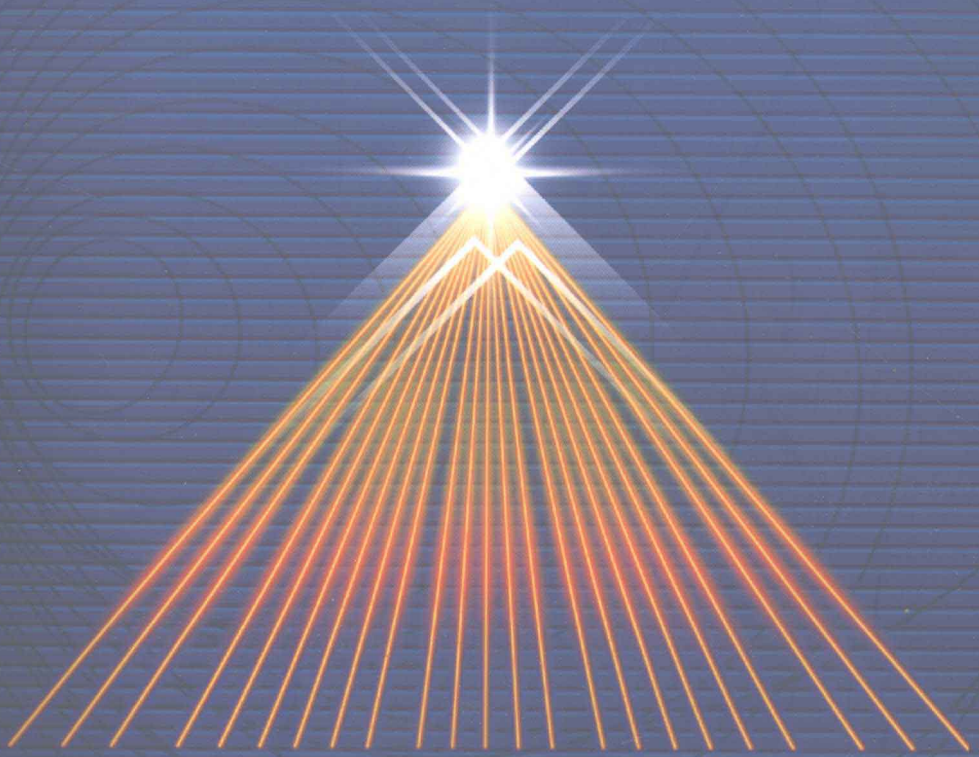
语义、句法模式识别及其应用

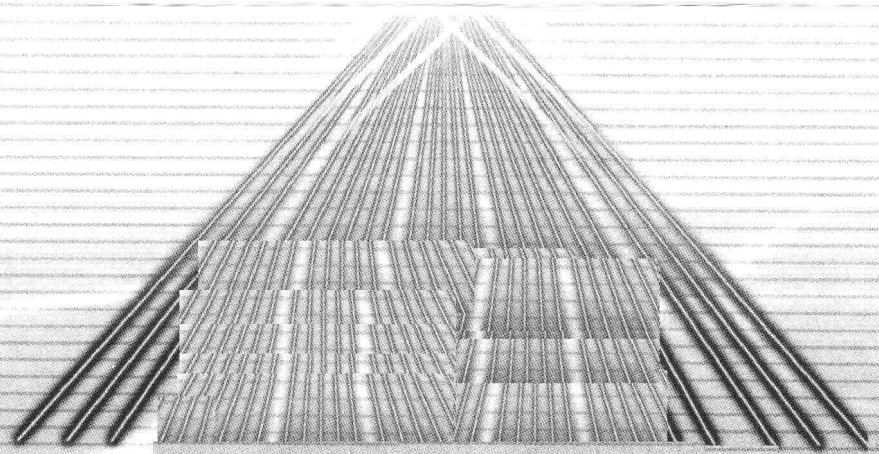
戴汝为



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

语义、句法模式识别及其应用

戴汝为



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

内容提要

本书各篇文献分别介绍了我国模式识别领域的发展。作者自上世纪70年代把当时新型学科模式识别介绍到国内，80年代在模式识别国际会议上发表论文，提出语义、句法模式识别，奠定了在线手写汉字识别的理论基础，从而形成了汉字识别的一系列发展；在此基础上，进一步从学科交叉研究，运用系统论的观点和综合集成的方法论，在离线汉字识别方面又取得了重要进展。这些工作为我国汉字识别领域做出了重大贡献。

图书在版编目(CIP)数据

语义、句法模式识别及其应用/戴汝为著. —西安：
西安交通大学出版社，2011.7
ISBN 978-7-5605-3814-3

Ⅰ.①语… Ⅱ.①戴… Ⅲ.①模式识别-研究 Ⅳ.
①TP391.4

中国版本图书馆CIP数据核字(2011)第004664号

书　　名　语义、句法模式识别及其应用
著　　者　戴汝为
责任编辑　任振国

出版发行　西安交通大学出版社
（西安市兴庆南路10号　邮政编码710049）
网　　址　http://www.xjtupress.com
电　　话　(029)82668357　82667874(发行中心)
(029)82668315　82669096(总编办)
传　　真　(029)82668280
印　　刷　西安新视点印务有限责任公司

