图象识别导论

程民德　沈燮昌等　编著

上海科学技术出版社

图象识别导论

程民德　沈燮昌等 编著

上海科学技术出版社

图象识别导论

程民德　沈燮昌等　编著

上海科学技术出版社出版

（上海瑞金二路 450 号）

新华书店上海发行所发行　江苏扬中印刷厂印刷

开本 850×1156 1/32　印张 10.75　字数 282,000

1983 年 7 月第 1 版　1983 年 7 月第 1 次印刷

印数 1—7,800

统一书号：13119·1075　　定价：（科五）1.35 元

内 容 简 介

本书从统计判决、语言结构法、模糊集论三方面提供了图象识别的理论基础。

第一章介绍了图象识别研究的对象及方法，它是本书的引论；第二章到第四章介绍了统计图象识别中的一些基本方法及理论基础；第五章介绍了图象识别的语言结构法；第六章介绍了用模糊集的方法进行图象识别。

本书可供从事有关图象识别的广大工程技术人员及科学研究工作者参考，也可以作为高等院校中有关专业的教科书或参考书。

序

近二十年来，图象识别这一课题在理论研究和工作实践中都有了迅速的发展。这一方面是由于图象识别的应用范围已渗透到国民经济的许多领域，另一方面则是由于计算机科学的飞速发展使这一应用的可能性也越来越大。除了早期的文字识别及语音识别以外，目前它在天气预报、质量控制、国防科学、指纹识别、遥感技术、地震探测、疾病诊断、细胞识别等各方面都有广泛的应用。

目前，在国际上，图象识别这一课题已受到极大的重视，有许多不同领域的科技工作者都在从事于这方面的研究。最近每年都有图象识别以及与它有关的图象处理、人工智能的专业性会议，并出版了各种会议的报告集以及有关的书籍。目前世界上有几十种杂志登载图象识别以及有关领域的文章，并且还有若干有关图象识别的专门杂志。在杂志"Computer Graphics and Image Processing"上每年都载有图象识别及有关领域全年发表的文章总清单。现在，国际上已成立了图象识别及人工智能的专业委员会。外国有的厂商已造出了图象识别的专用机。

图象识别的研究牵涉到很多学科，为了实现识别，需要综合各有关学科的知识和技术。仅就数学这一学科而言，就需要线性代数、矩阵论、规划论、信息论、函数论、概率统计、数理逻辑、形式语言等各方面的知识。本书主要论述图象识别的基本数学理论和方法，重点放在统计图象识别方面，同时也介绍用语言结构法及模糊集方法来进行识别。本书是近几年来在北京大学举办的有若干兄弟单位参加的一个讨论班的基础上写成的。尽管本书内容的取材及写法可能有许多不完善及不妥之处，但是由于国内有关介绍这方面的书籍很少，因此编者希望本书能向读者提供一个在图象识别方面的初步介绍。

本书第一章介绍了图象识别的研究对象、基本概念及几个简单方法。这一章是由沈燮昌及周民强两同志合写的。第二章着重讲 Bayes 判决及其应用，同时介绍了 Neyman-Pearson 判决方法。第三章介绍了统计图象识别中的几个判决方法，如 Fisher 判决、Wald 判决、概率密度函数的估计、一些非参数方法等。这两章是由沈燮昌同志执笔写成的。第四章介绍特性提取与特征选择的一些方法，主要是由程民德同志执笔写成的，其中§4.7由石青云同志执笔写成。第五章介绍用语言结构法来进行图象识别。第六章介绍模糊集方法在图象识别中的应用。这两章是由钱敏平同志执笔写成的，由沈燮昌同志作了若干补充，并由石青云同志进行了校正。在一些章节后面加了附录，以便读者进一步查阅。我们在编写过程中曾进行多次讨论，最后由沈燮昌同志对全书作了统稿工作。

