

# 自然场景下基于区域检测的文字识别算法

陈梓洋,王宇飞,钱侃,张超,孙知信

(南京邮电大学 物联网学院,江苏 南京 210003)

**摘要:**由于自然场景的图像具有复杂性和不确定性,导致自然场景下文字识别相比于文本识别更加复杂和困难,利用现有识别算法对自然场景图像进行识别,其识别效率较低,识别效果不理想。为提高识别率,文中首先利用分水岭算法对原图像进行初处理,再对预处理后的图像进行特征提取,划分区域,最后利用 ISODATA 算法进行第二次处理,最终对图像中的文字进行识别。对提出的算法进行仿真和实验,实验结果表明,采用该算法后,自然场景下文字的识别效果明显,识别率高。

**关键词:**分水岭算法;ISODATA 算法;基于区域的检测算法;自然场景;文字识别

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2015)07-0230-04

doi:10.3969/j.issn.1673-629X.2015.07.052

## Character Recognition Algorithm Based on Region Detection in Natural Scene

CHEN Zi-yang, WANG Yu-fei, QIAN Kan, ZHANG Chao, SUN Zhi-xin

(College of Internet of Things, Nanjing University of Posts and Telecommunications,  
Nanjing 210003, China)

**Abstract:** Because the natural scene text has the characteristic such as the uncertainty and complexity, the natural scene is becoming difficult and important in character recognition. By using the existing recognition algorithm for identification of natural scene images, the recognition rate is low and the recognition effect is not ideal. In order to improve the recognition rate, firstly use watershed algorithm to finish the first original image processing, then extract feature and partition the image after preprocessing, finally utilize ISODATA algorithm to process second times and identify the text in the image. By simulating and experimenting on the proposed algorithm, the experimental results show this algorithm can recognize the text in natural scene efficiently with the high recognition rate.

**Key words:** watershed algorithm; ISODATA algorithm; detection algorithm based on region; natural scene; character recognition

## 0 引言

OCR 即文字识别,是利用电子仪器对打印在纸上的文字进行扫描,然后用字符识别方法将扫描到的文字图形在计算机终端上进行文字显示的过程。就是对文本资料先进行扫描,然后对扫描的资料进行计算及分析,最后重新获取被扫描的文字的过程<sup>[1]</sup>。

自然图像是对外在客观世界的客观表现,对于同一事物必将获得相同或几近相同的图像结果,所以自然图像具有特定的规律与性质<sup>[2]</sup>:

(1) 自相似性和尺度不变性;

(2) 非高斯性;

(3) 边缘主导性和高维奇异性。

自然场景中的文字往往携带着非常重要的信息,它能够被用来描述该图像的内容。自动地获取图像中的文字信息可以帮助人们更有效地理解图像和对图像进行存储、压缩、检索等处理。在很多具体应用中,例如车牌定位、视频字幕提取等,最终的目标就是文字信息的获取。基于上述原因,图像中文字信息的自动提取已经成为计算机视觉领域中一个重要的研究方向。

正是因为自然场景下的图像具有以上三种特点,导致自然场景下的文字识别相比于文本识别更加复杂

收稿日期:2014-04-20

修回日期:2014-08-05

网络出版时间:2015-06-23

基金项目:国家自然科学基金资助项目(60973140,61170276,61373135);江苏省产学研项目(BY2013011);江苏省科技型企业创新基金项目(BC2013027);江苏省高校自然科学研究重大项目(12KJA520003)

作者简介:陈梓洋(1992-),男,硕士,研究方向为图像处理和模式识别;孙知信,博士,教授,研究方向为计算机网络与安全、计算机仿真、软件工程。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150623.1031.023.html>

和困难。其中文字定位<sup>[3]</sup>和文字识别,又分别是文字识别系统中的关键技术难点。

为提高自然场景下文字识别的准确率,文中主要研究自然场景下基于区域的文字检测与识别算法,对已有的分水岭算法和基于区域的检测算法进行分析,并提出一种适用于自然场景下的分水岭区域算法,以适用于移动终端。

