果。通过实验发现，整张图片进行识别的效果要比分割的单字符图片分别进行识别再组合的识别效果要好，因此，文中方法可以找到灰度图像的字符

a 原彩色图像

b 子图像

c 拼接二值图像

图 2　图像的分割结果

Fig.2 The segmentation result of the image

分割处之后分割灰度图像，分别进行二值化后，将其拼接起来，整体输入到 OCR 中进行识别。

由于场景图像的光照不均等因素，二值图像可能未完全将背景与字符分离，因此，OCR 识别得到的结果可能会出现乱码或者点、短横线等奇怪的字符。这样的结果不是希望得到的，对这种类型的结果，需要进行一些处理，将乱码或奇怪字符从识别结果中删除。另外，还有将字母识别成相似的数字或者将数字识别成字母的情况。出现这种情况，需要结合识别结果中其他字母或数字的数量信息来判断这个字符到底应该是字母还是数字，然后对识别结果进行相应的修正。经过修正后的字符串结果为最终的识别结果。

3　结果与讨论

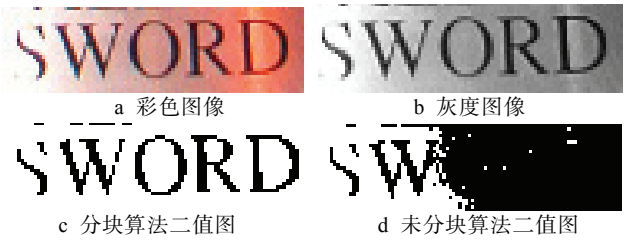
在 ICDAR2013 数据集中测试文中提出的方法，得到的总编辑距离为 474.5，优于 Lukas Neumann 提出的 TextSpotter 方法和直接用 ABBY OCR 进行识别的 Baseline 方法。不同方法结果的比较见表 1。编辑距离是通过计算一个字符串转变成另一个字符串需要变化的最小次数来量化 2 个字符串之间的相似程度的指标。总编辑距离是 ICDAR2013 数据集中 1095 幅图像的识别结果与正确结果之间的编辑距离之和。实验结果里的总编辑距离越小，说明在这个数据集里的单个字符的正确识别率越高，识别效果越好。单词正确识别率是完全正确识别出的单词数目在数据集中图片总数中所占的比例，单词正确识别率越大，识别效果越好。在字符识别方法的比较中，一般优先考虑总编辑距离，在单词正确识别率相差不大时，总编辑距离越小，认为该方法越好。文献[11]中的 TextSpotter 方法的总编辑距离不比文中方法高很多，但是识别率却低很多。经过分析，发现在 TextSpotter 方法中单个字符的识别率还不错，就整个单词而言，不能很好地把整个单词正确识别，但是文中方法并没有出现这个问题。文中的实验方案具有普适性，与不同算

法进行比较的结果是数据集中 1095 幅图片的综合结果。每一种算法都对 1095 幅图片做处理，综合起来计算出平均单词正确识别率和总编辑距离，因此文中算法是在整体上优于 Baseline 和 TextSpotter 方法的。换而言之，与这 2 种方法比较时，对于同一幅原稿，在大多数情况下，文中方法都更好，更具有普适性。文中方法使用 C++实现，在 64 位 Windows 7 操作系统，处理器为 Intel Core i3 3.3GHz 的 PC 上输入大小为 282×82 像素的图片进行字符识别，运行时间为 0.171 s。

表 1　文中方法与其他方法的比较
Tab.1 Comparison of the proposed method and other methods

方法	总编辑距离	单词正确识别率/%
Field's Method	390.6	52.33
文中	474.5	46.03
Baseline	517.9	46.58
TextSpotter	597.3	28.13

分块思想是文中算法的重点之一，在整个处理过程有 2 处应用了分块思想：在最初二值化的时候，使用了 2 行 3 列的分块，而不是直接对整幅图片进行二值化；将图片尽量分割成单个字符的图像再重新进行字符像素和背景像素的分类。使用分块算法是因为在自然场景图像中，同一幅图像中不同区域的背景像素之间的差异会很大，如果进行二值化的图像块过大，很难将字符像素和背景像素完全分开。图 3a 是原图，图 3b 是灰度图，在这 2 幅图像中可以很明显地观察到，同一幅图像中左右两端的背景像素（灰度值）的差异非常大。图 3c 是文中算法二值化的结果，图 3d 是不使用分块算法得到的二值化结果，从结果的对比中，可以清晰地看到在自然场景图像背景变化很大时文中算法的优越性。



a 彩色图像　b 灰度图像
c 分块算法二值图　d 未分块算法二值图

图 3　分块算法与未分块算法的比较
Fig.3 Comparison of block algorithm and un-block algorithm

图像锐化和词典的使用是影响识别结果的 2 个重要因素。锐化在一定程度上能改善某些图片的识别效果，也会使某些图片识别效果变差。如果一幅图像的边缘不够清晰，那么对图像进行锐化处理后可以改进识别效果，见图 4。通过人眼观察，图 4 的正确识别结果应该是 e。图 4c 是没有经过锐化处理得到的二值图分割结果，将其输入到文中的字符识别程序中，得到的识别结果为 C。图 4d 是经过锐化处理得到的二值图分割结果，将其输入到文中的字符识别程序中，得到的识别结果为 e。通过图 4 可知，锐化操作可以有效改进识别结果。因为对分割结果的锐化处理既会改善一些图片的识别结果又会使另一些图片的识别结果变差，所以文中算法分别对锐化后的图片和锐化前的图片进行字符识别，在词典中对 2 个结果进行比较，选择更好的那个结果作为最终的结果。

