

[英] 厄尔曼 著

文字 图形 识别 技术

刘定一译 闫育苏校

WENZI TUXING SHIBIE JISHU 人民邮电出版社

文字、图形识别技术

[英] 厄尔曼 著

刘 定 一 译

闫 育 苏 校

人 民 邮 电 出 版 社

Pattern Recognition Techniques

J. R. Ullmann

1973

London Butterworths

内 容 提 要

本书主要介绍文字自动识别的各种基本技术和方法,并有一节介绍语言识别技术。内容包括:模板匹配,文字识别前的预处理,线性技术,分段线性技术、位势法和随机近似,多项式判别和N-数组法,布尔判定和序贯判定,特征,上下文的、语言的和阵列方法,系数分析,学习,共十章。可供从事文字识别、模式识别科技工作者阅读参考。

文字、图形识别技术

[英] 厄 尔 曼 著

刘定一 译 闫育苏 校

*

人民邮电出版社出版

北京东长安街27号

北京印刷一厂印刷

新华书店北京发行所发行

各地新华书店经售

*

开本: 787×1092 1/32 1983年7月第一版

印张: 14 8/32 页数: 228 1983年7月北京第一次印刷

字数: 323千字 印数: 1—7,000册

统一书号: 15045·总2692-有5283

定价: 1.75元

译 者 前 言

在半个世纪前，如果有谁说人能教机器识字，人们将斥之为无稽之谈。然而在今天，机器识字已变成了活生生的现实。

日益扩大的技术社会产生了浩如烟海的各种文件，用机器自动处理这些文件的要求越来越迫切。就以日常生活中来说，为数以吨计的邮件分选地址需要付出十分紧张繁重的劳动。现在我国已研制成功信函自动分拣机，在邮政自动化的道路上迈出了可喜的一步，“邮政编码”就是为适应信函自动分拣而产生的东西。

用机器进行**图形识别**(更确切地说，是**模式识别**)是人工智能的一个重要方面，是一类崭新的技术。它可能涉及人类活动的任何领域，可称是本世纪最激动人心的科学探险之一。但是，向国内读者介绍的这方面的书籍尚不多见。我们谨向广大读者推荐此书，作为填补空白尝试的一部分。

本书概括地介绍了文字识别(兼及语言识别)的各种基本技术和方法。文字识别是图形识别技术中研究得最深入的方面，然而远远谈不上彻底，还有大量的工作要做。我国的文字识别工作者还面临着—项最为艰巨的任务，即汉字识别。由于汉字数量繁多、形状复杂，远非阿拉伯数字与拉丁字母所能比拟，所以识别汉字最为困难。无论如何，本书对此可以有所启发。

本书基本上自成体系，适于自学。我们相信，不但是文字识别工作者，还有计算机科学家、仿生学家、人工智能研究者、控制论学者、心理学家、哲学家，他们从各种不同角度对本书

都会感到兴趣。

本书原名为“*Pattern Recognition Techniques*”，为了更好地反映其内容和重点，译名取为《文字、图形识别技术》。

本书作者厄尔曼(1936～)是位知名的图形识别专家，此书在他于英国国立物理研究所任首席研究员时写成，后来他任职于英国谢菲尔德大学应用数学与计算机科学系。

本书的翻译，得到邮电部第三研究所卓忠翰等同志、长沙工学院许织新同志以及其他不少同志的关心和大力支持。北京邮电学院的闫育苏老师在百忙中对译稿作了逐字逐句的精心审校，在此一併表示衷心的感谢。

译者才疏学浅，译文错误之处在所难免，敬请广大读者指正。

书中文字的右上角所注数码是参考文献的编号，可查阅书末作者所列的“参考文献”。

译者 1980 年 10 月

作 者 序

本书相当概括地、非正规地介绍了自动识别图形的技术。本书是为没有这方面专门知识的工程师、计算机科学家和生物学家写的。书中只是粗浅地涉及到统计学、开关理论、傅里叶理论、图论、语言学和光学之类的领域，而且只是用到这些学科的基本原理，因此这本书实际上是自成系统的。

在识别的图形很多、或者所要求的识别速度很高以致现有的人力难以胜任的实际情况下，图形识别的自动化可能很有用。例如识别染色体、血细胞、指纹、相貌、语言、心电图、脑电图、地震信号、泡沫室或火花室现象、导弹、印刷电路或玻璃制品中的缺陷、航空照相中的目标等。在图形自动识别的所有可能应用中，文字识别是最熟悉的和最成熟定型的。