复旦大学唐国兴同志为本稿提了不少宝贵意见，谨表示感谢。

限于我们的水平及时间仓促，不足之处可能很多，欢迎读者批评指正。

编　者

于北京大学数学系 1982 年 5 月

目　录

第一章 导 引

§1.1 图象识别简介

图象识别(Pattern Recognition，也称模式识别，这里沿用习惯名称)，粗略地说：就是要把一种研究对象，根据其某些特征进行识别并分类。例如要识别写在卡片上的数码字，判断它是0，1，2，…，9中的哪个数字，这就是将数码字图象分成十类的问题。因此，这种识别早已存在于人们的生活实践中。然而，随着实践活动的扩大、深入和更加社会化的需要，人们不仅需要识别分类数很多的事物，而且被识别的对象的内容也越来越复杂。例如邮局每天需要识别大量信件上的编码，以便送到各个地区(分类)；又如对某地区数十万人口进行某种疾病的普查，以便进行预防和治疗等。特别是由于科学技术水平的提高，可以使得各种不同的研究对象"图象化"或"数字化"，也就是说，可采用某种技术把考察的对象转换成照片(如各种高空照片及X光照片)、波形图(心电图、地震波等)以及若干数据(遥测遥感中用多光谱扫描所得的数据)，这些数据就可以代表所研究的对象。因此，"图象"一词的含意决不只是指通常意义下的图或照片。人们自然希望采用各种仪器及设备来代替繁重的劳动，并且能够多快好省地进行图象识别。在有了大型、快速电子计算机的今天，数值化处理的手段已显示出很大的优越性。因此，数学的理论和方法已日益显示出它在这个领域中所起的作用。

我们所研究的对象——图象是千差万别的，它们都蕴含有本身固有的特性，因此，有可能把它们区别或分类。所以除了对图象进行"数值化"以外，还需要通过一些手段，将各类图象的重要特性

用数字刻划出来，这就称为特性提取。实际上，反映一类图象特性的数目往往是比较多的，这样，一方面在用计算机处理时，也必须花费很多时间，另一方面，由于这些特性的提取往往是不精确的，会带有一定的误差。因此，有必要对这些特性进一步进行选择，使得尽量设法去掉一些误差，而又保留原来特性中的信息。这往往是利用原来的特性，通过一些方法，找出某些(比原来特性数目要少)综合性指标来，这就称为特征选择。这样一来，每个图象就由一组数来表示。进一步的问题，就是要设计识别方案，使得对任何一个未知类别的图象，根据方案就可以判定它属于哪一类。由于同一类图象往往可以用不同的一组数来表示，也就是说，这组数往往是随机的。因此，就可以用统计方法来设计识别方案。有时也用统计方法来进行特征选择。此外，也经常用已知类别的一些图象(样本)来设计识别方案，使得这个方案对原来已知类别的图象能正确地识别，或者在某种意义下使得错误识别的可能性最小。这就是统计图象识别的主要思想。

现在我们举例说明：如在细胞识别中，需要判断哪些细胞是癌细胞，哪些是正常细胞。这就是一个两类图象的识别问题。根据医生的经验，癌细胞有一系列的异常情况：如(1)细胞核大；(2)细胞核染色增深；(3)细胞核形态畸形(正常细胞核呈圆形或卵圆形)；(4)核浆比倒置；(5)核内染色质出现粗颗粒，结成团状或有核膜、核仁，染色质分布不均匀；(6)整个细胞呈长条(纤维状)、串状等各种畸形。这样，我们把每一个细胞放入某一类仪器中，这个仪器能按一定的间隔测出细胞每一点附近的透光值，称消光系数。如果在每一个细胞的纵向与横向都取19个点，这样就得到361个数据，这就是细胞的"数值化"。然后，根据上面的六个特性，用一些方法，从这361个数据中设法找出每一个特性的数，这就是特性提取。例如，可以通过消光系数值的大小及分布，判断出哪些是细胞核，哪些是细胞浆。计算细胞核所对应的消光系数的点的个数就能知道细胞核的面积，这个数就可以反映第一个特性。也可以用细胞核边缘的周长平方与其面积之比来刻划它的畸形情况。

这是因为在周长一定的情况下，圆的面积最大。这里取周长的平方是为了表示与面积的量纲一样。当然也可以取别的方幂。这样，又得到了一个数，它刻划了第三个特性。最后，每个细胞就对应着刻划这六个特性的六个数，这些数构成一组有次序的数或向量。从这六个特性可以看出，它们彼此之间不是孤立的，是有一定关系的，因此，没有必要用这六个特性来进行识别。可以用一些方法（往往是用最优化方法及统计方法）从这六个特性中选出两个综合性指标。这样，一个细胞就对应着两个有序的数 x_1、x_2，它构成平面上一个向量 $x=(x_1, x_2)^T$（今后我们认为向量是列向量，T 表示转置）或在平面坐标系中对应着一个点 (x_1, x_2)。很多细胞就对应着平面上的很多点。

