

认知基础

Cognitive Foundation

第四章

视觉和注意

Vision and Attention

史忠植

中国科学院计算技术研究所
<http://www.intsci.ac.cn/>

内容提要

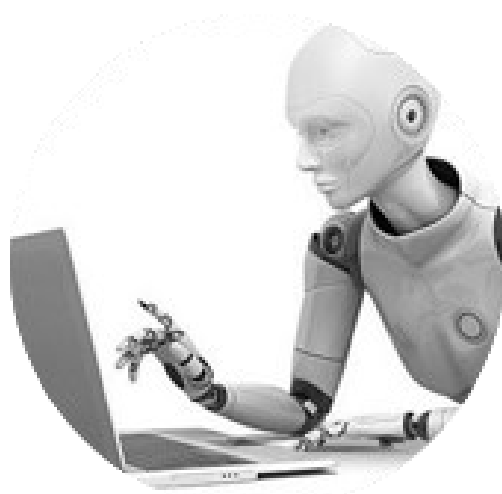
- 感知过程
 - 视觉的生理机制
 - 视皮层信息处理
 - 格式塔视觉理论
 - 拓扑视觉理论
 - 马尔的视觉计算理论
 - 计算机视觉
 - 注意

智能



感知智能

能听会说、能看会认



认知智能

能推理会理解



决策智能

能思考会决策

感知智能

感知智能是指将物理世界的信号通过摄像头、麦克风或者其他传感器的硬件设备，借助语音识别、图像识别等前沿技术，映射到数字世界，再将这些数字信息进一步提升至可认知的层次，比如记忆、理解、规划、决策等等。而在这个过程中，人机界面的交互至关重要。

感知



感知是客观外界直接作用于人的感觉器官而产生的，
在发展中经历三种基本形式：

- 感觉
- 知觉
- 表象

感知过程

- 感觉是客观事物的个别属性、特性在人脑中的反映。
- 知觉是各种感觉的综合，是客观事物整体在人脑中的反映，它比感觉全面和复杂。在知觉的基础上，产生表象。
- 表象即印象，是通过回忆、联想使这些印象再现出来。它与感觉、知觉不同，是在过去对同一事物或同类事物多次感知的基础上形成的，具有一定的间接性和概括性。但表象只是概括感性材料的最简单的形式，它还不能揭露事物的本质和规律，

感 觉

- 感觉的涵义

感觉是人脑对客观事物的各个属性、某种具体特征的直接反映。是一个人接受外界信息的过程，也是一个人觉察和获取刺激信息的重要渠道，是人们关于世界一切知识的重要源泉

- 感觉信息加工

它是指感觉器官与刺激信息相互接触（或作用）之后，感受器将物理刺激转化为生物电信号，并通过动作电位把外部事件的信息传递到大脑中枢特定区域的过程

- 感觉信息 (Sensory Information)：外部信息在人脑内瞬间滞留，是尚未经过诠释与归类的信息。

知 觉

- 知觉是人脑对客观事物的各种属性、各个部分及其相互关系的综合的、整体的反映
- 认知心理学认为，知觉是人脑对感觉信息的组织和解释，即获得并说明感觉信息的意义的过程
- 知觉信息加工过程
 - ☞ 包含互相联系的几个方面：觉察、辨别、确认
 - ☞ 直接知觉和间接知觉

