

# Bank Card Number Intelligent Identification Based on Digital Image Processing Technology

Longfei Han, Wan Li, Shuguang Zeng\*, Huiling He

College of Science, China Three Gorges University, Yichang Hubei  
Email: \*zengshuguang19@163.com

Received: Jul. 8<sup>th</sup>, 2016; accepted: Jul. 24<sup>th</sup>, 2016; published: Jul. 27<sup>th</sup>, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The digital recognition system is an important part of the modern intelligent management field. Its role is particularly reflected in the intelligence and digital aspects of daily life. With the bank card of China's Banks as the research object, this paper proposed a bank card number automatic identification technology based on digital image processing. The main purpose of this technology is to achieve the automatic extraction of bank card number information as a text file output for practical applications. Its main components are image acquisition, image preprocessing, digital number positioning, character feature extraction, character recognition, character segmentation and character recognition, etc. Through the experimental test, it is found that the technology can achieve more than 85.6% accuracy.

## Keywords

Bank Card Number, Intelligent Identification, Digital Image Processing, Character Segmentation

---

# 基于数字图像处理的银行卡号智能识别技术

韩龙飞, 李婉, 曾曙光\*, 何慧灵

三峡大学理学院, 湖北 宜昌  
Email: \*zengshuguang19@163.com

---

\*通讯作者。

收稿日期：2016年7月8日；录用日期：2016年7月24日；发布日期：2016年7月27日

## 摘要

数字识别系统是现代化智能管理领域的重要组成部分，其作用特别体现在日常生活的智能化和数字化方面。本论文以我国各银行的银行卡为研究对象，提出了一种基于数字图像处理的银行卡号自动识别技术。该技术的主要目的是实现银行卡号信息的自动提取并将卡号以文本文件形式输出，以切合实际应用。其主要组成部分为：图像采集、图像预处理、数字号码定位、字符特征提取、字符标记、字符分割、字符识别等。通过实验测试，发现该技术可以实现85.6%以上的准确识别率。

## 关键词

银行卡号，智能识别，数字图像处理，字符分割

## 1. 引言

随着计算机技术的发展，信息处理逐步向数字化转变，自动化的能力和水平不断提高。数字图像处理(Digital Image Processing)又称为计算机图像处理，它是指将图像信号转换成数字信号并利用计算机对其进行处理的过程。利用 MATLAB 等软件的强大处理能力可以迅速实现该过程，例如图像压缩、图像增强、图像复原等[1]-[4]。阿拉伯数字作为全球通用数字，在经济、生活等各个领域都扮演着重要角色，例如作为密码、身份认定等。如何高效、精密、安全的智能识别图像中数字信号已成为学术界的重要目标之一。

自动数字识别技术在国内外取得了飞速发展，具有较高的实用性价值。在农业方面，可用于粮食、种子纯度检测[5]。在畜牧业方面，可用于奶牛数字化管理[6]。在经融方面，可用于金融票据的面值识别[7][8]。在社会方面，可用人口普查，仪表读数，手写数字与印刷数字识别等[9][10]。研究者一方面探索更多可识别的数据库，将其应用于交通管理汽车牌照识别系统[11][12]，公寓管理门牌号识别[13]等。另一方面采用模式识别方式[14]、图像二值化阈值法[15]、算子优化[16]等方法建立更加精密的处理系统。

将自动数字识别技术应用于银行卡号与社会生活息息相关。现有较为成熟的卡号识别系统仅限于磁条读取，磁条的寿命有限，刷取一定的次数后会出现不同程度的磨损。而本文采用图像处理光学方法非接触式测量，对卡面几乎没有任何损耗，可以无限次、大批量的对卡号进行提取统计，同时解决了现有的银行卡号信息传输手动输入工作量大的问题，大大提高了工作效率。其处理方式基于通用的计算软件，拥有智能识别银行卡号的优势，可用途在银行自动存取款、企业财务部统计等社会金融方面。但是由于我国银行卡种类繁多，卡面背景相当复杂，使得银行卡号识别在单纯识别数字基础上增加了很大难度。本文所构建的自动识别银行卡号系统的主要任务是对获得的图像进行预处理、卡号定位、字符分割、字符识别和输出。采用 MATLAB 软件对每个模块依次进行编程和优化，并运用一系列图像处理手段排除背景干扰，以达到快速、准确识别并输出卡号数字的目的。最后通过一系列实际测试来检测该技术的实用性和准确性。