现在我们取一批已知类别的细胞；如取 100 个癌细胞及 100 个正常细胞（称为训练样本），它们在平面上就对应着两类点集。我们可以找一条曲线把这两类点分开（见图 1-1），这样整个平面就被这条曲线分成两个区域。对于任何一个未知类别的细胞，按上述特征选择方法也可以对应着一个点 $M(x_1, x_2)$，如果 M 落入区域 I，则就可以认为这个细胞是正常细胞；如果 M 落入区域 II，则就认为这个细胞是癌细胞。上述曲线就称为判决边界，其方程记作 $g(x_1, x_2)=0$；上述识别细胞的方法称为判决。从数学来看，可以认为：若 $M(x_1, x_2)$ 使 $g(x_1, x_2)<0$，则判决 M 表示正常细胞；若 $M(x_1, x_2)$ 使 $g(x_1, x_2)>0$，则判决 M 表示癌细胞。函数 $g(x_1, x_2)$ 称为判决函数。当然，这样找到的判决函数形状可能会很复杂，在计算时会带来很大的不便。我们也可以选择 $g(x_1, x_2)$ 为一个线性函数 $\bar{g}(x_1, x_2)$，而用 $\bar{g}(x_1, x_2)=0$ 作为判决边界，对前面给出的 200 个已知类别的细胞可能会有误判，此时，可以指定

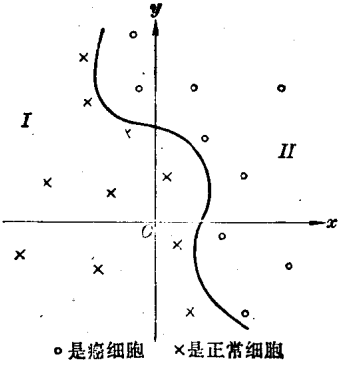
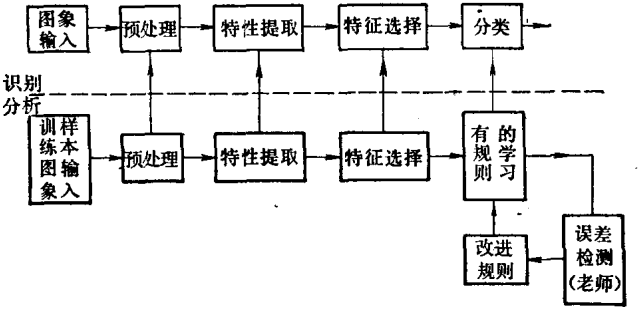


图 1-1

某个准则，使得在这个准则下，误判的可能最小。

现在可以用下列框图来描述统计图象识别的大致过程：



图中上半部分是识别部分，即对未知类别的图象进行分类；下半部分是分析部分，即对已知类别的图象样本制定出判决函数及判决规则（有规则的学习），使得对未知类别的图象能够进行分类。由于所输入的图象需要进行数字化，这就会产生误差；又如在高空中所摄的照片，由于大气扰动的影响或飞行器的移动等原因，都会使照片模糊；遥测遥感照片或多光谱扫描所得的数据也需要进行某些校正。所有这些都需要进行预处理。框图右下角部分是自适应处理部分。当用训练图象样本根据某些准则制定（学习）出一些判决规则后，再对这些训练样本逐个进行检测，观察是否有误差（这相当于请老师进行指导），如果有的话，再进一步改进判决规则，直到比较满意为止。

在一些图象识别中，往往需要了解的是图象的结构信息，且识别的目的不仅是需要安排图象属于哪一类，而且还要描写图象的形态。这方面的例子有指纹识别、场景分析等。近十多年来，用语言结构法来识别图象也有不少研究。由于一些图象的结构比较复杂，且特性的数目非常多，因此，要简单地判断它属于哪一类是不实际的。这样，自然要想到：能否将复杂的图象用一些相对比较简单的，图象的组合来表示，而这些子图象又用一些更为简单的，图象来表示，…，最后用一些最简单的，图象（称为基元）来表示，且所有这种表示又都按一定的规律组成。