表象

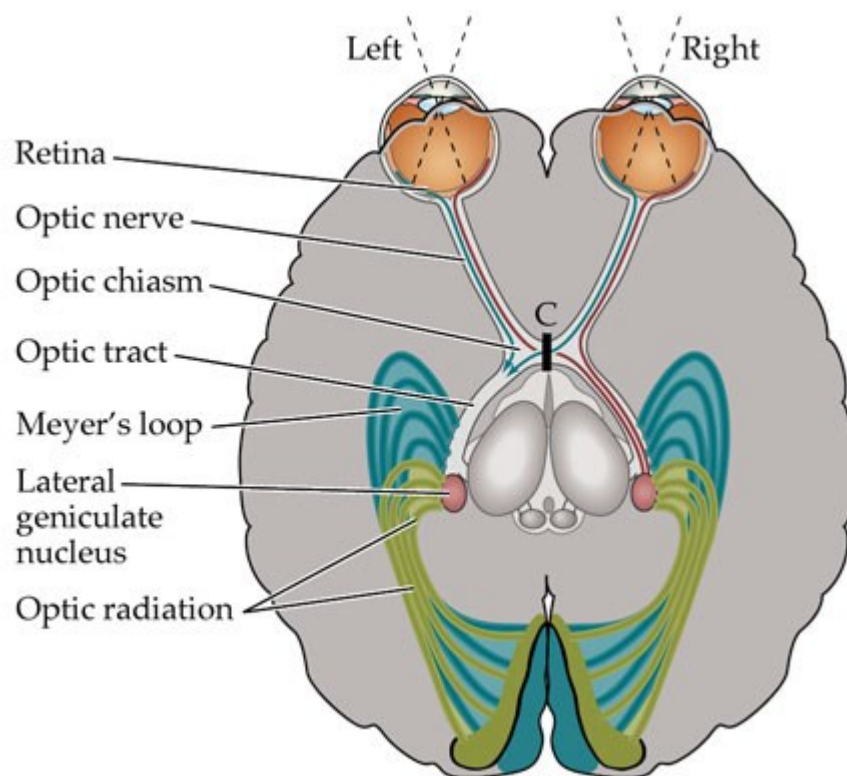
- 表象 (representation) 是客观对象不在主体面前呈现时，在观念中所保持的客观对象的形象和客体形象在观念中复现的过程。人们在头脑中出现的关于事物的形象。从信息加工的角度来讲，表象是指当前不存在的物体或事件的一种知识表征，这种表征具有鲜明的形象性。

内容提要

- 感知过程
- 视觉的生理机制
- 视皮层信息处理
- 格式塔视觉理论
- 拓扑视觉理论
- 马尔的视觉计算理论
- 计算机视觉
- 注意

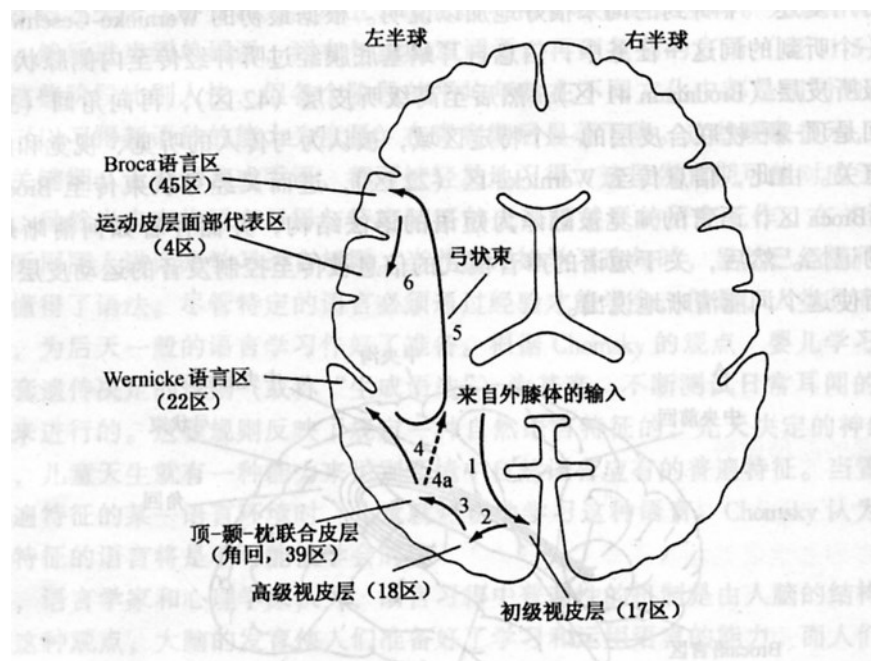
视觉的生理机制

- 眼是视觉的外周器官，是人体所有感官中最复杂的。来自外界的信息很大部分是通过眼接收的。外界物体的光线射入眼中，聚焦后在视网膜上成象。从视网膜发出神经冲动，经视神经达到大脑皮层视区，产生视觉。