## 2. 图像识别系统的建立

### 2.1. 银行卡号自动识别理论知识

银行卡号自动识别系统的建立采用 MATLAB 软件，其自带库函数与手动编程结合便可得到用户需要的模块函数。其优点是其功能强大，支持的文件格式很多，如\*.BMP、\*.JPG、\*.GIF 等普遍应用格式，

基本包括了现代摄像设备的输出格式，所以可以很方便的获得图像采集后需要的数据。在定位字符时，采用 Canny、Log 等算子[17]进行边缘检测、并采用开闭运算子进行形态学操作，数字长宽比特特征识别等能准确捕捉。银行卡卡号以阿拉伯数字组合，简单通用。虽字体不一样，但可根据不同银行卡的字体预先建立不同的图片库，以适应不同银行卡数字的识别。以上技术支持为自识别银行卡号码系统的建立创造了很好的条件。只要解决其复杂背景的影响，就可以很好的完成识别功能。

## 2.2. 系统构造

整个银行卡号自动识别系统的设计分块如图 1 所示，主要包括以下几个步骤：1) 图像采集：用手机拍摄纯色背景下的银行卡图片，保证像素较高、图片较为清晰。2) 图像预处理：目的是使图像中 useful 信息与背景噪声区分开，需要用到图像灰度化、图像增强等操作。3) 字符定位：目的是将有用信息定位，便于下一步提取。将整个图形分块，连通每个区域后再将区域信息进行下一步识别。4) 字符分割与识别：在上一步得到的连通区域后进行图像增强、水平矫正、字符分割等操作；将每个字符图片，与预先设置好的模板进行对比确认；如果与数字库模板的误差在设定范围以内，则判断为数字域，反之亦然；最后将真正有效的数字提取出来按照银行卡号顺序转化为文本文档输出。

## 3. MATLAB 图像处理流程

### 3.1. 图像预处理

#### 3.1.1. 灰度化

拍摄出来的图片是\*.JPG 格式的三维彩色图像。由于图像是彩色，背景颜色丰富，不利于定位出数字串。所以将原始图像转化为灰度图像不仅内存减小而且有用信息不会丢失，处理更加简单。由于银行卡的卡号都是黑色字体的，故灰度转换编程原理是增强黑色字体的对比度来排除彩色背景对识别的障碍。但 RGB 空间颜色区别度不高，故本文变换到 HSV 空间，提取图像亮度信息。灰度转换后图像是数字灰度图像如图 2(b)，灰度值在 0~255 之间。在实际情况下，图片拍摄光照不均，导致二值化时有些灰度值较小的暗背景出现丢失现象，此时可采用顶帽运算[17]，处理方法是原有的图像减去开运算后的图像。

#### 3.1.2. 反相

由于灰度化后的银行卡号一般为黑色，若直接对其进行处理，会增加程序算法的复杂性，所以有必要将包含卡号信息的区域反相，变换为白色方便处理，如图 2(c)，该处理可减少计算量。

#### 3.1.3. 去除背景

去除背景的目的是仅提取卡面信息，排除背景干扰。本文采用方法是首先边缘提取，由于卡面图像背景复杂，选用 Canny 算子可根据特定要求进行调整以识别不同的边缘特性[12]。故本文采用 Canny 算子将卡的外边框提取出来，进行膨胀防止断线，然后对封闭线进行填充。填充后的图片与反相后图片相乘，根据  $1 \times 1 = 1$  的原理，即可将背景变为全黑。这种方法不仅很好的保留了卡面数字信息，也将复杂背景简化了一大部分。