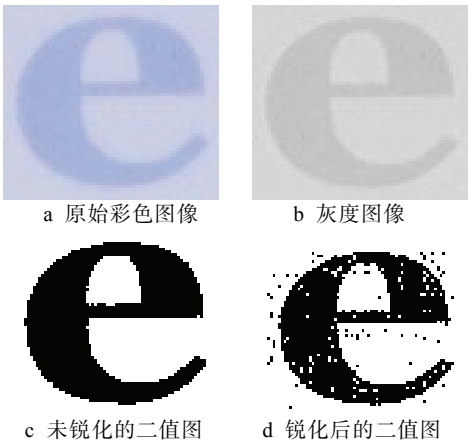

a 原始彩色图像　b 灰度图像
c 未锐化的二值图　d 锐化后的二值图

图 4　锐化算法与未锐化算法的比较
Fig.4 Comparison of sharp algorithm and un-sharp algorithm

词典是另一个影响识别效果的重要因素。使用词典来修正识别结果是因为在场景图像中出现的单词一般是常见的。如果一个单词中有大部分字符都是正确识别的，只有极少数字符没有识别出来，那么此时用词典对识别结果进行修正，有很大的概率可以得到正确的识别结果，因此，在算法中设定如果识别结果不是词典中的单词，那么这个结果应该被修正以接近正确的结果。其前提是词典必须足够大，大到至少包含在场景图像中出现的所有单词，只有满足这个条件，才能保证对识别结果的修正是使其更接近正确的结果，而不是离正确结果越来越远。

4　结语

提出了结合词典的基于 Otsu 和 OCR 的场景文本识别方法，将图片切分成单字符图像进行二值化，再拼接起来将包含整个单词的二值图片输入 Tesseract OCR 中进行识别。该方法的实现较简单，且能达到较好的识别效果。

参考文献：

[1] KUMAR D, RAMAKRISHNAN A G. Power-law Transformation for Enhanced Recognition of Born-digital Word Images[C]// 2012 International Confer-

ence on Signal Processing and Communications (SPCOM), 2012: 1—5.

[2] KUMAR D, PRASAD M N A, RAMAKRISHNAN A G. Nonlinear Enhancement and Selection of Plane for Optimal Segmentation and Recognition of Scene Word Images[J]. Proc SPIE 8658, Document Recognition and Retrieval, 2013, 6: 8658—8659.

[3] FIELD J L, LEARNED-MILLER E G, SMITH D A. Using a Probabilistic Syllable Model to Improve[C]// 2013 12th International Conference on Document Analysis and Recognition, 2013: 897—901.

[4] FIELD J L, LEARNED-MILLER E G. Improving Open-Vocabulary Scene Text Recognition[C]// 2013 12th International Conference on Document Analysis and Recognition, 2013: 604—608.

[5] 黄敏. 场景文字识别方法研究及其软件实现[D]. 成都: 电子科技大学, 2014.
HUANG Min. Research on Scene Text Recognition Method and Software Realization[D]. Chengdu: School of Electronic Engineering, 2014.

[6] 姚聪. 自然图像中文字检测与识别研究[D]. 武汉: 华中科技大学, 2014.
YAO Cong. Research on Text Detection and Recognition in Natural Images[D]. Wuhan: Huazhong University of Science and Technology, 2014.

[7] 尹芳. 场景文本识别关键技术研究[D]. 哈尔滨: 哈尔滨理工大学, 2012.
YIN Fang. Study on Key Technologies of Scene Text Recognition[D]. Harbin: Harbin University of Science and Technology, 2012.

[8] 李新良. 基于模板匹配法的字符识别算法研究[J]. 计算技术与自动化, 2012(2): 90—93.
LI Xin-liang. The Research of Character Recognition Algorithm in Template Matching Method[J]. Computing Technology and Automation, 2012(2): 90—93.

[9] MILYAEV S, BARINOVA O, NOVIKOVA T, et al. Image Binarization for End-to-end Text Understanding in Natural Images[C]// 2013 12th International Conference on Document Analysis and Recognition, 2013: 128—132.

[10] BISSACCO A, CUMMINS M, NETZER Y, et al. PhotoOCR: Reading Text in Uncontrolled Conditions[C]// 2013 IEEE International Conference on Computer Vision, 2013: 785—792.

[11] NEUMANN L, MATAS J. Real-time Lexicon-free Scene Text Localization and Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(9): 1872—1885.

[12] 覃晓, 元昌安, 邓育林, 等. 一种改进的 Ostu 图像分割法[J]. 山西大学学报(自然科学版), 2013, 36(4): 530—534.
QIN Xiao, YUAN Chang-an, DENG Yu-lin, et al. A Improved Ostu Image Segmentation Method[J]. Journal of Shanxi University(Nat Sci Ed), 2013, 36(4): 530—534.

[13] 李了了, 邓善熙, 丁兴号. 基于大津法的图像分块二值化算法[J]. 微计算机信息, 2015(21): 76—77.
LI Liao-liao, DENG Shan-xi, DING Xing-hao. Binarization Algorithm Based on Image Partition Derived from Da-Jing Method[J]. Microcomputer Information, 2015(21): 76—77.

[14] OTSU N A. A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Trans on System Man and Cybernetics, 1979, 9(1): 62—66.