## 1 相关研究

现有的文字识别检测技术中,主要有以下几种算法。

### 1.1 分水岭算法

分水岭算法就是基于形态学的分割算法,是对图像的梯度分割。而图像分割的目的是将图像分割成不同的特体,即提取出物体的边缘。文献[4]中对分水岭算法进行分析,通过实验表明,分水岭算法具有以下优点:①计算速度较快;②物体轮廓线的封闭性;③定位的精确性,并且分水岭算法对微弱边缘也具有良好的响应。但采用分水岭算法进行图像分割时,通常会产生过度分割的现象且耗时较长。由于自然场景下文字的特点,如果单一使用分水岭算法进行处理,就会产生过度分割的情况。

### 1.2 纹理检测算法

利用纹理检测算法可以筛选出图片中的文本区域,现有的纹理检测算法一般有:像素域法<sup>[5]</sup>、共生矩阵法、Tamura 纹理表达法、小波变换法、Gabor 小波等。Tamura 等提出了 Tamura 纹理表达法,将 Tamura 纹理特征与心理学相结合<sup>[6]</sup>。小波变换<sup>[7]</sup>也是一种常用的纹理分析和分类方法。

文献[8]使用 Gabor 滤波对纹理进行分析,首先对图像进行 RGB 图像空间转化,再构成相应的 RGB 颜色通道,利用二维 Gabor 滤波器在每个通道上进行高频转换,将每种纹理对应的主要频率成分经过频域和时域的转换,最终提取出该图像的纹理特征。根据得到的纹理特征进行合并分析,得到文本区域。此算法可有效避免各个纹理特征之间的频率混叠现象,适用于离散纹理频率的图片中。

纹理检测方法虽然可以通过文本区域定位,然后进行识别,但是该方法只适用于处理简单场景下且文字与背景的融合度不高的图片,例如文本文字,不适宜在复杂多变的自然场景下的文字识别。

### 1.3 基于区域算法

区域图像算法<sup>[9]</sup>的基本思想是先确定一个分裂合并的准则,即区域特征一致性的测度,当图像中某个区域的特征不一致时就将该区域分裂成多个相等的子区域,当相邻的子区域满足一致性特征时则将它们合成

一个大区域,直至所有区域不再满足分裂合并的条件为止。

常见的区域算法主要有  $K$  均值聚类算法、ISODATA 算法<sup>[10]</sup>。由 Mac Queen 提出的  $K$ -均值聚类算法<sup>[11]</sup>,其主要思想是通过迭代过程把数据集划分为不同的类别,使得评价聚类性能的准则函数达到最优,从而使生成的每个聚类内紧凑,类间独立。其原理简单、便于处理大量数据。这一算法不适合处理离散型属性,但是对于连续型具有较好的聚类效果。与  $K$ -均值算法相比,ISODATA 算法<sup>[12]</sup>则更加灵活;ISODATA 算法加入了一些试探步骤,并且可以进行合并和分裂,使其能利用中间结果所取得的经验更好地进行分类,但该算法对分割光照敏感的图像时,会产生较大误差。