#### 3.1.4. 二值化

图像的二值化算法是一种阈值算法，将图像划分为两个部分，图像非黑即白，在数据处理时非 0 即 1。二值化取最佳阈值的算法如下：1) 找出图像中灰度值的最小值 `min_gray` 和最大值 `max_gray`，相减得灰度值最大范围。2) 将此范围四等分，取比 `max_gray` 小一等分的值作为阈值。该处理只留下灰度值相对靠近于最大值 `max_gray` 的部分，也就是数字部分和其他一些亮度较高的取为 1，亮度较低的取为 0，因此在保留卡号信息的基础上尽可能去除了周边噪点。去除背景及二值化后的效果如图 2(d)所示。

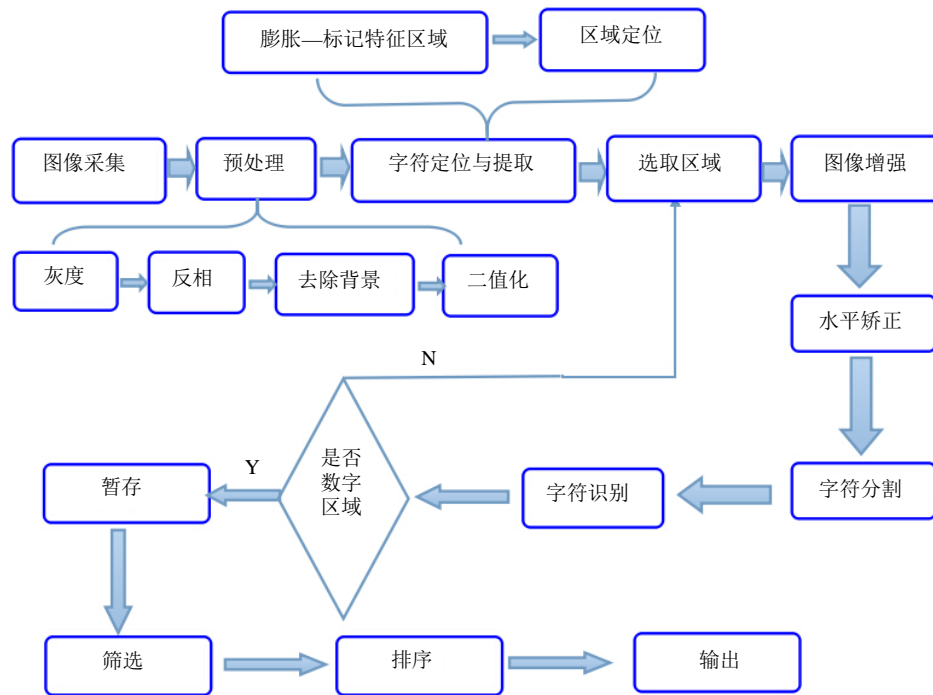


Figure 1. The System structure diagram  
图 1. 系统结构图



Figure 2. The bank card number automatic identification process map  
图 2. 农行卡号自动识别过程图

### 3.1.1.5. 去除噪点

由于二值化结果会导致图像中有许多像素点很小的高频区域，也就是噪点。噪点不仅会对图像质量造成影响，还会在定位提取时产生干扰，所以去除噪点可以得到更高质量图片。本文通过标记区域，判断其区域长宽特征值进行区域筛选，去除区域长宽很小的区域即噪点，此操作可清除一大部分无效区域。图 2(d)经过去除噪点后变成了图 2(e)。

## 3.2. 字符定位与提取

### 3.2.1. 膨胀

通过膨胀图像, 将靠得较近的字符连通成一个区域(如图 2(f)所示)。进行膨胀操作时, 其结构元素取元素全为 1 的向量, 进行横向膨胀, 避免银行卡号码上下的图像噪声也被连通, 影响数字的识别。在 BoundingBox 中提取膨胀后各区域的长宽信息。

