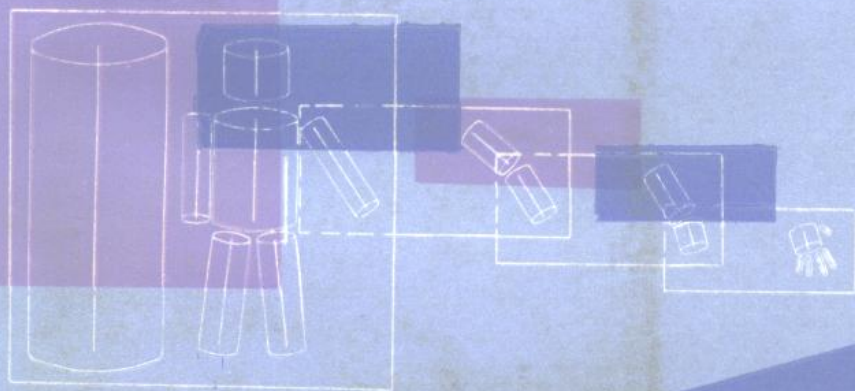


视觉计算理论

〔美〕D. 马尔 著



科学出版社

视觉计算理论

(美) D. 马尔 著

姚国正 刘 磊 汪云九 译

科学出版社

1988

内 容 简 介

本书以计算机科学为核心,以视觉神经生理学和心理物理学近年来取得的重要成果为基础,全面论述了视觉计算理论。全书共七章,分别论述:视觉研究的基本原理和方法;图像的表象:从图像恢复表面,可见表面的直接表象;用于识别的形状表象;视觉计算方法提要;为视觉计算方法辩护,还有术语解释和索引,为帮助读者掌握视觉计算理论的基本概念,了解它的意义并对视觉计算理论的最新进展有所了解,书末收入了其他作者的三篇论文及麻省理工学院脑和认知科学系教学大纲作为附录。

本书可供心理学、生理学、生物物理学、计算机视觉、人工智能等方面的研究人员和高等院校有关专业的教师、研究生等参考。

David Marr

VISION

*A Computational Investigation into the
Human Representation and Processing of
Visual Information*

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

1988年5月第一版 开本: 850×1168 1/32
1988年5月第一次印刷 印张: 14 3/4
印数: 0001—3,120 字数: 381,000

ISBN 7-03-000023-4/Q·5

定价: 6.00元

译者的话

视觉是一个古老的研究课题，但视觉计算理论却是在本世纪70年代中期才开始出现，并在80年代初就大放异彩的一个崭新的先进理论。随着电子计算机技术的发展，从60年代起图像识别在国外蓬勃地发展起来，但到70年代却普遍地发生了“危机”。人们发现：在一幅图像中任何事情都可能发生，而且所有的事情也的确在图像中发生了。用计算机实现图像识别这一人类轻而易举就能完成的任务所遇到的这种困难，几乎把研究人员逼入了绝望的境地。本书作者所创立的理论为摆脱这一困境开辟了一条道路，而计算机视觉这几年的发展形势的确也越来越好。现在，国内从事图像处理和模式识别工作的大学和研究机构很多，视觉神经生理学、视觉心理学、人工视觉(计算机视觉、机器人视觉或视觉代偿)等方面的工作也正在逐步开展，但它们需要得到视觉理论上的指导，以便把工作引向深入。我们翻译本书的目的，也就是为了推动国内有关工作的进一步发展。

视觉计算理论的创始人D. Marr教授原籍英国。他于1973年到美国，应邀在麻省理工学院人工智能实验室创建并领导一个以博士研究生为主体的研究小组，主要从事视觉理论方面的探索。Marr在视觉方面最初的开发性工作是用计算机实现人的立体视觉功能。本书所论述的计算理论的整体框架是他在1977年的一个报告中提出来的。他接着对计算理论做了许多充实和提高的工作。在他生命快结束的时候，Marr才动手写我们现在读到的这本书(最后是由他的一些学生完成的)。本书把他提出的视觉整体框架和当时已经取得的重大研究成果自然地结合了起来，深刻地阐明了视知觉研究的一个非常有用的新方法。它不仅为视觉的理论研究提供了有力的思想武器，而且还为视觉的应用研究开创了广阔的实用

