

<<脱机手写汉字识别若干关键技术研究>>

图书基本信息

书名：<<脱机手写汉字识别若干关键技术研究>>

13位ISBN编号：9787563533909

10位ISBN编号：7563533907

出版时间：2013-1

出版时间：任俊玲 北京邮电大学出版社 (2013-01出版)

作者：任俊玲

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<脱机手写汉字识别若干关键技术研究>>

内容概要

《脱机手写汉字识别若干关键技术研究》从脱机手写汉字识别为大类别数模式识别问题出发，以在基本不降低识别速度的基础上较大地提高识别的精度为研究目的，就手写汉字库的建立、脱机汉字识别系统的构建、基于统计分析的汉字识别算法中汉字特征的描述、样本选择等问题进行了详尽阐述。

书籍目录

第1章 绪论 1.1 课题的提出 1.2 研究现状分析 1.3 研究内容和主要工作 1.4 本书的结构及安排 参考文献
第2章 脱机手写汉字识别系统的实现 2.1 脱机手写汉字识别系统 2.1.1 识别原理 2.1.2 基于模板匹配的两级脱机手写汉字识别系统 2.2 手写汉字识别系统的特征提取 2.2.1 方向线素的定义 2.2.2 方向线素特征的提取 2.3 基于模板匹配的识别算法 2.4 汉字识别系统的粗分类算法 2.4.1 粗分类特征提取 2.4.2 粗分类时的识别算法 2.5 汉字识别系统的细分类算法 2.5.1 细分类特征提取 2.5.2 细分类时的识别算法 2.6 本章小结 参考文献 第3章 HCL2004脱机手写汉字库及相关研究 3.1 研究背景及现状 3.2 HCL2000手写汉字数据库 3.2.1 数据库系统模型 3.2.2 汉字样本信息的组织 3.2.3 书写者信息的管理 3.2.4 两种信息的互查方法 3.2.5 HCL2000的数据分布 3.3 对HCL2000数据库进行更新的原因 3.4 手写汉字数据库HCL2004系统模型及实现 3.4.1 汉字样本信息的组织形式 3.4.2 样本集的划分 3.4.3 HCL2004手写汉字数据库的实现 3.5 基于HCL2004数据库的分析 3.5.1 实验系统 3.5.2 训练样本数与识别率 3.5.3 样本质量的选择与识别 3.5.4 基于单字的识别性能分析 3.5.5 关于距离测度分类器性能的分析 3.6 本章小结 参考文献 第4章 基于统计分析的手写汉字识别算法研究 4.1 引言 4.2 几种常用的平均数 4.2.1 均值 4.2.2 中位数 4.3 样本数据的分散程度描述 4.3.1 标准差 4.3.2 极差 4.4 HCL2004数据库样本特征分析 4.5 基于平均数的手写汉字标准模板 4.5.1 基于均值的标准模板 4.5.2 基于分位数的标准模板 4.6 引入数据分散程度参数的距离测度 4.6.1 引入极差的距离测度 4.6.2 引入标准差的距离测度 4.7 实验 4.7.1 不同标准模板分类性能分析 4.7.2 引入不同分散程度参数的距离测度分类性能分析 4.8 本章小结 参考文献 第5章 基于高阶统计量的距离测度 5.1 引言 5.2 在距离测度中引入高阶统计量 5.3 基于二阶标准差的距离测度 5.3.1 二阶标准差的定义 5.3.2 用二阶标准差刻画特征分布的可行性分析 5.3.3 基于二阶标准差的距离测度 5.3.4 实验 5.4 基于高阶统计量的距离测度 5.4.1 3种高阶统计量 5.4.2 基于高阶统计量的距离测度 5.4.3 实验 5.5 本章小结 参考文献 第6章 基于样本聚类的多级汉字识别系统 6.1 引言 6.2 基于DB准则的K均值聚类算法 6.2.1 K均值算法 6.2.2 DB有效性准则 6.2.3 基于DB准则的K均值算法 6.3 多模板匹配算法 6.3.1 多模板匹配算法原理 6.3.2 多模板匹配算法的设计方案 6.4 实验 6.4.1 实验系统 6.4.2 系统实现 6.4.3 实验结果及分析 6.5 本章小结 参考文献 第7章 基于广义置信度的样本选择算法 7.1 引言 7.2 字符识别的置信度分析 7.2.1 分类器的置信度和广义置信度 7.2.2 分类器的置信度估计 7.3 基于广义置信度的边界样本定义 7.4 基于广义置信度的样本选择算法 7.5 实验结果及分析 7.6 本章小结 参考文献 第8章 结束语

章节摘录

版权页：插图：图7.3(c)给出了通过由正确识别样本集中的边界样本生成的标准模板对整个训练样本集作识别的识别结果，由图可见，随着挑选出的训练样本的减少，系统的识别率一直在下降，这说明通过该方法对样本进行选择是不正确的，也起不到提高分类器识别率的作用。

究其原因，主要在于在训练样本中只包含了被正确识别的样本，而没有包含误识样本的信息。

因此尽管所选出的样本为被正确识别的样本中的边界样本，但系统的识别率却始终在降低，从而说明在训练分类器的过程中，误识样本是必不可少的。

为了更好地分析各种不同样本在训练样本集中的作用，我们给出了由方法(a)和方法(b)进行样本选择后得到的不同训练样本数下的标准模板对测试样本进行识别的识别率变化曲线，如图7.4所示。

图中，横坐标为训练样本数，纵坐标为识别率(以百分比为单位)，曲线a代表训练样本中被正确识别的样本和误识样本成比例，曲线b则表示包含全部的误识样本，只剔除了被正确识别的样本中有较大置信度的样本的情况。

<<脱机手写汉字识别若干关键技术研究>>

编辑推荐

《脱机手写汉字识别若干关键技术研究》是北京信息科技大学任俊玲博士在从事脱机手写汉字识别相关研究基础上的撰编成果。

《脱机手写汉字识别若干关键技术研究》论述严谨，实验数据详尽，可供脱机手写汉字识别的研究人员使用，对模式识别和机器学习的研究人员也有参考价值。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>