例如考虑下面的场景 A(见图 1-2)，它是由一些物体及背景所组成，而物体又是由一个长方体及一个三角体组成；背景是由地板与墙所组成；长方形体是由看得到的三个面所组成；三角形体是由两个看得到的面所组成。这样，我们就可以逐级地描写这种结构如下(见图 1-3)。于是，这样一种逐级描写的结构方法与日常所用的句子分析有类似之处(见图 1-4)。当然，这里每一个字还可以再分解成一些字母的组合。这里的字用语法规则连接起来构造出短语，最后再构成一句完整的句子。对照一下上述的场景，取最简单的子图象(基元)，用一定的规则即可构成较为复杂的子图象，再根据一定的规律，可从子图象逐步地构成一幅场景。在句子中字与字之间有语法规则联接；在图象的基元与基元之间也有一定的规则连接，这种规则也可以称为文法。这种文法就称为图象文法。

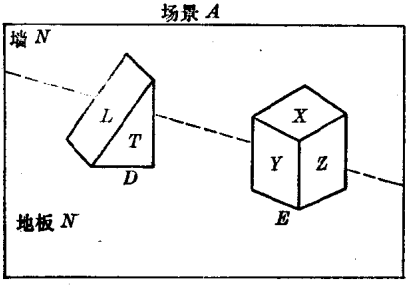


图 1-2

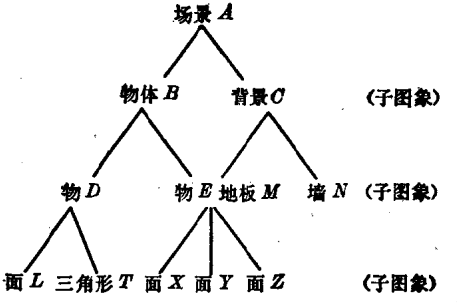


图 1-3

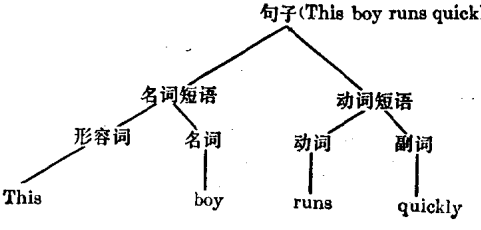
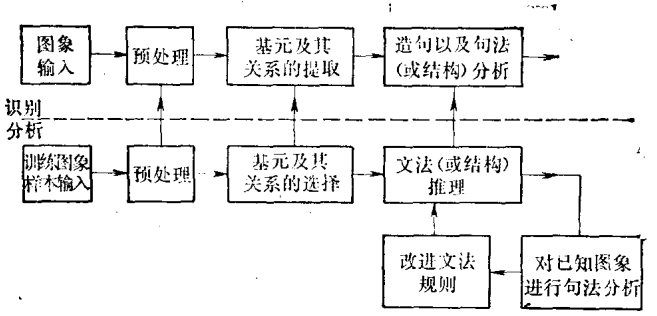


图 1-4

用基元及其关系(文法)能描述图象结构的语言称为图象描述语言。

语言结构法中图象识别系统的方框图如下：



这里,象用统计方法进行图象识别一样,也分成两部分;上半部分是识别部分;下半部分是分析部分。在分析部分中,用一些已知结构信息的图象作为"训练样本"构造出一些文法规则,再用这些文法对未知结构信息的图象所表示的句子(经常可以由字链所构成)来进行句法分析,这实际上就是识别。如果能够被已知结构信息的文法分析出来,那么这个未知图象也有这样的结构信息,否则,它就不是具有这种结构信息的图象。

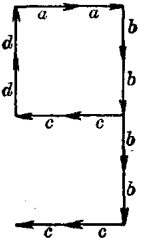


图 1 5

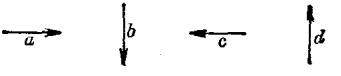


图 1-6

在基元及其关系的提取阶段类似于前面的特性提取及特征选择这两个阶段,当然在具体实现时是完全不同的。例如,我们要描写数字 9(见图 1-5),就可以用四个基元来描写(见图 1-6)。这四个基元是四个向量,其长度都一样(见图 1-6),而连接的方式只能有两种,即向量的首尾相连接。它的结构图如图 1-7。