开　　本　890mm×1 240mm　1/16　**印张** 21.625　**彩页** 1页　**字数** 643千字
版次印次　2011年7月第1版　2011年7月第1次印刷
书　　号　ISBN 978-7-5605-3814-3/TP·543
定　　价　66.00元

读者购书、书店添货，如发现印装质量问题，请与本社发行中心联系、调换。
订购热线：(029)82665248　(029)82665249
投稿热线：(029)82664954
读者信箱：jdlgy@yahoo.cn

戴汝为院士

戴汝为院士简介

戴汝为，云南昆明人，1951 年考入清华大学，1955 年毕业于北京大学，分配到中国科学院力学研究所工作，师从著名科学家钱学森，后到中国科学院自动化研究所工作至今。1980 年作为国家首批派出赴美访问学者师从著名模式识别大师傅京孙(K. S. FU)教授；1991 年当选中国科学院院士。现任中国科学院自动化研究所学术委员会主任、学位委员会主任、中国自动化学会理事长、国际自控联委员；曾任中国科学院学部主席团成员、信息技术科学部副主任、道德委员会委员。长期从事自动控制、系统科学、思维科学、模式识别、人工智能等方面研究工作；1956 年将钱学森教授 1954 年在美国出版的经典著作 *Engineering Cybernetics* 译成中文《工程控制论》(1958 年出版)。

20 世纪 70 年代初，将“模式识别”引入中国，提出语义、句法模式识别，成为“汉王”核心技术的理论基础，获国家科技进步一等奖；上世纪 90 年代初，通过人工智能的途径，在钱学森先生的直接指导下，跨入对“开放的复杂巨系统及其方法论”、“以人为主、人-机结合，从定性到定量的综合集成研讨厅”等领域的研究，对相关前沿领域进行学科交叉整合，完成“信息空间综合集成研讨体系”，在我国经济、军事及社会发展领域重大问题决策中，正在发挥着重要作用。近年来先后出版了《语义、句法模式识别及其应用》、《智能系统的综合集成》、《人-机共创的智慧》、《汉字识别的系统与集成》、《论信息空间的大成智慧》、《社会智能科学》、《系统学与中医药创新发展》、《社会智能与综合集成系统》等专著；发表学术论文二百余篇；已培养博士、硕士百余名。现任《模式识别与人工智能》、《复杂系统与复杂性科学》学术杂志主编，兼任清华大学、北京师范大学等多所大学教授及名誉教授。

1991 年 获中科院自然科学一等奖。

1999 年《智能自动化丛书》(主编共 6 册)获国家图书奖。

2001 年 获国家科技进步一等奖。

2002 年 获“何梁何利”科技进步奖。

2010 年 获中国模式识别科技终身成就奖。

序　言

西安交通大学校长
中国工程院院士　郑南宁

2006年8月，在香港举办了第18届国际模式识别大会(ICPR)，这是国际模式识别大会成立以来首次在中国举行。戴汝为先生应邀在大会做主题报告《汉字识别的历史、现状和前景》，受到大会尤其是亚洲同行的欢迎。

戴汝为先生是闻名海内外的控制论与人工智能专家，早在1955年师从钱学森先生从事工程控制论研究。上世纪70年代，他首开国内模式识别研究的先河，80年代初，把“统计模式识别与句法模式识别”结合起来，提出了“语义、句法方法”。80年代后期，率先在国内开展了人工神经网络研究，并主持完成了“人工神经元网络软件包”，从而将模式识别和人工神经网络及“形象思维”联系起来，并应用于手写汉字的识别。90年代，他把“综合集成”的思想应用于模式识别，提出了“集成型模式识别”。三十多年来，戴先生潜心研究于模式识别和汉字识别领域，做出了大量开创性的研究工作。此次将其31篇学术文章结集出版，不仅可以让读者了解国内模式识别及其在汉字识别领域的发展轨迹，更能了解戴汝为先生在这一领域所做出的贡献。

1. 当年的新学科——模式识别

20世纪70年代日本实施“模式情报、信息的处理系统”计划，这个计划突出了模式识别研究的重要性，引起国际学术界的关注，也引起了中国学者重视，本文集收录的文章1至文章5关于图形、文字识别等，现在看来很粗浅，但它们却是中国早期对模式识别的介绍。1978年改革开放伊始，中国科学院派出由北京自动化研究所和沈阳自动化研究所六人组成的代表团前往日本京都，参加第4届国际模式识别大会，从此开始模式识别领域和国际间的学术交流活动。

2. 语义模式识别、结构模式(或者句法模式)识别——模式识别的发展

模式识别一般来说可以概括为统计模式识别和结构模式(或者句法模式)识别两类，前者已经进行大量的研究，为人们所认识；而后者的研究先驱者公认是美籍华人、原美国普渡大学傅京孙(K. S. Fu)教授，他把模式识别和形式语言紧密地联系起来，生前出版与结构模式(或者句法模式)识别紧密相关的五本著作和大量的学术论文。遗憾的是，他55岁英年早逝。为纪念他对模式识别的贡献，国际模式识别学会专门设立“傅京孙奖”，以奖励对模式识别做出贡献的科学家。