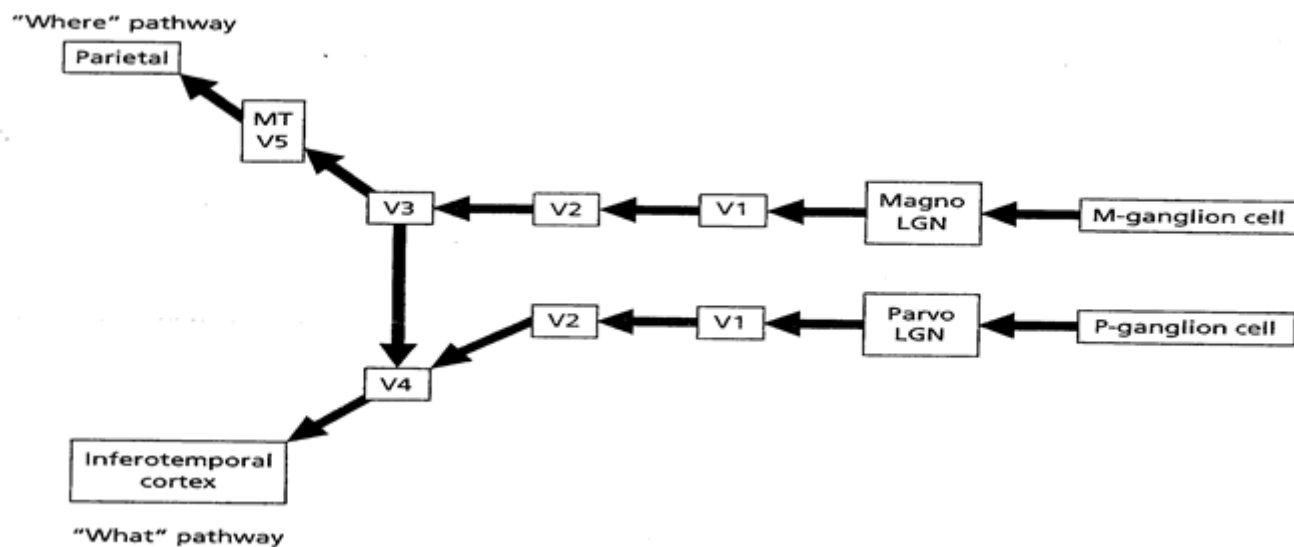


视觉的神经通路

- 视皮层中17区被称为第一视区 (V1) 或纹状皮层
- 和视觉有关的皮层还有纹前区 (18区) 和纹外区 (19区)。
- 18区内包括三个视区, 分别称为V2, V3和V3A
- 19区深埋在上颞沟后壁, 包括第四 (V4) 和第五视区 (V5)



