多体蒙古文（混排汉英）印刷文档识别暨统一平台少数民族文字识别系统

关 键 词：文档识别 少数民族文字 文字识别

所属年份：2007	成果类型：应用技术
所处阶段：	成果体现形式：
知识产权形式：	项目合作方式：
成果完成单位：清华大学	

成果摘要：

该项目提出了一种基于统计鉴别分析的多字体多字号蒙古文字符识别的方法，对蒙古文的常见多种字体（白体、黑体、报体、哈旺体、楷体等二十余种字体）具有良好适应性，单字测试平均识别率达99%以上；项目提出并实现了多字体蒙古文文本切分方法，在提取单词基线位置后，通过定义代价函数，综合分析字符几何位置、形状等多种信息，取代价函数极小值处作为切分点；提出了蒙汉英语种判别方法，结合汉、英文识别技术，实现蒙汉英混排文本识别。文本识别率可达96.8%以上。该系统具有蒙古文印刷文档版面分析方法，可分析处理文本、图片、表格区域。项目还研制了字符形码到Unicode标准蒙古文编码转换方法。

成果完成人：

完整信息

行业资讯

新疆综合信息服务平台
准噶尔盆地天然气勘探目标评价
维哈柯俄多文种操作系统FOR ...
社会保险信息管理系统
塔里木石油勘探开发指挥部广...
四合一多功能信息管理卡MISA...
数字键盘中文输入技术的研究
软开关高效无声计算机电源
邮政报刊发行订销业务计算机...
新疆主要农作物与牧草生长发...

成果交流

推荐成果

· 液压负载模拟器	04-23
· 新一代空中交通服务平台、关...	04-23
· Adhoc网络中的QoS保证(Wirel...	04-23
· 电信增值网业务创意的构思与开发	04-23
· 飞腾V基本图形库的研究与开发...	04-23
· ChinaNet国际(国内)互联的策...	04-23
· 电信企业客户关系管理(CRM)系...	04-23
· “易点通”餐饮管理系统YDT2003	04-23
· MEMS部件设计仿真库系统	04-23

Google提供的广告