戴先生曾经于1980年—1982年在美国普渡大学作为访问学者和傅京孙教授共同工作，并得到傅京孙教授的指教和帮助。戴先生的“Semantic Syntax—Directed Translation

for Pictorial Pattern Recognition"一文收录于第六届国际模式识别大会论文集，他提出，语义和文法间的关系很重要，并致力于语义方法、句法方法的研究。1982 年归国后，又结合汉字识别的研究，把结构模式(或者句法模式)识别方法拓展为语义方法、句法方法，并应用于汉字识别的研究。1985 年在美国 Palo Alto 举行的中美双边学术研讨班上，戴先生做了题目为《对复合汉字识别方法》的报告，论述语义方法、句法方法，并且证实对联机手写汉字识别的有效性，得到傅京孙先生的首肯。

3. 在线手写汉字识别的理论和方法的建立

1985 年中美双边学术研讨班以后，戴先生一方面把结构模式(或者句法模式)识别加以扩充，应用语言的语义信息，减少句法的复杂程度，如对于联机手写汉字识别应用模糊属性有限状态自动机，从而大大简化类型汉字识别型问题。本文集的文章 11、文章 14、文章 20 均有论述。另一方面，戴先生提出应用数性文法把统计模式识别和结构模式(或者句法模式)识别两类方法的结合，形成语义模式识别、结构模式(或者句法模式)识别。统计模式识别和结构模式(或者句法模式)识别则是语义模式识别、结构模式(或者句法模式)在一定形式下的特殊体现。文章 6 至文章 20 中介绍了这些研究成果，形成了本文集的核心内容。

汉字是由一些子模式，进一步是由笔画组成，结构的信息很重要。应用语义方法和句法方法，通过人工神经网络处理基本元素和子模式间的关系，从而奠定了汉字识别理论基础，并且找到了有效处理的方法。戴先生在 1984 年第七届国际模式识别大会上报告了该研究成果(文章 13)。其后，戴先生提出了在线手写汉字识别的方法，并在 1988 年第九届国际模式识别大会进行了报告，该报告作为一章(文章 20)被收录在 *Syntactic and Structural Pattern Recognition—Theory and Application* 一书中(1989 年由世界科学出版社(World Scientific Publishing Co . Pte . Ltd)出版)。这些工作取得了两个方面的成果，一是奠定语义模式识别、句法模式识别的基础，二是推动中国对在线手写汉字识别的产品化。

4. 学科交叉研究、综合集成的思想在模式识别中的应用

众所周知，以钱学森先生为代表的中国科学家在系统科学方面做了有益而重要的贡献。1990 年为构建系统科学中的基础学科——系统学，钱学森等提出了"开放复杂巨型系统和处理这类问题的系统方法论"，即"从定性到定量的综合集成法(Metasynthesis)"问世。戴先生参与了这方面的工作，并且把综合集成的构思应用于处理类型模式识别型问题的设计和应用。戴先生与他的学生、同事一起，按综合集成的思想，在离线汉字识别(3755 类汉字)方面做出了出色的工作。本文集的文章 28 至文章 30 就是类型识别的例子。戴先生所提出的汉字识别的学术新思想在许多领域得到了广泛应用，并形成产品，进而发展为中国汉字识别技术产业，这是戴先生的重要贡献之一。

本书所展示的研究成果凝聚了戴汝为先生默默的耕耘和长期的学术追求，虽然他已到耄耋之年，依然活跃在学术前沿，提携青年学子。这本书的出版一定会促进我国文字识别和语言识别的研究。

2011 年 3 月于西安交大

目　录

图像识别*

戴汝为
（中国科学院自动化研究所）

1. 引　言

图像识别(Pattern Recognition)或称模式识别是近二十年发展起来的一门技术科学。这里所指的图像具有广泛的含义，不仅仅是对图片或相片，还包括对文字、不同形状的物体、图形、声音的自动分类，以及医学与工程方面的诊断问题。关于分类，大致有两种情况，一种称为聚合(Clustering)，就是只告诉若干对象和它们的性质，把特性相同的归为一类，事先往往并不知道究竟有多少类。另外一种是把对象、性质以及对象所属的类型都告知，然后，对于一个新的对象分析它的性质，决定它属于哪一类，这个过程叫做学习过程。

这门新技术属于所谓人工智能(Artificial Intelligence)的范畴，它包括下面内容：建立大脑中枢神经系统的“神经网络模型”、“神经元”模型和自动机模型，用计算机进行原理验证、博弈、制定策略和决策等等，对文字、声音、波形和图形自动识别，用自然语言和计算机进行人机联系，研制具有识别、分析与综合等功能的系统等等。它们的共同特点是着重于实现部分脑力劳动自动化，因此对实现四个现代化有重要意义。例如人类生活在地球上，为了了解地球表面的资源，从而有效加以利用，发射地球资源卫星、发展航空遥感就需要解决处理和识别大量遥感图片的问题。利用文字识别，通过识别信函上的城市地名代码，用信函分拣机来自动分信，既减轻了劳动，又大大加快了速度。目前癌症对人的威胁较大，能够把癌细胞和正常细胞加以区分的细胞识别装置将为癌症的普查提供新工具。一个机械手配备上能够把周围物体加以区分的识别系统，就会大大增加功能和使用价值。另外随着计算技术的发展，复杂系统中使用计算机是必然的趋势，信息处理的要求日益提高，人与机器“对话”进行信息交流就比较突出，把声音、图形、文字作为输入，加强人机联系显得十分重要。