视觉的神经通路



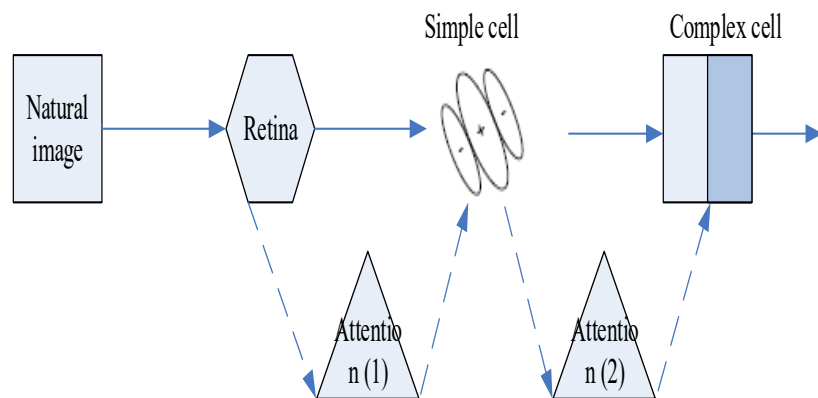
- 视觉主要有两个功能：一是目标知觉，即它是什么？二是空间知觉，即它在哪里？

内容提要

- 感知过程
- 视觉的生理机制
- 视皮层信息处理
- 格式塔视觉理论
- 拓扑视觉理论
- 马尔的视觉计算理论
- 计算机视觉
- 注意

视皮层信息处理

- 根据皮层神经元感受野结构的不同，休贝尔(Hubel)和维塞勒(Wiesel)对猫和猴的视皮质中单一神经元的激发模式进行的研究，发现有四种类型视皮层神经元——简单细胞，复杂细胞，超复杂细胞和极高度复杂细胞。



特征选择性

视皮层神经元对视觉刺激的各种静态和动态特征都具有高度选择性，包括

- 方位/方向选择性
- 空间频率选择性
- 速度选择性
- 双眼视差选择性
- 颜色选择性

功能柱

- 视皮层功能柱的发现是近年来视觉中枢研究中最引人注目瞩目的进展之一。由于这些柱系统正好与各种特征检测功能一一对应，所以也被认为是对特征检测理论的一个有力支持。到目前为止所发现的各种功能柱都垂直于皮层表面，排列成片层状。
- 在猴和猫的17区和18区存在着方位柱结构。Hubel和Wiesel用微电极倾斜地穿刺视皮层。方位柱是与皮层表面垂直排列的。

内容提要

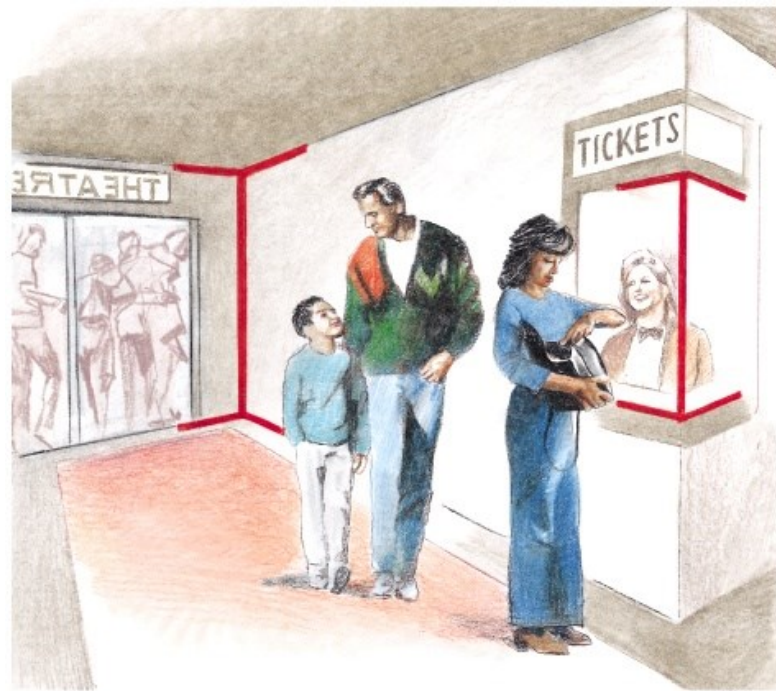
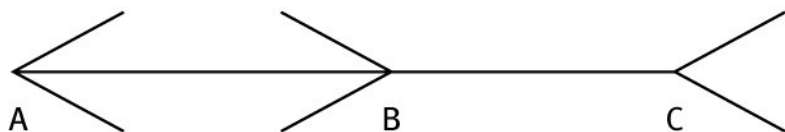
- 感知过程
- 视觉的生理机制
- 视皮层信息处理
- 格式塔视觉理论
- 拓扑视觉理论
- 马尔的视觉计算理论
- 计算机视觉
- 注意