### 3.2.2. 数字区域定位

在连通区域之后, 需要确定哪些区域是数字区域。其方法是将每个连通区域的原始图片进行识别预处理即字符分割后再识别(分别见下文 3.3 和 3.4)。若识别结果在既定误差范围内, 则判断为数字区域, 将其数字信息暂存下来; 若超过误差范围则定为非数字区域, 进行下一个区域识别。直到完成所有区域的识别, 数字区域被保存进行下一步操作, 最终筛选出为银行卡号码的数字。

## 3.3. 字符分割

### 3.3.1. 图像增强

在字符分割之前需要消除背景使目标图像更加清晰, 解决如图 3(a)中字符图片自身不连续或者与其它字符串连通的问题。本文采用了边缘提取图像与原图像相加的方法来增强图像效果, 得图 3(b)。

### 3.3.2. 水平矫正

因图像拍摄可能会有一定倾斜角, 故需对卡号区域提取后的图进行水平矫正。水平矫正原理是取第一个字符边缘左上角坐标和最后一个字符左上角坐标, 计算出斜率再转换为弧度[11]。利用旋转函数将增强后的图像旋转相应角度即可得到水平图像。旋转过程可能会导致像素变化, 可再次进行二值化处理。图 3(c)给出了图 3(b)经过水平矫正后的图像。

### 3.3.3. 字符分割

图 4 中显示了待分割的字符, 包括各类银行卡号。结果证明该系统能够准确提取银行卡号区域位置, 只要有相应的字符图库就可以识别并输出信息, 具有普遍使用的可能性。下一步则是分割, 在字符分割时, 首先需要滤除非目标数字。因噪声图像高度较小, 故设定一个高度阈值, 对单个小区域高度进行筛选。具体步骤: 先对图像区域化标记。如图 4(a)和图 4(b)此时都出现了数个非目标小区域。设定图像所有小区域中最大高度  $high\_max$  的二分之一为阈值, 大于阈值则分割提取, 即可完整保留目标数字, 去除干扰部分的小区域。图 5 给出了图 4(a)中分割出来的各个卡号数字, 非常清晰, 有利于下一步的字符识别。

## 3.4. 字符识别

### 3.4.1. 归一化处理

字符分割得的字符图像大小各异, 需对字符的大小归一化处理。本文采用图像点对点异或法来判断其数值范围是否匹配模板值, 需要目标图像与模板像素大小相似。故对模板和分割后图片也归一化处理。归一化采用按比例线性缩放来进行归一化, 根据银行卡上数字长与高的实际比例为 5:7, 将分割后的图片缩放为指定的大小: 长  $\times$  高 =  $35 \times 50$  个像素点。

### 3.4.2. 字符比较

将切割后的第一幅字符图像与模板 0~9 的图片取异或后 1 的数量即为图片与数字模板的误差值。采用 sort 函数排序后找出其中值最小的, 取它对应的模板图像为其目标图像匹配图, 输出匹配图像所对应的 0~9 之间序号数保存。依次下去, 取第二幅, 第三幅。用 for 循环实现所有字符比对。方法为: 预设