设计识别系统还得依赖于数学方法，它可以概括为两种。一种称为统计方法(或几何法)另一种称为语言方法(或结构法)，前者着眼于找出能反映图像特点的特征度量。把图像的数据进行信息压缩，以抽取特征，注意到所选的特性对于通常遇到的干扰和畸变来说尽可能具有不变性，或者很不敏感。如果抽取 N 个特征能够基本描绘原来的图像，那么图像就可以应用 N 维向量代表。对图像分类就相当于把特征空间划分为若干部分，当输入一个图像时，就根据相应的特征向量属于特征空间中的哪一部分而决定属于哪一类，如图 1 所示。

以往十多年来，统计方法研究得比较充分。语言方法是近来正在研究的新方法，对于某些

* 本文载：国外科学，1977，第 1 期，52－58

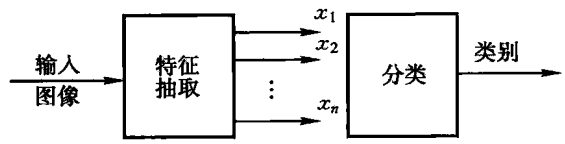
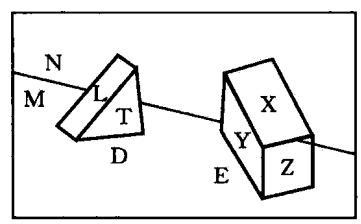


图 1

图像识别问题，描绘图像结构的信息非常重要，不仅要求判断图形属于哪一类，而且能够描绘使该图形不属于其他类的性能。图片识别或者更广泛的景象分析(Scene Analysis)就属于此，在这类问题中，面对的图像非常复杂，所需描绘它的特性度量的数目非常大，于是着眼于借助简单子图像，如何形成复杂图像的描述，因而，识别的要求也只能是每一图像满足一种描述，而不是简单的分类，如图 2 所示图画 A 可以应用相应的结构体系加以描述。



图画 A

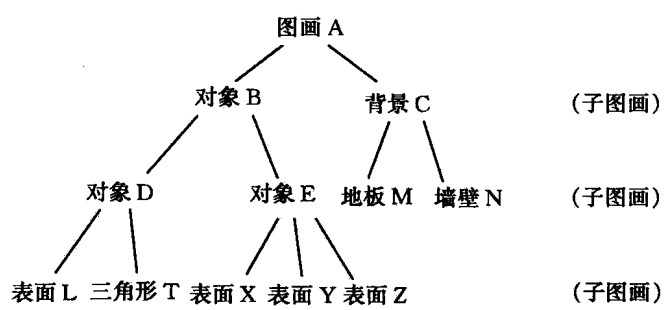


图 2

图像描述着重于图像如何由比较简单的子图像构成，这些简单的子图像又如何由更加简单的子图像构成等等，最基本的子图像称图像元素(Pattern Primitive)。这就类似语言的组成，例如英文句子由短语组成，短语又由单词组成，一个图像相当于由某种文法规则组成的句子，图像元素就相当于组成句子的基本单元即单词。由图像元素构成图像的规则，就相应地叫做图像描述语言的文法。用语言方法，首先就是识别图像元素，然后图像识别的过程就是剖析描述该图像的句子，确定它对于特定的文法来说，在语法上是否正确，识别图像元素当然要比识别图像本身容易得多。这种方法考虑用简单图像元素的集合加上文法规则，而文法规则可反复应用，于是就可以描述大量复杂图像，由于语言方法还处于研究初期，目前能解决的问题仍非常局限，还有待进一步研究与发展。

当针对一项有关图像识别的任务进行系统的设计时，一般说来，包括着三方面互相有关但有明显分界线的过程，即数据获得、图像分析以及图像分类三个过程。(1)数据获得：使物理对象所包含的信息，通过一定的转换装置，变成数字计算机接受的信号，以便做进一步的处理，各

种敏感器、扫描装置，都是为达成这一任务而采用的。(2)图像分析：数据获得后，就要对数据加以分析，包括进一步数据处理，进行信息压缩。典型的工作是特征抽取、特征选择以及聚合的分析等等。(3)图像分类：对于新的对象，判断它应该属于的类别，按照满足某种性质指标从而设计识别逻辑，通常就是分类过程的关键所在。图像分析和分类的工作可以借助计算机来进行。当然，人是主要的因素，设计一个系统、抽取什么特性、识别逻辑的优劣，都由设计人员决定。为了能充分发挥人的积极因素，并有效利用计算机作为工具，这就需要配备适当的输入输出设备，增强人机对话功能，做到输入灵活，能及时修改，很快就能够观察到效果。甚至于经计算机验证，确定了方案，就可以用适当硬件设备加以实现。