前景。除作者富有成果的创造性思想外,从本书中,我们还可以看到他的广博的知识和出众的才华。Marr 因白血病于 1980 年 11 月在波士顿病逝,当时仅 35 岁。

Marr 的视觉计算理论立足于计算机科学,系统地概括了心理物理学、神经生理学、临床神经病学等方面业已取得的所有重要成果,独树一帜,从而为一门新的视觉信息处理学科奠定了基础。这个理论具有鲜明的时代特征。它对神经科学的发展可能是极重要的,对人工智能已经产生深远的影响。事实上,这个理论一出现就深受神经科学家和人工智能专家们的推崇。在美国,1983 年还专门召开了 Marr 学术思想讨论会。由于本书是迄今见到的最完善的一部视觉理论著作,凡从事自然视觉和人工视觉工作的研究人员、教师、研究生都可以从中得到教益(有的甚至把它奉为至宝),因此本书自然就成了了一本抢手的畅销书。在阅读本书的时候,我们希望读者能注意计算理论的下列要点:

1. 首先,该理论认为,视觉是一个复杂的信息处理任务,其主要目的在于认知外部世界中有什么东西,它们在什么地方。

2. 要完整地理解视觉,就必须在三个不同的层次上对它进行解释。这三个层次是:计算理论(注意,这是 Marr 首创的)、算法和硬件实现。

3. 从计算理论这个层次来看,视觉信息处理必须用三级内部表象来加以描述。这三级表象是:要素图(图像的表象)、2.5 维图(可见表面的表象)和三维模型表象(用于识别的三维物体形状表象)。

4. 视觉信息处理是按照功能模块的形式组织起来的。换言之,在图像中被编码的表面信息是由一些几乎相互独立的处理过程分别实现解码任务的。

5. 最后,在研究各个处理过程、以便从图像中恢复可见表面的几何结构时,用公式表示计算理论的一个关键性的步骤,就是要根据外部客观世界的一般特性,找出有关问题的约束条件,并把它们变成精密的假设,从而得出确凿的、经得起考验的结论。这种结

果将具有永恒的性质,就像物理学中的结果一样,因为它们是在外部世界的物理学和图像的公式化的基础上得出来的。视觉计算理论的重要特征,就在于它使视觉信息处理的研究变得十分严密,把视觉研究从描述的水平提高到数理科学的研究水准。

本书分三个部分,共七章。第一部分(总导言、第一章)论述视觉研究的基本原理和方法,这部分内容很重要。第三部分(第七章)是作者特意安排的一组问答,目的是要帮助读者掌握这一理论在思想方法上的特点。这部分内容针砭了人工智能及认知科学中各个流派的理论,是值得玩味的。第二部分(第二章至第六章)则对视觉问题作了具体的分析。作者在这里论述了他所倡导的研究方法和理论框架在视觉中是如何得到体现的,同时也向读者展示了业已取得的成果。当然,视知觉问题在这里并没有完全被解决,但这部分内容的确表明视觉表象这个方法是一个十分有用的方法。

在此我们应当指出,自本书于1982年面世以来,视觉计算理论的研究又取得了一大批值得称道的新成果。例如,零交叉理论和人的体视算法已经大大向前发展;许多初级处理模块有了一个统一的理论形式;对2.5维图的结构提出了新的见解;从视觉算法出发,为所谓“信息生物物理学”规定了基本的内容,等等。总之,Marr书中的空白正在迅速地得到填补。这是值得引起读者注意的。

为了有助于读者掌握和了解视觉计算理论的基本概念和它的意义,为了使读者对视觉计算理论的最新进展有所了解,我们选择了Marr的亲密合作者之一T. Poggio教授的二篇论文及国际脑研究学会主席伊藤正男教授的一个评论及麻省理工学院脑和认知科学系教学大纲,作为附录收进本书。

读者也许会问:为什么创立视觉计算理论的是Marr而不是别人呢?对此,我们的回答是:恰恰是Marr具备了创立理论的各种条件。我们一开始就提到,以图像识别为主要内容的动物和机器视觉的研究经过一阵蓬勃的发展之后,从七十年代初开始就出现了严重的危机。这一形势要求研究工作者对以往走过的道路进行深刻的反省,同时,这也为创立新的理论提供了机会。而Marr

