论文

基于多信息融合的中文手写地址字符串切分与识别

付 强 丁晓青 蒋 焰

智能技术与系统国家重点实验室 清华大学电子工程系 北京 100084

摘要
该文提出了一种有效的中文手写地址字符串的切分与识别方法。首先，利用笔划提取与笔划合并将字符串图像进行过切分，得到"字根"图像序列；然后综合利用几何信息、识别信息和语义信息挑选最优的"字根"合并路径，得到最优的切分结果及对应的最优识别结果。其中，几何信息是根据当前字符串自身的特点统计得到，因此可适应不同书写风格的字符串。识别信息由单字分类器给出，包括10个候选识别结果及其相应的置信度；单字分类器采用MQDF分类器。语义信息用基于字的bi-gram模型进行描述，模型参数是从包含18万条地址数据的数据库中统计得到的。用3000个实际的手写地址样本做试验，单字识别正确率达到88.28%。

关键词　　地址识别；字符串切分；手写字符串识别

分类号　　TP391

Segmentation and Recognition Algorithm for Chinese Handwritten Address Character String

Fu Qiang Ding Xiao-qing Jiang Yan

State Key Laboratory of Intelligent Technology and Systems, Department of Electronic Engineering,
Tsinghua University, Beijing 100084, China

Abstract
An effective segmentation and recognition method of Chinese handwritten address strings is proposed. Firstly, over-segmentation is applied to character string images by extracting stroke and merging them to obtain "radical" sequences. Next, the optimal segmentation and recognition result is achieved by synthesizing geometric analysis, isolated character classifier and semantic information together. The geometric information is estimated on current character string to adapt to various writing styles of character strings. The isolated character classifier adopts MQDF classifier with ten candidate results and their confidence. The semantic information is described by a character-based bi-gram model, parameters of which are estimated from a database containing 180,000 addresses items. The algorithm is tested on 3,000 actual handwritten address samples and the single-character recognition rate is 88.28%.

Key words　　Address recognition　Character string segmentation　Handwritten character string recognition

通讯作者　付 强

作者个人主页　　付 强 丁晓青 蒋 焰

扩展功能

本文信息

▶ Supporting info
▶ PDF(265KB)
▶ [HTML全文](0KB)
▶ 参考文献[PDF]
▶ 参考文献

服务与反馈

▶ 把本文推荐给朋友
▶ 加入我的书架
▶ 加入引用管理器
▶ 复制索引
▶ Email Alert
▶ 文章反馈
▶ 浏览反馈信息

相关信息

▶ 本刊中 包含"地址识别；字符串切分；手写字符串识别"的 相关文章

▶本文作者相关文章

• 付 强 丁晓青 蒋 焰