图像识别的范围很广，下面着重讨论几个比较重要方面的情况，包括文字识别、遥感图像识别、声音识别、生物医学图像识别等。

2. 文字识别

文字是人们交流信息的重要工具，欧美等国习惯用打字机，对于识别打印体的字符，进行过不少研究，为了便于机器识别，曾经把使用广泛的英文字母及数字改成特殊字体，按照各个字符黑白笔道的分布情况，分别设计“模板”转变成电的信号存储起来，当输入新的字符后，与所有各种模板匹配，以符合的程度最大判定该属于相应的类。只要稍加注意就会发现，平时用的字符，往往包括着各种风格的字体，为了能有效地使用，必须解决识别多字体字符的要求。通过十多年的发展，国外已经研制出各种类型光学文字识别机，简称 OCR，开始用于邮政、银行、商业、航空线以及印刷等方面。此外，研究识别手写体字符有更大现实意义。由于不同的人使用不同的书写工具，手写体字符的字形变化非常大，识别的困难也大。下面准备对能够识别 10 个手写数字系统做一扼要介绍。写在规定矩形框内的数字，字体可以比较自由。数据的获取是用示波管扫描，如图 3 所示。

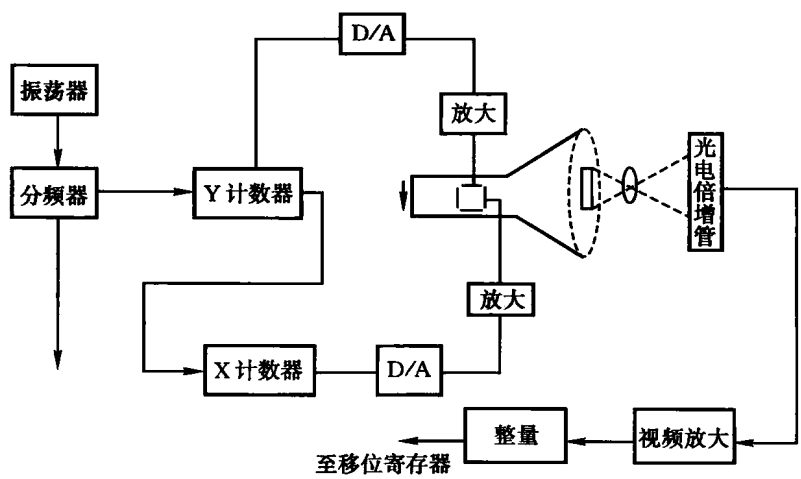


图 3

示波管以 X＝18，Y＝30 的光点矩阵进行扫描，几经聚焦投影在文字上并用光电倍增管接收反射信号。经视频放大的视频信号是随着纸面对光的反射程度而变化的模拟信号，在背景处

反射率高、视频信号大，在笔道处反射率低、视频信号小。为了将连续的模拟信号转成“0”信号、“1”信号，需要具有一定的阈值以作标准，经整量后的信号送到按平面排列 X＝18，Y＝30 的 540 位移位寄存器（寄存器由低速组件做成），这相当于一个工作区。信息反复地从中取出经过适当处理，再重新存入寄存器中。处理是逐步进行的。首先对存入的字进行平滑，也就是根据每一个点周围若干个点的情况，确定该点究竟是“0”或“1”。平滑的目的在于去掉纸面上孤立的污点，补上文字笔划中的飞白点，补齐文字边缘的小凹凸以及保留文字中间原有的小孔。平滑之后，进行细化。通过细化逻辑逐步去掉文字的边缘点，直到获得宽度只有一个点的文字骨架。根据文字骨架的端点节点数，把文字进行粗分类。在粗分基础上，从端点到端点沿着骨架进行跟踪。每跟踪一步，就得到这一步的八个方向特性，由于这一步和以前所构成的凹、凸、平三种线特性以及所到达点的点特征（分成端点、节点和连点三种），共计 14 种特征。其结果是把二维平面文字图形变成了一串特征序列，这在一定程度上体现了语言结构方法的思想。

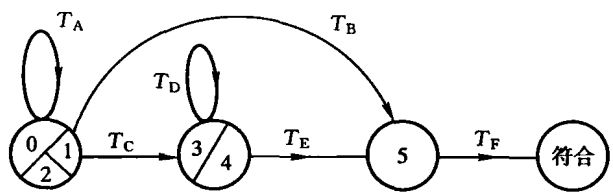


图 4

识别的过程主要考虑特征序列是否能通过预先设计的顺序逻辑。如图 4 所示圆圈为所处的状态，由序号小的标号加以检验。带箭头的连线表示从一个状态到另一状态的转移。转移条件 T 可以有适当的变化范围以适应字形的变化。识别开始时，认为处于标号为 0 处，并从端点到输入文字的骨架进行跟踪。每跟踪一步，抽取一组特征作为输入加以比较。如转移条件符合，则转移至另一状态。如不符，再检验下一个标号的有关转移条件是否能转移。如检验至该状态的最大标号仍不符合转移条件，说明逻辑不能通过。在检验中，只要进行了转移，则考虑处于所转移到的状态，然后再跟踪走一步，重复从该状态检验，是否能转移，最后判定能否通过这一顺序逻辑。实现时采用了微程序控制，这是因为一方面具有较大的灵活性，另外软硬设备可以同时进行。微指令和顺序逻辑放于只读存储器中。如果要改变逻辑，就相应更动只读存储器穿线即可。整个系统方块图如图 5 所示。