正好又具有创立新理论的素质和外部条件。Marr 早年在剑桥大学 Trinit 学院攻读数学（获得数学硕士学位），而他的兴趣则在神经生理学方面（获得神经生理学博士学位），同时他还受过神经解剖学、生物化学、心理学等方面的严格训练。在英国，他曾从事过新皮层、海马、特别是小脑方面的理论研究。到美国后，他既搞视觉的理论研究，又任麻省理工学院心理系的教授。这个经历表明，在 Marr 这位学者身上集中体现了多种学科互相交叉、互相渗透这种现代科学发展的趋势，而这正是创造性思想的极重要的来源。非但如此，Marr 还在计算机上做了整整十年工作，具有极丰富的编程经验和感性知识。正是他设计和实现计算机程序所积累的经验，才使他得以总结出信息处理系统的“四大组织原则”这样的真知灼见，提出这些原则对神经系统的信息处理也是适用的这一远见卓识。除此之外，Marr 的工作条件是很理想的。如他在序言中写的那样，上峰允许他便宜行事，周围有一批杰出的人才，政府和社会团体给予了慷慨的财政资助；这些无疑也是他取得成功的一个重要条件。我们中国人爱看《三国演义》，“占天时、据地理、得人和”已成了一句老生常谈。译者顺便借用一句，即 Marr 之所以能创立视觉计算理论，是因为他尽占天时、地理、人和的缘故。但是，最基本的，Marr 首先是一个杰出人才。因此，任何一个有志于有所发明、有所创造的科学工作者，特别是年青的科学工作者，都应当不断地完善自己。

本书是我们几个人断断续续翻译出来的，前后差不多延续了两年。不少关心本书出版的读者曾再三来函催稿，我们实在愧对这些热心的读者。译书的分工是：姚国正译总导言，第一章的 1.2—1.3 节，第二章，第三章的 3.1—3.2、3.4—3.6、3.9—3.10 节；刘磊译第一章的 1.1 节，第三章的 3.3、3.7—3.8 节，第四章；汪云九译第五章至第七章。全部译文最后由姚国正统一整理。本书几乎全用直译，力求忠于原作的内容和风格。由于这部著作内容博大，“新词”频仍，语言明快，文字流畅，而译者这方面学问既小，闲空的时间又少，因此不大可能按照“信、达、雅”那样一个高标准来做。退

而求其次的一个做法,就是力求译文的准确和通顺。在统一译文风格、进行汉语规范处理的时候,对本书第一和第三部分(以及其他有关方法论的部分),我们使用汉语习惯用法较多,努力保持原作的文采和风度;对第二部分,我们使用平铺直叙的译法较多,努力保持原作平顺通畅的叙述方式。这是因为我们希望本书能成为一本颇可一读的、内容正确的视觉计算理论入门书;而这也正是原书作者所追求的。

在本书翻译过程中,程子习、丁晓鸿帮助翻译了附录Ⅱ,沈德兴、王克非、邢静帮助抄写了部分译稿,译者向他们一并致以谢忱。限于译者水平,书中错误不当之处是难免的,我们恳切地希望读者提出宝贵的批评和建议。

译 者

识于中国科学院生物物理所

1986年11月17日

序 言

本书不能算是一本论述严谨的教科书，它只记述了我近几年的探索和追求。这一探索性工作开始于1973年；是年，我应 Marvin Minsky 教授和 Seymour Papert 教授之请，到麻省理工学院人工智能实验室工作。由于 Patrick Winston 出色的行政管理工作，由于国防部高级研究计划局和国家科学基金会提供的慷慨资助，由于 Whitman Richards 在 Richard Held 的关照下允许我便宜行事，因此工作条件是很理想的。我有幸能结识一批杰出的人才——特别是 Tomaso Poggio——并和他们一起工作。在这批人中，有许多原先是研究生，他们后来才成为我的同事，从他们那里我学到了许多东西。他们是：Keith Nishihara, Shimon Ullman, Ken Forbus, Kent Stevens, Eric Grimson, Ellen Hildreth, Michael Riley 和 John Batali。Berthold Horn 使我们能跟上不断发展的光物理学，Whitman Richards 则不断使我们了解在研究人之所能及不能方面所取得的最新成果。