通过对比分析,并结合各个算法的优缺点与自然场景的特点,文中提出了一种基于分水岭的区域检测算法。

## 2 算法概述

### 2.1 算法总体设计

文中采用了一种基于分水岭的区域检测算法,流程如图 1 所示。

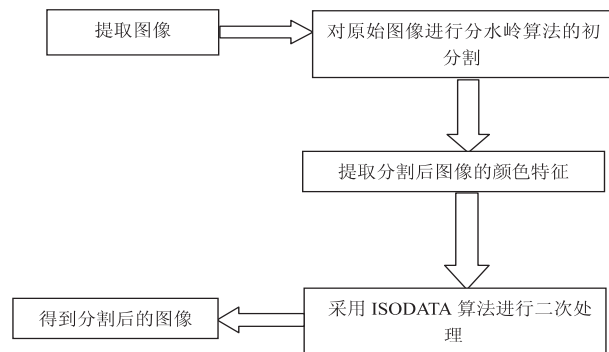


图 1 区域检测算法流程图

下面对该算法进行阐述。

由于自然场景下的图片具有光照不确定性等未知因素,经由对已有算法的比较和分析可知,利用分水岭算法计算速度较快,具有良好的轮廓封闭性和定位的精确性的优点,可以对原图进行第一次分割,经过分割后的图片可以去除绝大多数的毛刺,间接地减少了自然场景下所造成的对图片的干扰。

虽然分水岭算法对微弱边缘具有良好的响应,但是其会造成过度分割,所以再利用 ISODATA 算法对初分割的图片进行二次分割。ISODATA 算法的分割合并,和启发式的特点,可以对初分割的图片进行二次分割,以此将分水岭算法中产生的过度分割进行合并,大幅提高了分割的精度,最终得到分割后的图像。

### 2.2 算法思想

分水岭算法的基本思想是:将一幅图像看作是自

然界的地形地貌,图像中的每一像素的灰度值表示该点的海拔高度,其每一个局部极小值及其影响区域称为集水盆,而集水盆的边界则为分水岭,通常描述分水岭变换有“雨滴法”,即当一滴雨水分别从地形表面的不同位置开始下滑,其最终将流向不同的局部海拔高度最低的区域(称为极小区域),那些汇聚到同一个极小区域的雨滴轨迹就形成了一个连通区域,称为集水盆。不同区域的水流相遇时的界限,就是期望得到的分水岭。

文中算法先对原始图像采用分水岭算法进行初分割,将原图像转换成一个标记图像,其中所有属于同一集水盆的点均被赋予同一个标记,并用一个特殊的标记来标识分水岭上的点。经过分水岭算法的初分割之后,可以明显减少光照对图片的影响。

### 2.3 颜色空间提取

许多实验表明,采用颜色的特征值提取,对于最终的图像检测有更好的效果,因此文中对图片的颜色空间特征进行提取。

常用的颜色空间有以下几种:RGB, YUY, HIS 和 HSV。这几种颜色空间都有各自的特点和适用范围。

RGB 是目前运用广泛的颜色系统之一<sup>[13]</sup>。但 RGB 不适用于图像处理之中,处理效率低。因此文中采用 HSV 颜色空间。

HSV (Hue, Saturation, Value) 是根据颜色的直观特性创建的一种颜色空间,这个模型中颜色的参数分别是:色调( $H$ )、饱和度( $S$ )、亮度( $V$ )。其中, $H$ 表示不同的颜色, $S$ 表示颜色的深浅, $V$ 表示颜色的明暗。HSV 对用户来说是一种直观的颜色模型。由于自然场景下对图片影响最大的是光照因素,这就意味着,HSV 颜色空间基本可以满足用于处理自然场景下拍摄的图片。

文中先将初分割的图片利用 RGB 到 HSV 转换公式进行转换<sup>[14]</sup>。最终得到三个颜色特征:

$H =$

$$\begin{cases} 60 \times \frac{G-B}{S-V}, S \neq 0 \text{ 且 } \max(R, G, B) = R \\ 60 \times (2 + \frac{B-R}{S \times V}), S \neq 0 \text{ 且 } \max(R, G, B) = G \\ 60 \times (4 + \frac{R-G}{S \times V}), S \neq 0 \text{ 且 } \max(R, G, B) = B \end{cases}$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)}$$

$$V = \max(R, G, B)$$

其中,若  $H < 0$ , 则  $H = H + 360$ ;  $R, G, B$  分别为 RGB 模型中的三个分量。

### 2.4 ISODATA 算法

在经过了分水岭算法的初分割,颜色空间的特征

值提取之后,将进行算法的第三步处理,即对初分割的图片进行二次分割处理。ISODATA 算法具有的分割合并,和启发式的特点,可以较好地对分水岭算法分割所产生的过度分割区域,进行二次处理合并。