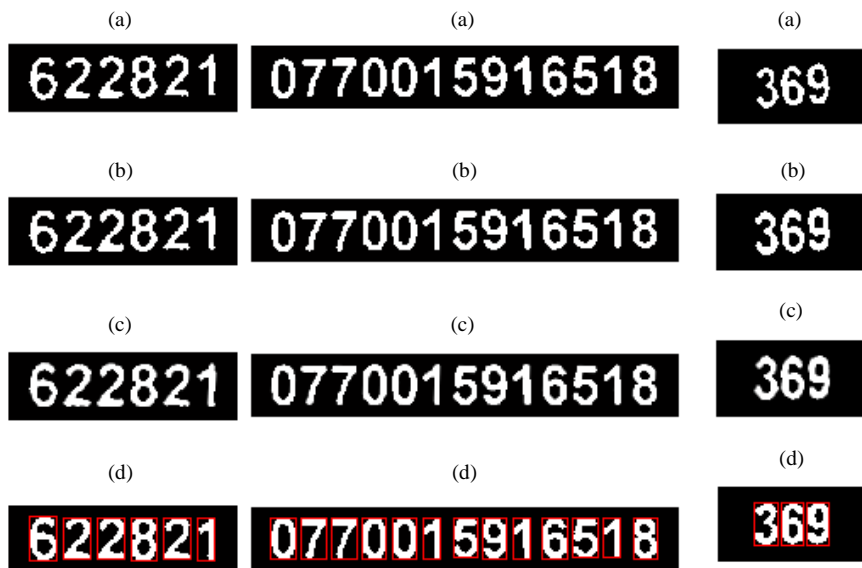


Figure 3. Character segmentation preparation process, (a) original card number area, (b) the enhanced image, (c) correction of the image, (d) the marked image

图 3. 字符分割预备过程, (a) 原始卡号区域, (b) 增强图像, (c) 水平矫正图像, (d) 标记图像

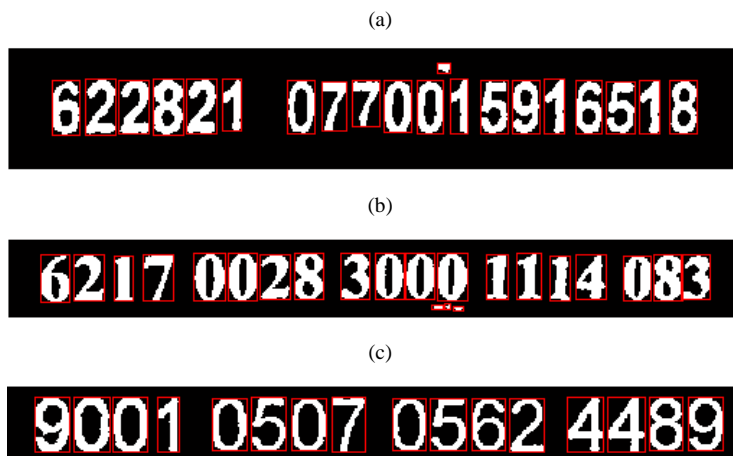


Figure 4. Pinpoint stay after a single character segmentation pattern, (a) ABC card number (b) CCB card number (c) ICBC card number

图 4. 定位出单个字符后的待分割图样, (a) 农行卡号(b) 建行卡号(c) 工行卡号

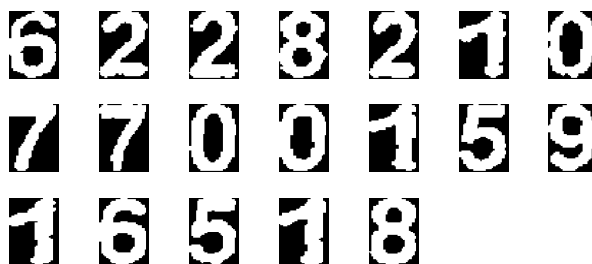


Figure 5. ABC card number character segmentation pattern

图 5. 农行卡号字符分割图样

一个误差值，若所有字符图片与其匹配完成的模板误差值中的最小值大于预设误差值，则将其判断为非数字区域。因为当定位区域中字符不是所要求的数字时，进行模板匹配后其误差值会较大。根据此原理取待测字符与所有数字模板的误差值最小值来判断，防止识别判断错误。在 3.2.2 中，各个连通域的字符比较完成后即可找出银行卡上存在的所有数字区域。