写成的句子就是 $ccddaabbbbcc$，或简单地写为 $c^2d^2a^2b^4c^2$，这也称为链码。

对于手写体文字，一笔一划可以作为基元；在形状分析中，一定特性的直线段或曲线段都可以作为基元；在语音识别中，单音可以作为基元；…。

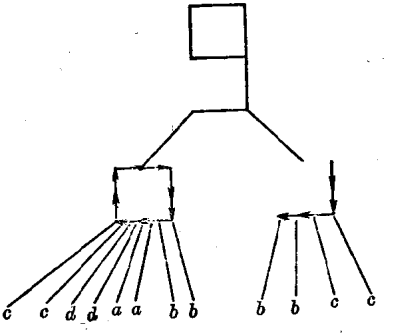


图 1-7

在选择基元时，如果选择得非常简单，其优点是容易把它找出来，但缺点是不易用紧凑的文法来描述一个图象；反之，如果基元选择得比较复杂，虽然容易用紧凑的文法来描述图象，但对基元本身却不容易识别。这二者往往是矛盾的，所以就需要适当选取，二者兼顾。有时可以用统计图象识别方法来识别基元，然后再用紧凑文法来描述图象。

有了基元后，必需对各种训练图象样本构造文法，这样才能产生语言，并用它来描述图象即造句。当然，最理想的是从已给定的基元来自动地产生文法，如对"桥"的基元能自动地推出生成"桥"的文法。目前，还不可能普遍实现。因此就需要根据预先的知识及经验进行人工编制。当编制出来几条文法规则以后，再用已知结构信息的图象来进行句子分析，若能够分析出是这种结构，则这几条文法规则是可采用的，反之，若分析出来不是这种结构，则就需要修改上述几条文法规则。这就是前面框图右下角一块的含意，这类似于统计图象识别中自适应修改判决规则的那一部分。对于一种文法来说，若规则很多，当然功能就大，但因此设备也大，花时间多，代价也大，而其优点是解决问题的范围也大；反之，如果功能小，则很多图象就无法描述。

例如，对前面的句子 "This boy runs quickly" 可以引入文法（即改写规则或生成规则）如下：

P: 〈句子〉→〈名词短语〉〈动词短语〉

〈名词短语〉→〈形〉〈名〉

〈动词短语〉→〈动〉〈副〉

〈形〉→ This　　〈名〉→ boy

〈动〉→ runs　　〈副〉→ quickly

其中〈句子〉是开始符，它连同〈名词短语〉〈动词短语〉〈形〉〈名〉〈动〉〈副〉都称为非终止符，而 This、boy、runs、quickly 这四个字称为终止符。

对前面的数字"9"，其非终止符为：  ；终止符为 $\overrightarrow{a}$，$\downarrow b$，$\overleftarrow{c}$，$\uparrow d$。改写规则为：

P:

$\square \to c^2d^2a^2b^2$

$\lrcorner \to b^2c^2$

显然，这两个文法（改写规则）都是非常简单的，且只能描述一个句子或一个数字"9"。如果想要描述一类句子或一批数字，则文法规则就复杂多了。对一般的图象，也可以用上述原则来构造文法，但是非常复杂。

有了各种文法后，对于未知结构信息的图象，在把它按一定的规则写成句子后，就可以用各种文法来进行句法分析。这往往是对一串终止符或链码通过由各种文法所对应的自动机来进行分析。如果能被某一种自动机通过，则这个图象就具有此自动机所对应的文法而产生的结构。这些就是用语言结构法来进行图象识别的大意。这将在第五章中作详细的介绍。由于这种方法只是在近十多年中才开始发展起来，因此，无论从理论上还是从实践上来看都还很不成熟。

在这十多年中，除了继续使用上述两种方法来识别图象外，还产生了用模糊集的方法来识别图象。由于客观世界中有很多概念不是确定性的，而是模糊的。例如年老、年轻、美丽、想象、…等。过去经典的集合论只反映确定性的概念：给定了一个集合，任何一个元素或者属于这个集合，或者不属于这个集合，二者必居其一。如果一个集合代表年老，则不属于这个集合的就代表年轻。显然，这样截然的分开方法也不是很科学的。