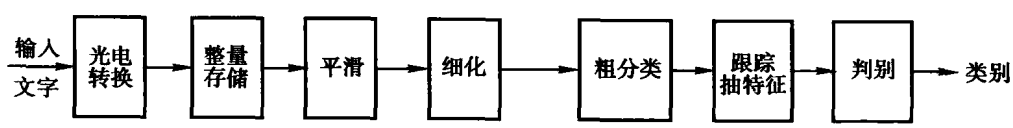


图 5

可以看出，在整个过程中，平滑、细化等处理比较费时间。顺序逻辑是逐条地去加以检验，也花费时间。对于自动分拣信函，每秒钟分三封信，信上有 6 个字符，即每秒识别 18 个字符可以满足要求。用上述方法，采用速度快的元件是可行的。

在文字识别的发展过程中，曾经希望研制高性能的文字识别机，作为快速电子计算机的输入，代替目前最常用的穿孔纸带输入。这就牵涉到速度、识别率、可靠性等技术指标。对于打印

体字符国外已经研制成指标相当高的机器，有的每秒可达2000字符，错误率为10^{-6}。然而这种装置本身就很庞大，价格非常昂贵，难以达到作为输入的希望。对于研制OCR，如何才能够达到推广使用是特别值得注意的。至于手写体字符，例如识别FORTRAN所用的50个字符，要达到较高的指标是不容易的，甚至还缺乏有效的方法。对文字进行分类主要是根据形状，通过形状所能抽取到的仅仅是为数不多的几何特征，要用来达到数目很多的分类就更不容易了，用统计方法难以取得效果。用语言方法，曾有过令人注目的试验结果。一种称为Cyclop—1的系统，考虑把文字之间各部分有效加以连接。为区分用光笔写的较规整的英文字和数字，大约包括42个基本线段。文字的描述方式有线段的长度和总长度之比、线段端点间与线段长度之比、线段与X轴的夹角以及线段的反弯点等。试验表明，能把重叠在一起的字区分开来。另外前面谈到过把文字图形通过细化后跟踪化为特征序列，然后通过适当文法规则来描述不同类的字，似乎可以解决问题。问题在于研究得比较充分的短语结构文法，往往难以解决字形变化的问题。根据近年来图像文法的研究表明，对于手写体字符，有效地描绘规则往往与图像元素的坐标有关。这已经超出了古典形式语言的范畴，需要发展和建立新的体系。

此外，日本和东南亚的一些国家也曾开展了汉字识别的研究。经过较大的努力，日本对于一种印刷字体881个汉字，用多级图形匹配方法，达到错误率低于10^{-6}，拒绝率为10^{-3}的结果，并研究实现多级匹配的专用硬设备。

3. 遥感图像识别

为了有效地利用和开发地球表面的资源，可通过安装在飞机或者航天飞行器上的遥感装置，以获得有关地球表面各种对象物理特征的信息。在农业方面，可以了解耕作面积、灌溉和施肥情况、病害虫害检查、农作物的分类等。在地质水文方面，如矿藏、石油的开发、河流泥沙淤积、地下水的调查、污染检查和测量、地形识别，另外森林防护、土地使用、水库和水坝调查等都与遥感技术密切有关。从目前已知的情况来看，在遥感方面应用图像识别是比较有希望的，利用不同物体反射和辐射不同光谱强度的特征，从而区别开这些物体，这是多波段遥感的基本思想。数据的获得是从高空或空间用多波段遥感仪器对地面摄影或扫描，则对应于地面景物有一点在不同波段上有一组光谱强度数据。这些数据在与波段数相等的多维度量空间中组成对应的一个点，因为环境条件的不同，如太阳照射角不同、大气途径的变化等，虽然是同样的地面物体，所反映的光谱强度却仍然有差别。因之在度量空间中虽属同类物体就不只对应于一个点，由许多点群聚在一起称为一个集群。这里就要考虑怎样把代表不同类别物体的不同集群间的分界面确定下来，并进而决定当输入一个未知的新样本，应该属于哪一个集群，这就是遥感方面遇到的分类问题。由于多波段遥感扫描装置所得到描绘对象的向量，一般说来维数都不算高，因此分类问题可以有效地应用统计方法来介绍。方法大体上可分成两种，一是监督学习法(Supervised Learning)，另一是非监督学习(Unsupervised Learning)，也即聚合法。在监督学习法中，首先要从被探索的地面区域中选定一些训练区，在这些训练区中，实际存在物体是已知的，它们的光谱特性可以实测得到，这就是所谓地面实况资料。从空中对这些训练区进行遥感，得到图像信息再经过数据处理，对应地面的实况，得到再度量空间中各种类别的集群所占的地位及形状以及它们的分界面。从训练区所得的结果再外推到未知区域进行判断，一个未知样本在