1977年12月发生的一些不幸事件使我不得不比原计划提前几年就动手写这本书。虽然书中有一些重大的空白——我希望这些空白不久就会得到填补——但关于视觉研究的一个崭新的理论框架已经崭露头角，并得到足够坚实的研究结果的支持，因此就值得把这两者结合起来作为一个统一的整体记述下来。

许多人帮助我渡过了这段相当困难的时期，特别是我的双亲、我的姐妹、我的妻子 Lucia，以及 Jennifer、Tomaso、Shimon、Whitman 和 Inge。他们为我做的一切远远超过我应得的。尽管仅向他们表示谢忱并不足以表达我的感激之情，但我还是由衷地感谢他们。William Prince 把我介绍给剑桥 Addenbrooke 医院的 F. G. Hayhoe 教授和 John Rees 博士，我感谢他们给了我宝贵的时间。

戴维·马尔

1979 年夏

目 录

译者的话 序 言

第一部分 导言和视觉研究的准备

总 导 言	1
第一章 视觉研究的基本原理和方法	6
1.1 视觉研究的背景	6
1.2 理解复杂的信息处理系统	18
表象和描述	19
处 理	21
三个层次	24
计算理论的重要性	27
J.J. Gibson 的方法	29
1.3 关于视觉的表象框架	31
视觉的目的	32
高级的视觉	35
必由之路	37

第二部分 视 觉

第二章 图像的表象	40
2.1 初级视觉的物理基础	40
图像的表象	43
基本物理假设	43
表面的存在 (43) ; 分层次的组织 (44) ; 相似性 (46) ; 空间的 连续性 (48) ; 不连续中的连续性 (49) ; 运动的连续性 (49)	
初级表象的一般性质	50
2.2 零交叉和原始要素图	53
零交叉	53
生物学意义	60

初级视觉的心理物理学(60); ∇^2G 滤波器的生理学实现(63);	
零交叉的生理学检测(65); 图像的第一个真正的符号表象(66)	
原始的要素图.....	68
方法论上的补白.....	74
2.3 图像的空间配置关系	79
2.4 光源与透明度	87
其他的光源效应.....	89
透明度	90
结 论.....	91
2.5 组合处理和完全的要素图	92
论证的要点.....	96
质地辨别的计算方法和心理物理学.....	97
第三章 从图像恢复表面	100
3.1 人的视觉信息处理机器的模块组织	100
3.2 处理、约束及图像合适表象	104
3.3 立体视觉	113
立体视差的测量	114
计算理论(114); 体视匹配的算法(121); 体视融合的神经	
系统实现(157)	
从视差计算表面距离和朝向	131
计算理论(161); 算法与实现(164)	
3.4 方向选择性	165
视觉运动引言	165
计算理论(172); 一个算法(174); 神经系统的实现(176)	
用方向选择性分离独立运动的表面	183
计算理论(183); 算法和实现情况(185)	
深度方向上的运动	189
3.5 表观运动.....	190
为什么要研究表观运动?	191
把问题一分为二	192

对应问题	197
实验的发现(197); Ullman的对应处理理论 (205); 对Ullman 理 论的评论 (207); 关于对应问题的新看法 (211)	
从运动恢复结构	215
问题 (215); 以往解决问题的方法 (217); 刚性约束 (219); 刚性假 设 (221); 关于透视投影的一个注解 (222)	
视觉流	222
输入表象 (223); 数学上的结果 (224)	
3.6 形状轮廓线	226
几个实例	227
由遮挡造成的轮廓线	229
几个约束性假设 (231); 这些假设的意义 (234)	
表面朝向的不连续性	236
表面上的轮廓线	238
表面轮廓线的疑难 (239); 确定轮廓线发生器的形状 (240); 多 于一条轮廓线的作用 (241)	
3.7 表面质地	245
质地元素的分离	245
表面的参数 (246); 可能作哪些测量; 直接估计按比例变化的距 离 (249)	
小 结	250
3.8 明暗和光度学体视	251
梯度空间	252
表面照明、表面反射率和图像强度	255
反射率地图	257
由明暗恢复形状	260
光度学体视	261
3.9 亮度、光度及颜色	263
Helson-Judd的方法	265
光度和颜色的锐化处理理论	267
算法(268); 推广到颜色视觉 (269); 对锐化理论的评论 (270)	