ISODATA 算法的原理比较简单,即通过设定初始参数,通过判断聚类中心,对样本进行归并与分裂,其中分裂和合并的规则分别如下:

(1) 当某两类聚类中心距离小于某一阈值时,将它们合并为一类;

(2) 当某类标准差大于某一阈值或其样本数目超过某一阈值时,将其分为两类;

(3) 在某类样本数目少于某阈值时,将分裂的两个类取消。

最后对参数和样本进行迭代运算,得到一个比较理想的分类结果。

文中根据 ISODATA 算法的基本原理,并结合已有参数,将初分割的图片进行区域的合并和分裂操作,给出以下算法流程:

Step1: 初始化。

输入  $N$  个 HSV 颜色空间中的  $H$  分量样本  $\{X_i, i = 1, 2, \dots, N\}$ , 预选出  $N_c$  个初始聚类中心  $\{m_1, m_2, \dots, m_{N_c}\}$  并对相关参数进行初始化:

$C$ : 预期的类数;

$\theta_N$ : 每一聚类域中最少的样本数目,如果少于此数即不作为一个独立的聚类;

$\theta_s$ : 一个聚类域中样本距离分布的标准差;

$\theta_c$ : 两个聚类中心间的最小距离,若小于此数,两个聚类需进行合并;

$M$ : 在一次迭代运算中可以合并的聚类中心的最多对数;

$K$ : 迭代运算的次数。

Step2: 初始聚类中心运算。

确定聚类  $S_j$  由  $N$  个样本组成,修正各聚类中心  $Z_i$ , 并计算出各聚类域  $S_j$  中样本与各聚类中心间的平均距离  $\bar{D}_j$  及全部模式样本与其对应聚类中心的总平均距离  $\bar{D}$ 。

Step3: 判断分裂,合并运算。

(1) 若迭代运算次数已达到  $K$  次,即最后一次迭代,结束算法;

(2) 若  $N_c \leq \frac{C}{2}$ , 即聚类中心的数目小于或等于规定值的一半,则对已有聚类进行分裂处理;

(3) 若迭代运算的次数是偶数次,或  $N_c \geq 2C$ , 进行合并处理;否则(既不是偶数次迭代,又不满足  $N_c \geq 2C$ ), 进行分裂处理。

Step4: 分裂运算。



计算各类内样本到类中心的标准差向量:

$$\sigma_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{nj})^T$$

计算其各个分量,然后求出每一类标准差向量  $\sigma_j$  中的最大分量  $\sigma_{j\max}$ 。若有某一类中最大分量大于  $\theta_s$ , 同时又满足下面两个条件之一:

(1)  $S_j$  样本总数超过规定值的一倍以上;

$$(2) N_c \leq \frac{C}{2}$$

则将该类  $Z_i$  分裂为两个类,且  $N_c = N_c + 1$ 。

两个新类的中心  $Z_i^+$  和  $Z_i^-$  分别是在  $Z_i$  中相应于的分量加上和减去  $k\sigma_{j\max}$ , 其中  $0 < k \leq 1$ 。

分裂后,转至 Step2。

Step5:合并运算。

计算各类中心间的距离  $D_{ij}, i = 1, 2, \dots, N_c - 1; j = 1, 2, \dots, N_c$ 。

比较  $D_{ij}$  与  $\theta_c$ , 并将小于  $\theta_c$  的  $D_{ij}$  按递增次序排列,从最小的  $D_{ij}$  开始,将相应的两类合并,并计算合并后的聚类中心  $Z_k^*$ 。