这个空间中所占的地位与其他已知类别集群间的关系，可以推定它应该属于哪一类。因为采用了训练区的已知资料，用计算机处理数据时，能先进行学习来识别不同类别的物体，所以称有监督学习，由于通过地面实况资料确定出一类的统计特性，每一个波段的光谱强度数据可以用直方图(Histogram)表示。值得注意的是在有些情况下，直方图是单峰分布的，于是可以用多度量高斯密度函数来描述对象所对应的特征向量。两个主要参数平均向量和协方差矩阵由训练区的样本得到，如果能够求得概率密度函数，那么通常就用典型的贝依斯判别法。它的判据是，如果

$$P(d/A)P(A) > (d/B)P(B)$$

则判定 d 点属于 A 类而不属于 B 类，其中 $P(d/A)$，$P(d/B)$ 是度量点 d 属于已知类 A 或 B 的条件概率。$P(A)$，$P(B)$ 是 A，B 可能发现在这个区域的先验概率。对于不止两类的多类判别来说，就需要两两加以比较来决定。

非监督学习法是利用不同物体的多光谱强度所形成的度量点，在多维空间中必然占据不同位置这样一个特性，也就是形成不同的集群，因而先对图像数据进行处理，在空间形成不同的集群后，再和地面实况核对，确定这些集群代表哪些类别的物体。要把图像信息归纳成多维空间中不同集群，通常利用多次迭代方法。先假设一初始情况，计算各种统计参数，计算每一点和其他点的距离。如果距离小于某一值，就认为是属于同一类别的集群点。如果集群过大，有时还需要把它分成两个。而两个过小的集群有时需要合并成一个。这样经过许多次迭代，计算结果趋于收敛稳定。如认为结果满意，就可以和地面实况核对，看某一集群代表什么物体。这类迭代方法已经编制成标准程序，如所谓 Isodata。

监督学习法和非监督学习法相比较，前者的缺点是(1)需要预先对所观察的地区有大体了解，才能选好训练区作为典型。(2)训练区的详细情况必须事先搞清楚。(3)这些训练区的情况随着条件不同而有差别，因而不是任何时候都适用，相比之后者就没有这些缺点。它不需要事先知道被观察区域的情况，而且等计算有了结果后可以更加好地根据结果选择具有代表性的地区作为核对和判别的根据，计算结果的正确性在条件可能下也不逊于前者。

为了增加分类判断的正确性，图像信息数据常常需要事先作预处理，除了校正图像的误差之外，重要的一步是消除各种因素，如云影、地貌变化、照射角差别等环境条件引起的图像中辐射强度的变化，常用的有比例成像法，也称为信息延伸技术，即是将每一段的辐射强度与相邻波段或全部波段的辐射强度相比等，经过这类处理，判断正确性增加，训练区也可以少用。

随着多波段扫描遥感的发展，波段数目已大为增多，最多的已达到 24 波段，进行分类处理时，如所有波段数据都参加运算，则信息量将大为增加，加重计算机的负担。有一种处理就是在不影响分类能力的条件下，减少输入信息的波段数，以便降低多维空间的维数，较有代表性的是主分量变换法，因为多光谱信息在各波段间相关性是很高的，一般在多维空间中大多形成狭长的分布，所以只要把数据分布向比较集中的几个轴变换一下，其他一些轴上数据分布就大大减少，可以忽略。这样选择头几个轴的数据，维数就降下来了，减少了运算的时间，如果降为二维，那么就表示成平面上的点，也就可以较直观地用显示装置把各类数据加以显示。另外一种降低维数的办法是非线性映射。考虑用低维空间的点代替高维空间的点而保持着互相间距离关系，这种思想可能来源于画地图，把球面上的点映射到平面上。具体处理可以归结为求极值最优化问题，然后应用反复迭代的方法来达到降低维数的要求。

4. 声音识别

交谈是人们传递和交流信息的另一种方式。谈话比写字等方式要方便，效率也高，这是人所共知的事实。人与计算机进行信息交流，研制方便灵活的人机对话系统，声音识别无疑是一个重要的方面，声音识别的数据获得过程相当简单，只要用一个话筒把声波通过模拟数字转换送入计算机，这对于推广使用是有利的条件。

国外开展声音识别的研究已经近三十年，以往进展甚微，近来才有了较显著的成绩，从系统方面大致可分为“孤立字音”的识别系统。孤立字音指的是作为一个单元的语调，或一个短语的语调，字音和字音间是分开的，具有一定空隙。1973 年以来，已经制成作为商品的孤立字音识别机，开始用于电视面板等检验，包括分拣、数字控制机床等方面，其成本较低，而且还努力进一步下降，另外国外还投入相当力量研究连续声音识别系统以及“理解系统(Understanding System)”。连续声音指的是字音之间没有明显的空隙，识别时首先要设法把字音区分开来。理解系统处理的也是连续声音，主要在于还采用了有关语言含义的信息，这种系统往往着重完成某种特定的功能。看来有关识别连续声音的研究，不仅仅是做出实用装置，或是获得人机通信的有效工具，而且还有重要的科研价值，因为它反映了机器识别的一些有代表性的基本问题。