说明同时对比具有重要意义的一些物理学依据	272
关于强度的非线性变化的表面根源假设	275
测量三色图像的意义	275
本方法小结	278
3.10 本章总结	278
第四章 可见表面的直接表象	282
4.1 引言	282
4.2 图像的分割	284
4.3 从另一个方面表述问题	287
4.4 被表达的信息	289
4.5 2.5维图的一般形式	292
4.6 可能被表象的信息	294
4.7 可能使用的坐标系	298
4.8 内插、延拓和不连续性	301
4.9 内插的计算问题	304
不连续性	305
内插方法	307
4.10 其他内部计算	308
第五章 用于识别的形状表象	312
5.1 引言	312
5.2 由形状表象提出的若干问题	313
判定形状表象有效性的几个准则	313
可达性(313); 范围和唯一性(314); 稳定性和敏感性(314)	
形状表象的设计原则	315
坐标系(315); 基元(317); 组织结构(319)	
5.3 三维模型表象	319
自然坐标系	320
基于轴线的描述	321
三维模型表象的模块组织	322
三维模型的坐标系统	324

5.4	自然的推广	326
5.5	推导和使用三维模型表象	330
	三维模型描述的推导	330
	以观察者为中心的坐标系与以物体为中心的坐标系之间的关系	334
	检索和三维模型的类目	335
	推导和识别之间的相互作用	338
	找出图像和类目模型之间的对应关系(339); 约束分析(340)	
5.6	心理学方面的考虑	342
第六章	提要	346

第三部分 后 记

第七章	为视觉计算方法辩护	350
7.1	引言	350
7.2	对话	350
	术语解释	385
	参考文献	393
附录 I	Marr 的视觉计算理论	409
附录 II	计算视觉和正则化理论	416
附录 III	神经生理学家该向何处去?	435
附录 IV	麻省理工学院脑和认知科学系教学大纲	439
	索引	441

第一部分 导言和视觉研究的准备

总 导 言

看见指的是什么意思?普通人(以及亚里斯多德)对这个问题的回答是:通过观看,认知有什么东西在什么地方。换言之,视觉乃是一种处理过程(process),是要从图像中发现外部世界中有什么东西,它们在什么地方。

因此,视觉首先是一个信息处理任务。但是我们认为视觉和处理并不完全相同。因为如果我们能够认知外部世界中有什么东西在什么地方,那么我们的脑就必须以某种方式表象(representing)这种信息——在绚丽的色彩、多种多样的形状、美感、运动、细节等所有各个方面表象这种信息。所以视觉研究不仅必须探讨我们是如何从图像中提取对我们有用的外部世界的各个方面的,而且还必须探讨内部的表象性质,因为正是借助于这种内部表象,我们才能把握这种信息,并使之成为我们思想和行为决策的基础。表象和信息处理这一双重性质,乃是大多数信息处理任务的核心所在,并深刻地影响着我们对视觉提出的特殊问题所进行的研究。

对理解信息处理任务和信息处理机器的需要是最近才产生的,在人们起而设想、继而着手建造这样一些机器之前,根本就没有深入考虑这些问题的迫切需要。但是,人们一旦真正开始认真思索这些任务和机器,就很容易发现我们周围世界的许多方面都受益乎信息处理的观点。许多与我们人类休戚相关的现象——例如,生命和进化的奥秘,知觉、感情和思想的奥秘——本质上都是信息处理现象。因此,如果我们当真要充分理解这些现象,那么在我们对于这些现象的思考中就必须把这个观点也包括进去。