### 3 实验结果

#### 3.1 仿真环境

文中在 MATLAB 7.0 中进行实验仿真。为了验证文中算法适用于自然场景,对算法每一步进行实验。

#### 3.2 实验结果分析

图2为自然场景下拍摄的原始图像,图3为经由第一次初分割,即分水岭分割后得到预处理后图像。



图2 自然场景下的原始图片



图3 第一次初分割图片

由图3可以看出,经过分水岭算法的初分割后,对图像文字区域有了大致的区域识别,但是产生了过度分割的现象。

图4为经由 ISODATA 算法的第二次分割,可以将文字区域完整识别出来,分割效果显著。



图4 二次分割图片

### 4 结束语

文中针对自然场景下图片的特点进行分析,提出一种基于区域的 ISODATA 和分水岭算法,并能够快速有效地对自然场景图像的文字进行识别。

#### 参考文献:

- [1] Ali H, Ali Y, Hossain E, et al. Character recognition using wavelet compression[C]//Proc of 13th international conference on computer & information technology. Dhaka: IEEE, 2010:452-457.
- [2] 高新波. 视觉信息质量评价方法[M]. 西安:西安电子科技大学出版社,2010.
- [3] 宋 砚,刘安安,张勇东,等. 基于聚类的视频字幕提取方法[J]. 通信学报,2009,30(2):136-140.
- [4] 王国权,周小红,蔚立磊. 基于分水岭算法的图像分割方法研究[J]. 计算机仿真,2009,26(5):255-258.
- [5] Angadi S A, Kodabagi M M. Text region extraction from low resolution natural scene images using texture features[C]//Proc of IEEE 2nd international advance computing conference. Patiala: IEEE, 2010:121-128.
- [6] 施智平,胡 宏,李清勇,等. 一种快速有效的图像纹理谱描述子[J]. 计算机辅助设计与图形学学报,2004,16(12):1703-1707.
- [7] Grigorescu S E. Nicolai features based on Gabor filters[J]. IEEE Transactions on Image Processing, 2002, 11(10):1160-1167.
- [8] Nirmala S, Nagabhushan P. Foreground text segmentation in complex color document images using Gabor filters[J]. Signal, Image and Video Processing, 2012, 6(4):669-678.
- [9] 张 引,潘鹤鹤. 复杂背景下文本提取的彩色边缘检测算子设计[J]. 软件学报,2001,12(8):1229-1235.
- [10] 周书仁,殷建平. 基于 Haar 特性的 LBP 纹理特征[J]. 软件学报,2013,24(8):1909-1926.
- [11] 罗 可,蔡碧野,吴一帆,等. 数据挖掘中聚类研究[J]. 计算机工程与应用,2003,39(20):182-184.
- [12] 张语涵,孙劲光,苗锡奎. 一种基于区域的彩色图像分割方法[J]. 计算机仿真,2010,27(6):271-274.
- [13] 左 飞,万晋森,刘 航. 数字图像处理原理与实践[M]. 北京:电子工业出版社,2012:260-264.
- [14] 邢伟利. 图像检索中颜色的特征提取及匹配算法[J]. 微机发展(现更名:计算机技术与发展),2004,14(2):86-88.

自然场景下基于区域检测的文字识别算法

作者：[陈梓洋](#)，[王宇飞](#)，[钱侃](#)，[张超](#)，[孙知信](#)，[CHEN Zi-yang](#)，[WANG Yu-fei](#)，[QIAN Kan](#)，[ZHANG Chao](#)，[SUN Zhi-xin](#)  
作者单位：[南京邮电大学 物联网学院, 江苏 南京, 210003](#)  
刊名：[计算机技术与发展](#)[ISTIC](#)  
英文刊名：[Computer Technology and Development](#)  
年，卷(期)：[2015\(7\)](#)

引用本文格式：[陈梓洋](#). [王宇飞](#). [钱侃](#). [张超](#). [孙知信](#). [CHEN Zi-yang](#). [WANG Yu-fei](#). [QIAN Kan](#). [ZHANG Chao](#). [SUN Zhi-xin](#) [自然场景下基于区域检测的文字识别算法](#)[期刊论文]-[计算机技术与发展](#) 2015(7)