汉字（自动）识别

本刊编者

自计算机系统出现以来，人机联系的研究一直是探讨的课题之一。

能够向计算机输入的数据，到目前为止，只限于符号、英文字母和数字、假名。就易于使用这一点来说，还存在不少缺点。因此图象识别可以说是迫在眉睫的研究课题。

大家知道，图象信息的处理不得不依靠光来进行，而光电子学的进步，又加速了这些工作的进展。

目前，在这方面取得显著进步的是汉字的识别技术。汉字识别，是把以文字作为媒体的信息进行输入输出，在实现最自然的数据处理中，汉字识别是不可缺少的一门技术。

与英文字母和数字相比，汉字识别困难得多，主要原因是汉字的种类多，文字图形复杂。

汉字，是我国的主要文字。随着我国社会主义革命和社会主义建设事业的迅速发展以及国际交往的扩大，对于汉字自动识别的要求日益迫切，呼声不断增长。汉字识别机对于我国的四个现代化将是一个必不可少的工具，其应用之广泛现在还很难尽言。

汉字识别主要用于各种用途的大、中型计算机的输入。这一方面是大、中型通用计算机和各类控制计算机的需要，一方面也是计算机进入国家各职能部门后一个十分突出的关键问题。例如出版印刷、新闻通讯、银行邮政、航空水运、铁路交通、工商业管理各类统计、图书、资料和政治、科技文献等的选择存储、检索和管理，文字翻译等部门使用计算机，都有一个大量输入汉字的问题，依靠打字穿孔输入费工费时，许多场合甚至使计算机无法使用。所以，汉字识别的突破必将显著提高上述部门使用计算机的水平，扩大其应用范围；同时，由于汉字数量庞大，构形复杂，远非英文字母和数字可比，故汉字识别的突破将使图象识别水平大幅度地提高。这又将使人工智能、自动控制和计算机应用提高到一个新水平。

目前，出版印刷行业对汉字识别的要求相当迫切。其最大问题是拣字速度慢、工效低、周期长、劳动强度大，因而急待出版的政治、历史、科技、卫生、工农业生产等方面的大量材料积压，大字版本的出版物不能及时提供。近年来承担的联合国文件资料中文版的印刷任务，时间要求也较紧迫。所以，为印刷出版业提供汉字识别机和电子照排机已是刻不容缓的战斗任务。

有了汉字识别机，就可使新华社能够更快地把华主席和党中央的指示和各种通讯资料送向全国各地，为缩短各地报刊出版周期提供了有利条件。

稍长远一点看，图书、资料和政治、科技文献等的选择、存储、检索和管理的现代化必将提到日程上来。目前的检索方法费延时日，不能齐全，直接、间接地拖了各方面工作的后腿。要用计算机检索和管理，就必须直接识别印刷体汉字，将各种出版物的标题、摘要或目录等自动输入计算机，实现选择存储。所以，汉字识别机和大型计算机是我国图书馆和情报中心现代化的物质基础。（集）