以往声音识别成效甚微的原因之一是信息压缩问题解决得不好。声音信号属于非平稳随机过程，描绘它的办法是在一段短时间内，确定能量、频率和时间三者的关系。如何通过少数几个量以表征原来的波形不是容易的事。近来提出了一种信息压缩的普遍方法，即用线性预测系数(Linear Prediction Coefficients)作为特征。这种方法计算方便，效果明显。所谓线性预测其实是维纳滤波(Wiener Filtering)的一种形式，最初是在有关炮瞄系统设计中为了对目标进行预测而确定伺服系统的参数开始得到应用，对于诸如脑电波、震动波、声波等都可应用，主要是把所处理的波形考虑为由某种模型所产生的信号，声音信号可以认为是一串近乎周期的脉冲或噪音源作用在一个滤波器上产生的信号。信号经过适当的采样，并考虑用过去 p 个时刻采样值线性组合预测下一时刻的值。如果这样的预测能够获得较好的效果，那问题就归结为确定 p 个系数。预测值与真正的值自然有所不同，可以按使平均平方误差最小来确定 p 个系数(一般来说，这些系数是时间的函数)。这 p 个参数就可以描述该段连续波形，叫做线性预测系数。由于声音具有非平稳的特性，需要分成若干段，逐步处理。在一段短时间内，估计系数可以看成常量，由 p 个系数组成向量就作为论述该段信号的特征，由各段的特征综合起来就得到声音信号的特征，例如一个长约 2 秒的声音，加以处理时，先通过低通滤波器保留 5k 以下分量，然后按每段长 50 毫秒当成 40 段。用 10k 速率采样，每段可得到 500 个采样数据。用它们作为度量直接可表示原信号，但维数太高。试验表明，通过这些数据求出线性预测系数，$p=12$ 时大致可以满足要求，这样维数大大下降，起到信息压缩的作用。整个长 2 秒的声音，共计由 40 组系数加以描述，成功地解决了把一个连续波形应用一些常数表征的要求，这样一来识别声音就可以用统计方法来解决。直观的考虑是用模板匹配，匹配需要逐步加以考虑，计算总的效果。由于模板的设计时间长度是固定的，不同的声音长短不同，分段数目也不一样。当一个声音输入，要与这些长短不同的模板匹配，在时间轴之间必须有适当的转换，才能判定输入究竟和哪个模板匹配的程度为好，简单的办法是两个时间轴之间用线性转换。另外匹配过程实际上是最优化过程，可用动

态规划(Dynamic Programming)做工具,动态规划是研究多步决策过程,使指定指标达到极值的计算方法。它是利用不变嵌入原理,把一个特定的极值问题嵌入一组性质类似的极值问题中加以解决,给出反复迭代的计算公式。当维数低时,用计算机可以求得最优策略,这种方法近来有效地用于识别声音,对于多个音节声音效果较显著,对于时间轴的转换用线性转换及动态规划如图 6 所示。

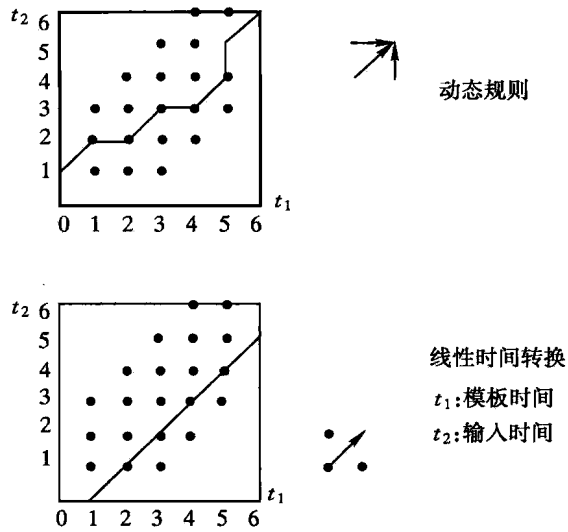


图 6

以上讨论表明,从声音分析到识别孤立字音,用线性预测系数和动态规划,使用的是一般性方法。问题是处理与识别所花时间太长,还需要进一步加以解决。当需要识别的声音数目多的时候,模板的数量相应增多,复杂程度增加。考虑到人们所用的语言是由"语音(韵母、声母、元音及辅音)"拼成,所以很自然地把语音当成基本单元,如果能识别语音,而语言就是由一串语音(或语音链)构成,通过拼音的办法来识别声音,这样做的好处一是通过识别语音可以进行粗分类,有选择地取出模板来加以匹配。二是利用语音使存储模板时需要的存储单元减少。当分类数目非常多时,优越性就显示出来。这样看来,似乎把语音作为基本单元,有可能解决语言识别问题。其实这只能解决部分问题,因为需要鉴别一个语音的信息通常不是只体现于该语音自身。由于舌头和其他发音器间物理惯性,语音往往受到它邻近语言的强烈影响。例如英语中辅音"t"在 twin 和 tin 中的发音就有所不同。前者唇收圆,后者唇分开。要鉴别一个语音,通常不是只取决于该单元本身,而需要了解它前后的情况,才能决定。这就是开头就谈到的基本问题,即局部含糊之处必须通过了解全局信息才可能加以消除。当一个人说话时,即使他的发音不准确,但根据上下文,听话的人还是能够听得懂。要是设计出来的机器具有这样的功能,这就是颇不容易的事了。如果把语音当成基本单元,那么连续的声音就变成语音链,于是语言识别就着重分析由基本单元组成的链。其实图像识别的语音方法就是由这点出发,即人工语言中借助一些概念,形成一个新领域。然而带有讽刺意味的是,以前两者间似乎还互相不相干,这也从一个角度表明连续声音识别以及语言方法是图像识别中非常值得注意和研究的方面,许多问题还有待于进一步地加以解决。