格式塔理论

格式塔（Gestalt）心理学诞生于1912年。它强调经验和行为的整体性，反对当时流行的构造主义元素学说和行为主义“刺激-反应”公式，认为整体不等于部分之和，意识不等于感觉元素的集合，行为不等于反射弧的循环。尽管格式塔原理不只是一种知觉的学说，但它却导源于对知觉的研究，而且一些重要的格式塔原理，大多是由知觉研究所提供的。

格式塔心理学

- 知觉的形成是基于先验知识的刺激组织



格式塔心理学

- 格式塔理论认为，知觉问题涉及到比较和判断。

Picasso's
panting

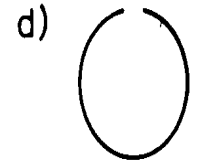
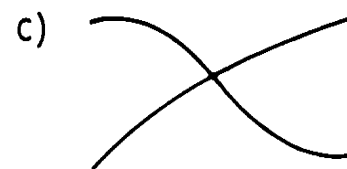
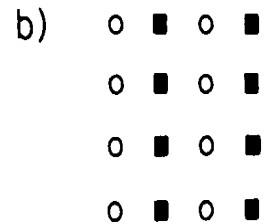
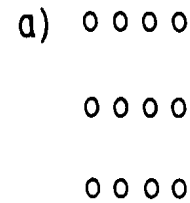


知觉组织法则

格式塔心理学家把人类知觉组织归结为几个重要的法则：

1) 邻近性。

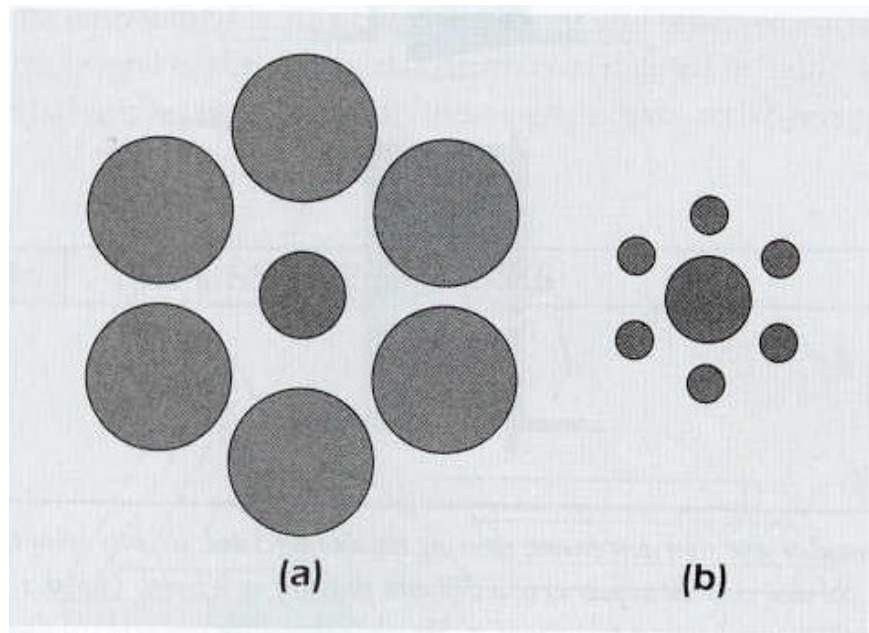
彼此紧密接近的刺激物比相隔较远的刺激物有较大的组合倾向。接近可能是空间的，也可能是时间的。例如夏天下雨时，雷电交加，我们就把它们知觉为一个整体，即知觉为同一事件的组成部分。