下面要着重指出的一点是：说一件工作“仅仅”是一个信息处理任务，或者说一个有机体“仅仅”是一个信息处理装置，这对所论事物既无限制，亦无贬低之意。更重要的是，我决没有用这种说法把那些必要的解释限制在某一种形式上的意思。事实上，我们的做法正好相反。信息处理机器的引人入胜的特点之一就在于：为了完整地理解它们，我们必须欢迎在许多不同的层次上对信息处理机器进行解释。

例如，让我们来看一看一个人，作为人类的普通一员或作为一个科学家，当他声称对视觉已经有所了解时，他的哪些观念必须被满足。首先，我认为也是最重要的，是作为一个普通的人的观念。他懂得所谓“看见”到底是怎么一回事，因此除非我们提出的论据和理论的要点与他凭直接经验确认是正确的东西大体上相符合，否则我们就可能错了（Austin于1962年曾以精巧而有力的方法指出过这一点）。其次，脑科学家、生理学家和解剖学家则有他们自己的见地。关于神经系统是如何构造的，其中一些部分的行为特点如何，他们知之甚多。对于这些科学家来说，令人满意的解释必须能联系知觉来解决并阐明他们所关心的问题：细胞是怎样联结起来的？它们的反应为什么是现在这个样子？Barlow(1972)提出的神经元数据到底怎么样？同样的论据也适用于实验心理学家的看法。

另一方面，买了一台小型家用计算机并摆弄过它的人可能会提出完全不同的要求。他可能会说，“如果视觉确实是一个信息处理任务，那么只要我的计算机具有足够强的功能，具有足够大的存贮器，而且有机会把它和家用电视摄像机联起来，我就可以用它来完成视觉任务”。因此，他要求得到的解释是相当抽象的，应当能够告诉他对什么东西编制程序，如果有可能的话，还应当能够提示一下对编制程序有用的最佳算法。他并不想了解视紫红质、外侧膝状体核、有抑制性作用的中间神经元等等。他要求知道的是怎样编制能计算视觉的程序。

至关重要的一个论点就是：为了解执行信息处理任务的装

置,我们需要多种不同的解释。本书第一部分所论述的就是这个观点,这一部分非常重要。虽然在最近的科学进展如分子生物学中,对如何构成现象的解释已有必要进行仔细的斟酌,但对于视觉问题就必须更加谨慎小心。意识到这一点,已经成为本书的一个基本观点。对于视觉这个学科来说,根本就没有哪一个公式或哪一种观点是能够包罗万象洞察一切的。视觉中的每一个问题都必须从几种不同的观点来加以论述——它可以被看成是一个信息的表象问题,是一个能推导出这种表象的计算问题,或者是一个能又快又可靠地执行上述两项任务的计算机的结构体系问题,等等。

如果我们牢牢记住解释的这个本质,记住这个必然是相当普遍的提法,那么我们就避免许多失误。例如,强调信息处理的一个结果可能会导致在人脑和计算机之间进行比较。当然,从某种意义上来说,脑就是计算机;但如果不加任何限制,这个说法就会把人引入歧途。因为人脑实质上不仅仅是一台计算机,而且是一台习惯于执行某些相当特殊的计算的计算机。计算机这个术语通常是指一种机器,这种机器具有相当标准的指令集,它在存入存贮器中的程序的控制下,进行通常是串行的运算(但当今的机器有时也可以进行并行的运算)。为了理解这样一台计算机,就必须了解与计算机有关的一系列问题:这台机器是用什么部件构成的?它是怎样装配起来的?机器的指令集是什么样的?它的存贮器有多大?怎样存取数据?如何把机器投入运行?等等。但对于理解一台正用来执行某一信息处理任务的计算机来说,上面提到的不过是应当理解的一小部分内容罢了。

这个看法是有代表性的。要知道为什么在人脑和计算机之间所作的大多数类比都失于肤浅而毫无价值,这便是原因所在。例如,让我们考虑一下由计算机飞机订票系统组成的国际网络。这个计算机网络的任务是为全世界千百万乘客指定航机。为了理解这个系统,只懂得现代的计算机是如何工作的还不够。我们还必须懂一点什么是飞机,它们是干什么用的;懂一点地理学、时区、票价、汇率、联运问题;懂一点政治、饮食,以及恰好与这一特殊任务有关