本书有一节讲了语言识别，但是为了统一和明了起见，本书介绍的绝大多数基本技术都针对着文字识别问题加以说明，虽然有许多技术可以用于更广泛的方面。本书是引论性的概要，并不是包罗万象的论文，并且也不打算追溯观念的历史起源。关于图形自动识别的文献浩如烟海，列在书末的参考书只是其中一个小样本而已。

本书尽可能用文献中最广泛采用的术语，即使象识别 (recognition)、学习 (learning)、训练集 (training set) 之类的术语在哲学上可能要不得，而像窥窗模板 (peephole mask) 之类的术语又可能是古怪的。

识别系统的性能取决于要识别的图形的可变性和多样性。

所以，如要定量地评价系统的性能，就应该测量数据的可变性和多样性，并测量系统的误差率。除了有限的场合之外，我们没有可科学地复现可变性和多样性的量度可用，所以无法精确地比较建立在不同数据上的不同系统的识别性能。一个识别系统的误差率不仅依赖于数据和系统的工作原理，而且也依赖于实现这些原理的技巧和努力程度。就是说，如果用两台根据不同原理工作的机器来对同一种数据进行试验，比较二者的误差率并不一定能衡量出这两种原理的真正价值。由于这些困难，本书对识别系统的实际定量性能谈得很少。

但是，对于可接受的成本和当前的技术水平来说，究竟什么是做得到的，什么是做不到的，还是有某些一致意见的，特别在文字识别方面²⁹。大体上说来，在印刷质量得到控制的情况下，有可用的令人满意的技术来识别以给定铅字印成的文字。在印刷质量和铅字都完全未受控制时（例如识别任意一批打印的或铅印的信封字样），当代机器的误差率往往很高（例如按字算大于1%）；但是即使误差率比较高，自动识别在商业上可能仍然值得一试。有些在商业上用得颇广的机器，它们能识别按一定规则写成的手写文字。以可行的成本和误差率来自动识别不受约束的手写文字的完整集合0, ..., 9, A, ..., Z（如信封上的字）看来还没有成功。识别草体字（非实时地）更显然超出1972年的技术水平。

总的印象是，文字识别与其说是一个问题，不如说是一系列越来越困难的问题。在这串问题中最容易的一头，文字是受到约束或控制的，识别这种文字，是牢固建立的、一门投资几百万英镑的工业的基础^{17, 67, 361}，在最困难的一头，文字所受的约束很少，识别这种字，还要进行许多研究²²²。

刘(C. N. Liu)博士和谢尔顿(G. L. Shelton)博士向我提

供了图 1.1，阿米塔奇(J. D. Armitage)博士和洛曼(A. W. Lohman)教授向我提供了图 9.8，海德(S. R. Hyde)先生向我提供了图 9.12, 9.13 和 9.14，为此我谨向他们表示谢意。本书不是我在国立物理研究所的公务的一部分，而是利用晚上和周末写成的，为此我非常感谢我妻子的宽容。

目 录

译者前言	1
作者序	1
第一章 模板匹配	1
1.1 光学模板匹配	1
1.2 采用模拟灰度的电子模板匹配	7
1.3 数字灰度	10
1.4 取值最大化	16
1.5 窥窗模板	18
1.6 负权	20
第二章 文字识别前的预处理	23
2.1 把视觉图形变为电子图形	23
2.1.1 尼普柯夫圆盘	23
2.1.2 光导摄像管	25
2.1.3 飞点扫描装置	26
2.1.4 光电管阵	27
2.1.5 传真发送器	28
2.2 二值化	29
2.2.1 阈由对比度确定	29
2.2.2 阈由测量笔划宽度来确定	31
2.3 对准	33
2.3.1 用测量来竖直对准	33
2.3.2 用穷举试验来竖直对准	35
2.3.3 水平对准和文字分隔	38