知觉组织法则

2) 相似性

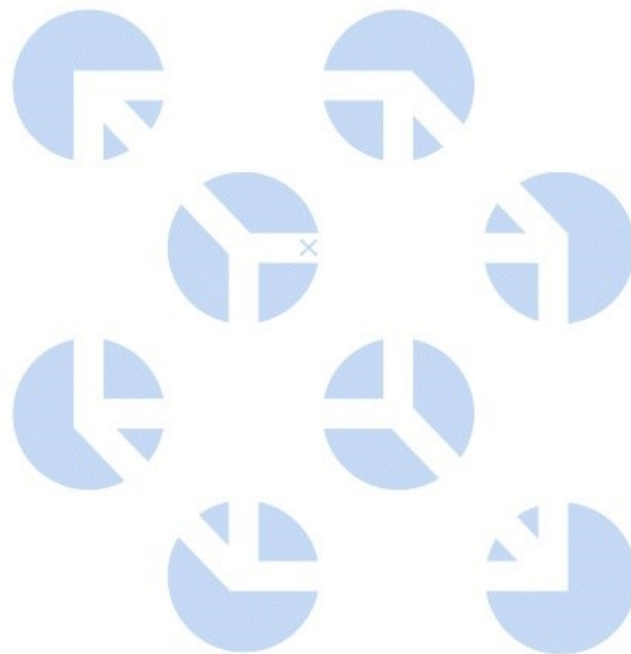
彼此相似的刺激物比不相似的刺激物有较大的组合倾向。相似意味着强度、颜色、大小、形状等这样一些物理属性上的类似。俗话说：“物以类聚，人以群分”，也就包含这种原则。”



知觉组织法则

3) 好的连续性

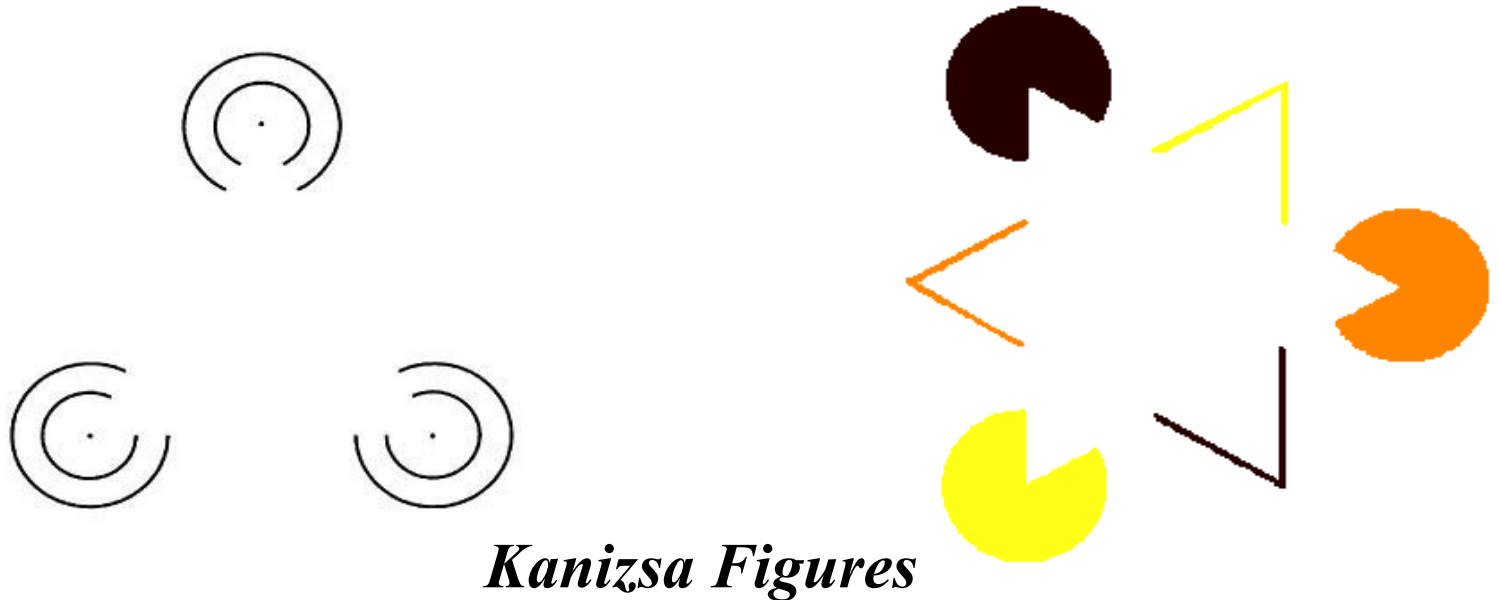
一种倾向，即一些成分和其它成分以这样一种方式连接在一起，以便有可能使一条直线、一条曲线或者一个动作沿着已经确立的方向继续下去。