### 3.4.3. 筛选银行卡号

在所有的数字区域中筛选出银行卡号码，如图 3 中最左列和中间列图片为真正的银行卡号码，最右列图片为干扰数字。根据我国银行卡特性，银行卡数字串基本都在同一行，也就是区域的纵坐标一致，并且每一串卡号数字一般都在 4 位以上。根据这两个特征，区域长度较长且区域纵坐标较接近的为真正的银行卡号码；反之判断为干扰数字域。最后，根据数字区域横坐标进行先后排序并保存信息。

### 3.4.4. 数据输出

将 3.4.3 中按照顺序保存的银行卡数字信息，转成 txt 格式输出。这种输出方式也可用于数字显示输出，满足实际需求。图 6 则是与图 5 对应的农行卡号码的数字输出，与原图卡号(见图 4(a))完全符合，实现了银行卡号识别并输出的目标。

## 4. 实验结果及分析

为了检验系统的可靠性，进行了大量实验检测。为保证样本的随机性和广泛性，实验样本图片来源分为网上随机搜索和人工手动拍摄两种。评估识别系统的指标有识别率、识别速度。其定义为识别率 = 识别正确的次数/实验的总次数，识别速度 = 识别一张卡所需要的平均时间。测试对象主要为中国四大银行：农行、中国银行、建行与工行卡图片。测试方法是将 txt 文件输出数据与卡号对比正确与否，记录每一次测试的时间并取平均值。这里需要说明的是由于网上图片存在光照不均匀、清晰度不够等质量问题，故在测试时其二值化阈值根据图片质量有所变化；而人工手动拍摄的照片质量较好、光照强度固定，测试时其二值化阈值可以取为一定值。经过若干次实验后，所得测试结果如表 1 所示。

从表 1 中分析可得：不管是手工拍摄图片还是网上随机搜索图片都能实现较高的识别率；农业银行卡的卡号识别率最高达到了 97.5%，其它银行的识别率较低；各银行卡平均识别率为 85.6%；农行的卡号平均识别速度较其它银行卡要快。影响卡号识别率和识别速度的重要原因在于卡面背景的复杂度。例如某些卡的卡号与其周围背景图案的颜色类似，导致两个相邻数字连在了一起。农行的卡面背景相对简单，故其识别率和识别速度较快；而其他银行卡种类繁多、背景复杂，存在数字与背景难以分割的困难，导致识别率较低、识别速度较慢。影响系统识别率的另一个原因在于卡号数字本身的质量。比如农行卡号字体粗细均匀；而建行等的卡号字体粗细不一致，比如“0”的上下两个连接处太细，在区域分割时，容易被分成两个区域，导致识别容易出错。

## 5. 结论

提出了一种基于数字图像处理的银行卡卡号自动识别技术。该技术是通过图像采集、图像预处理、

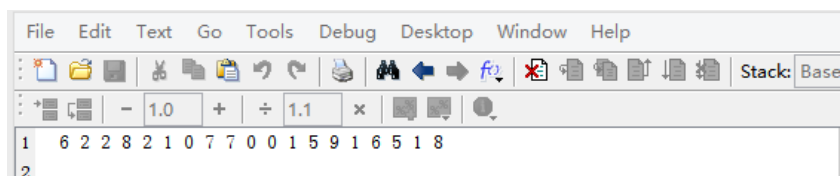


Figure 6. ABC card number output map

图 6. 农行卡号输出结果图

**Table 1.** The test results of bank card recognition rate  
**表 1.** 银行卡识别率测试结果

类别	测试次数 (次)	正确次数 (次)	识别率 (%)	总识别率 (%)	平均识别速度 (秒/次)	
农行卡	网上图片	20	19	95	97.5	8.643
	人工拍摄	20	20	100		6.543
中国银行	网上图片	25	22	88	88	9.642
	人工拍摄	25	22	88		9.642
工行卡	网上图片	30	26	86.67	86.67	10.293
	人工拍摄	15	13	86.67		15.498
其它	网上图片	30	25	83.33	81.67	13.615
	人工拍摄	15	12	80		14.500
合计	所有图片	125	107	85.6	85.6	11.246