2.3.4 尺寸和透视的规格化	41
2.4 平滑化、边缘检测和细化	42
2.4.1 平滑化	42
2.4.2 边缘检测	45
2.4.3 细化	47
第三章 线性技术	50
3.1 识别类	50
3.2 误差最小的贝叶斯分类器	51
3.3 统计独立性	53
3.4 高斯分布	56
3.5 与规格化平均模板互相关	61
3.6 线性判别函数	63
3.7 固定增量程序	67
3.7.1 两个类的情况	67
3.7.2 收敛性的证明	70
3.7.3 多类情况下的固定增量程序	72
3.7.4 滞区	73
3.7.5 边界矩阵	74
3.8 图形误差	75
3.9 对分方案	81
3.9.1 一/多对分	81
3.9.1.1 单类参数确定法	81
3.9.1.2 多类参数确定法	82
3.9.2 一/一对分	82
3.9.3 线性可分的机会	84
3.9.4 多/多对分	85
3.9.4.1 理想的 τ 比特图形组成法	90

3.9.4.2	统计选择	90
3.9.4.3	切割平面算法	91
3.9.4.4	公共子图形	92
3.10	卡伦—莱芙展开式	93
第四章	分段线性技术、位势法和随机近似	97
4.1	分段线性判别函数	97
4.2	直觉确定的子类	98
4.3	最近邻域法	99
4.4	菲尔沙因和费希勒的方法	100
4.5	分段线性固定增量程序	103
4.6	位势法	104
4.6.1	未知概率分布的近似	104
4.6.2	子类位函数	106
4.6.3	位函数的多项式近似	107
4.6.4	在位势法中的误差校正	109
4.7	在图形识别中的随机近似	112
第五章	多项式判别和 N-数组法	116
5.1	最小二乘方近似	116
5.1.1	规格化正交展开式	116
5.1.2	最小二乘方近似	121
5.1.3	随机近似的实现法	122
5.2	最大似然 n -数组法	123
5.3	布莱索—布朗宁法	127
5.4	多项式判别函数	132
5.4.1	系数由统计方式确定的多项式判别函数	132
5.4.2	系数由误差校正程序确定	133
5.5	根据一个信息准则作自动选择	135

5.6 移位窥窗模板系统	141
5.6.1 用于数字式图形	141
5.6.2 用于模拟式图形	143
第六章 布尔判定和序贯判定	145
6.1 布尔函数	145
6.2 使用布尔函数的识别系统	153
6.2.1 IBM 1418	153
6.2.2 小区域窥窗模板	156
6.2.3 IBM 1975	158
6.3 不完全规定的布尔函数	161
6.4 用数值函数来实现布尔函数	162
6.4.1 引言	162
6.4.2 阈函数	163
6.4.3 多项式判别式	166
6.4.4 多类情况下的数值判别式	167
6.5 非数值序贯识别	169
6.5.1 一棵简单的判定树	169
6.5.2 程序控制的序贯	171
6.5.3 数据控制的序贯	176
6.6 判定策略	180
6.6.1 序贯概率比试验	180
6.6.2 级联判定	181
6.6.3 规格化	183
第七章 特征	185
7.1 引言	185
7.2 区域特征	185
7.2.1 通过边缘检测来识别	185

7.2.2	利用笔划接头和笔划末端的位置的方案	189
7.2.2.1	实现手段	189
7.2.2.2	讨论	194
7.3	图表示法	197
7.3.1	引言	197
7.3.2	不用属性的表示法	199
7.3.3	用属性来表示图	201
7.3.4	图中的点对应于单个黑比特	205
7.3.5	黑比特和白比特	212
7.3.6	同构的一个必要条件	217
7.4	序贯检测特征	220
7.4.1	一个 IBM 曲线跟踪器系统	220
7.4.1.1	设计原理	221
7.4.1.2	实现特征检测器的硬件	223
7.4.1.3	讨论	226
7.4.2	线跟踪技术	229
7.4.3	序贯检测长条中的特征	232
7.5	特征的讨论	239
7.5.1	子图形类	239
7.5.2	否定系统	240
7.5.3	非否定系统	243
7.5.4	关系系统	245
7.5.5	畸变形状的识别	247
7.6	交截计数系统	251
第八章	上下文的、语言的和阵列的方法	257
8.1	上下文	257
8.1.1	使用完整词的上下文	257

