DOI:10.16262/j.cnki.1000-8217.1993.03.013

超级汉字识别系统的研究

中国科学院自动化所文字识别实验室积多年模式识别、专家系统研究经验，于1984年起，开始汉字识别的研究工作。在国家自然科学基金、国家“七五”计划及国家“863”计划的支持下，从识别方法到识别系统，都已取得了一批具有世界先进水平的科研成果。在此基础上，推出了超级汉字识别系统，该系统按其识别对象（简体汉字或繁体汉字）的不同，分为超级汉字识别系统（A）、超级汉字识别系统（B）。

超级汉字识别系统（A）在简体汉字操作系统下运行，并包括下列子系统：

（1）联机简体手写汉字识别系统；

（2）简体印刷体汉字识别系统；

（3）脱机特定人手写汉字识别系统；

（4）脱机非特定人手写汉字系统；

（5）自由手写体数字识别系统；

（6）表格识别处理系统。

超级汉字识别系统（B）在繁体汉字系统下运行，其他子系统类同于识别系统（A）。

上述的子系统既为超级汉字识别系统中的组成部分，也可以作为独立的系统与计算机及外设配套使用，以满足单项汉字识别之需求。如：汉字识别在国内较多遇到的问题是简体汉字的识别，手写汉字的识别可用联机简体手写汉字识别系统，印刷汉字的识别则可用简体印刷汉字识别系统。当然，如需识别繁体汉字，又可以选用单项繁体汉字识别系统。

超级汉字识别系统集自动化所汉字识别研究开发之大成，可处理国内外各种字体的汉字录入问题，对办公自动化、中文信息处理将有深远的影响。

一、概　述

计算机普及到办公室，标志着办公自动化的开始。但计算机在国内及东南亚华语地区的应用，还面临着一大难题，即汉字的输入、处理。尽管经过人们的艰苦努力，已经有了一些为人所知的汉字录入方法，但是这些方法与人们从小到大学习、掌握、使用汉字的习惯不符，以至限制了计算机在汉字领域的应用。如何解决这一难题，正是汉字识别系统所肩负的重任。

由于汉字识别本身的难度，日本、台湾、新加坡及其它地方的汉字识别研究者，一般只专攻联机手写汉字识别研究、脱机手写汉字识别研究或印刷体汉字识别研究中的一项，很少有人同时涉猎几个领域。而自动化所的技术人员具有深厚的理论基础，同时在汉字识别各个领域开展研究，达到国际领先或先进水平，并开发出产品。其中，“一种手写汉字的在线识别装置”获国家发明专利，“手写体汉字识别的理论、方法与实践”获1992年中国科学院自然科学一等奖。

超级汉字识别系统是汉字识别的必然趋势。在此产品出现以前，国内外已有的汉字识别

本文于1993年5月5日收到。

系统只能解决汉字识别领域的一部分问题。例如印刷体汉字识别系统只能解决印刷品的自动录入问题，联机手写汉字识别系统只能解决现写现识别手写汉字录入问题……。每个系统都有自己的应用范围，也都有一定的局限性。由于研制单位的不同，各个不同的汉字识别系统很难结合在一起，以形成应用范围更广的产品。但是用户碰到的汉字录入问题是多种多样的，既有印刷方面的，又有手写的。如果一样买一套，不仅财务支出大，使用起来也不方便。所以，超级汉字识别系统无疑是经济、方便的最佳选择，是汉字识别系统市场的最终发展方向。

二、项目的技术状况

超级汉字识别系统是一软、硬结合的产品。该系统运行在简体（或繁体）汉字操作系统之下。系统配置为386以上微机，鼠标器、页式扫描仪。各子系统的性能指标如下：

超级汉字识别系统（A）：

联机简体手写汉字识别系统：识别率：90—97%，识别速度：实时反应速度，识别字数：二级国标6763个汉字、常用异体字、繁通字共12000个汉字。

简体印刷体汉字识别系统：识别率：96.3%，识别速度：5字/秒，识别字数：一级国际3755个汉字及常用字符。

脱机特定人手写汉字识别系统：识别率80—95%，识别速度：3字/秒，识别字数不限，简、繁体不限。

脱机非特定人手写汉字识别系统：识别率70—90%，识别速度：1.5字/秒，识别字数：3000字。

自由手写体数字识别系统：识别率：98%，识别速度：5字/秒，识别字数：10个阿拉伯数字。

超级汉字识别系统（B）：

联机繁体手写汉字识别系统：识别率：90—95%，识别速度：实时反应速度，识别字数：5401个繁体汉字及常用字符。

繁体印刷体汉字识别系统：识别率：96.3%，识别速度：5字/秒，识别字数：一级国标3755个汉字及常用字符。

脱机特定人手写汉字识别系统：识别率：80—90%，识别速度：3字/秒，识别字数不限，简、繁体不限。

脱机非特定人手写汉字识别系统，识别率：70—90%，识别速度：1.5字/秒，识别字数：3000字。

自由手写体数字识别系统：识别率：98%，识别速度：5字/秒，识别字数：10个阿拉伯数字。

（信息科学部供稿）

STUDY ON THE SUPER CHINESE CHARACTER RECOGNITION SYSTEM