数字号码定位、字符特征提取、字符标记、字符分割、字符识别等步骤来有效地实现银行卡卡号自动识别与输出。对中国农业银行、中国银行、中国工商银行等几大银行的银行卡进行了测试，发现中国农业银行的卡号识别率最高，达到了 97.5%；各类银行卡的综合识别率达到了 85.6%。影响银行卡卡号识别率和识别速度的主要因素有卡面背景的复杂度、照片质量、卡号数字本身的质量等。通过制作专门的银行卡号识别装置可以获得较高质量的银行卡图片；而通过采用人工神经网络等机器学习方法可以进一步提高卡号识别率。总的来说，该技术有望取代传统的磁条读取卡号的方式，形成一种新的无接触的卡号读取方式。

## 基金项目

本论文得到宜昌市科技局自然基础科学研究与应用专项项目(A15-302-a15)的支持。

## 参考文献 (References)

- [1] 王秋雨. MATLAB 图像处理的几个应用实例[J]. 福建电脑, 2011, 27(11): 6-7.
- [2] 陈霞, 卫静婷. 基于 MATLAB 的图像增强处理与应用[J]. 科技广场, 2013(3): 73-76.
- [3] 涂望明, 魏友国. 施少敏. MATLAB 在数字图像处理中的应用[J]. 微计算机信息, 2007, 23(6): 299-300.
- [4] Chickerur, S. and Kumar, M.A. (2010) Image Restoration: Past, Present and Future. *Recent Patents on Computer Science*, 3, 108-126.
- [5] 赵新子. 玉米种子纯度数字图像处理与识别技术的应用研究[D]: [硕士学位论文]. 长春: 吉林农业大学, 2005.
- [6] 郭卫. 奶牛数字化管理关键技术的研究[D]: [博士学位论文]. 保定: 河北农业大学, 2010.
- [7] 陆锋, 伊卫星, 袁晓辉. 神经网络模式识别对金融票据中数字特征的识别方法研究[J]. 苏州大学学报: 工科版, 2003, 23(4): 44-51.
- [8] 孙鹏辉. 基于数字图像处理的人民币号码检测算法研究[D]: [硕士学位论文]. 长春: 吉林大学, 2010.
- [9] 钱稷. 基于图像处理的字符识别系统研究[D]: [硕士学位论文]. 保定: 河北农业大学, 2007.
- [10] 吴忠, 朱国龙, 黄葛峰, 等. 基于图像识别技术的手写数字识别方法[J]. 计算机技术与发展, 2011, 21(12): 48-51.
- [11] 赵珊, 裴亮, 刘翠等. 基于 MATLAB 图像处理的汽车牌照识别研究[J]. 测绘工程, 2014, 23(1): 70-72.
- [12] 袁卉平. 基于 MATLAB 的车牌识别系统的设计与研究[J]. 工业控制计算机, 2010, 23(10): 73-74.
- [13] 侯艳丽, 杨国胜, 黄春艳. 基于字符识别的门牌号识别算法研究[J]. 河南大学学报: 自然科学版, 2004, 34(1): 76-79.
- [14] 黄玲, 李冬梅, 王焕灵. 数字字符识别的新方法[J]. 数字技术与应用, 2011(8): 220-221.



- [15] 袁璐, 李仁发. 一种改进的字符图像二值化算法[J]. 微计算机信息, 2012(3): 140-141.
- [16] 罗勇, 张华. 基于 Canny 边缘检测算子和去除小物体算法的熔池图像处理[J]. 焊接, 2005(5): 17-20.
- [17] Gonzalez, R.C. (2007) Digital Image Processing. 3rd Edition, Prentice Hall, New Jersey.

**期刊投稿者将享受如下服务:**

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>