8.1.2 使用 n 联字的上下文	260
8.2 景象分析	264
8.2.1 给定物体可能有多种分类法	264
8.2.2 形状	265
8.2.2.1 不根据形状	265
8.2.2.2 可解析描述的形状变化	265
8.2.3 景象分段问题	265
8.3 图画句法	268
8.3.1 原生语法	268
8.3.1.1 有限状态语法	268
8.3.1.2 上下文无关语法	271
8.3.1.3 上下文有关语法	273
8.3.1.4 变换	276
8.3.2 图形语法	277
8.3.2.1 链	277
8.3.2.2 二维语法	282
8.3.2.3 图语法	287
8.3.2.4 语法推断	289
8.4 用综合来分析	289
8.5 叠代阵列技术	292
8.5.1 微单元阵列	293
8.5.2 宏单元阵列	298
第九章 系数分析	303
9.1 高阶矩	303
9.2 缝隙扫描技术	306
9.2.1 加权过滤器缝隙扫描	306
9.2.2 用于特殊字体的缝隙扫描法	307

9.3 傅里叶变换	313
9.3.1 离散的傅里叶级数	313
9.3.2 傅里叶积分	318
9.3.3 傅里叶定理	319
9.4 用傅里叶光学来识别图形	321
9.4.1 用夫琅和费衍射实行傅里叶变换	321
9.4.2 识别字的傅里叶变换	323
9.4.3 在图形识别中的夫琅和费全息照相术	327
9.5 自相关	334
9.5.1 二项自相关	334
9.5.2 多重自相关	337
9.6 语言识别	339
9.6.1 引言	339
9.6.2 通过识别时-频图形来识别口语单词	343
9.6.3 波谱段的独立分类	344
9.6.4 交零	349
9.6.5 语言的自相关	351
9.6.6 语言识别中的分段问题	352
第十章 学习	359
10.1 不加管理的学习	359
10.1.1 二级系统	359
10.1.2 完全不加管理的学习	361
10.2 特征的自动确定	363
10.2.1 引言	363
10.2.2 否定系统	364
10.2.2.1 选择技术	364
10.2.2.2 构造 f -类的一种方法	365

10.2.3 非否定系统·····	368
10.3 一种关系系统·····	369
10.3.1 图形的畸变·····	369
10.3.2 单重分块·····	374
10.3.3 多重分块: 分离的训练集·····	376
10.3.4 多重分块: 混合训练集·····	382
10.3.5 训练集·····	385
10.3.6 退化·····	385
10.3.6.1 识别时的退化 ·····	385
10.3.6.2 在确定 M_0 时的退化·····	387
10.3.6.3 在确定 M_0 时的退化·····	388
10.3.7 讨论·····	391
10.3.7.1 n 的值 ·····	391
10.3.7.2 一个语法推断问题 ·····	391
10.3.7.3 通过穷举试验来规格化 ·····	391
10.4 学习的迁移·····	395
10.5 联想存储器·····	397
10.6 自动图形识别的科学基础·····	402
10.6.1 设计无自动学习能力的识别系统·····	402
10.6.2 把分类表存储在联想存储器中·····	404
10.6.3 在图形识别中的学习·····	405
参考文献·····	409

第一章 模板匹配

1.1 光学模板匹配

1929年陶舍克(Tauschek)⁴³⁰描述了这样一台机器，它能识别用单副铅字印成的数字。陶舍克的机器用了10块模板，分别对应于10个数字。模板就是象空心花纹板和照相底片那样的东西。例如，如果在“2”字的模板上投射“2”的正象，那么被投影的“2”的黑色部分就复盖住模板的所有透明部分，只有很少的光线能通过模板。被投影的“2”的黑色部分却复盖不了其他模板的所有透明部分。为了识别文件上的一个数字，陶舍克的机器就把这个数字的象依次投射到10块模板上。透过每一块模板的总光量用光电管测量后，这个数字就被识别为其模板透过的总光量比规定的最小光照量还小的那个数字。换句话说，把一个数字识别为与它匹配的那块模板的数字。

汉德尔(Handel)¹⁹⁸对这种类型的另外一台机器作了描述，在这台机器中每个数字都有一个输出通道。例如，如果文件上的一个数字与模板“2”匹配，那么在“2”输出通道上就加上一个电压，而在所有别的输出通道上则加上另一个不同的电压。如果一个数字与模板“3”匹配，那么在“3”输出通道上就加上一个电压，而在所有别的输出通道上则加上另一个不同的电压，等等。如果一台机器产生符合一个字的输出电压，这台机器就识对了这个字。举例来说，如果机器对数字“2”产生了符合“3”的输出电压，就说机器把“2”误识了。