知觉组织法则

4) 封闭性

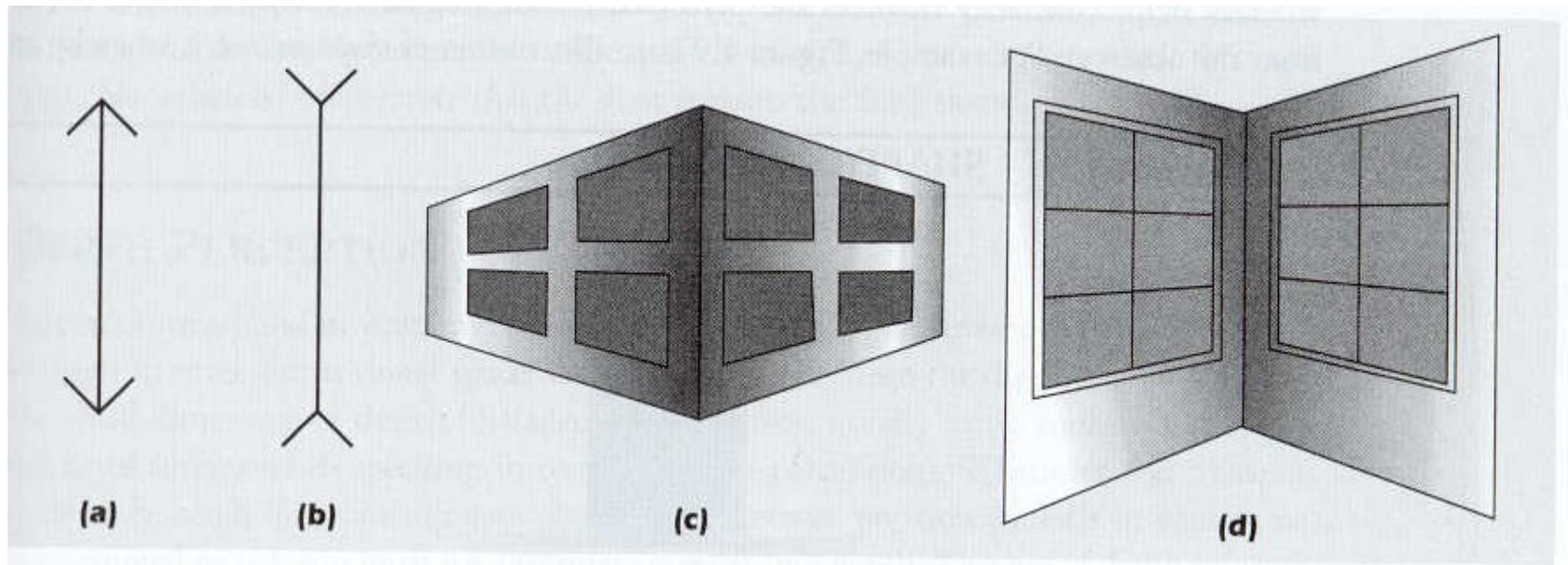
一些成分以这样一种方式组合，以便有助于形成一个更加紧密和更加完整的图形。



知觉组织法则

5) 对称性

一种条件，有利于这样一些组合，即导致对称或平衡的整体而不是非对称的整体。



知觉组织法则

6) 共同的命运：在一个共同的方向上移动或变化的那些成分的组合，而与视域里朝其它方向运动或变化的成分相区别。这个组合原则本质上是相似组合在运动物体上的应用，它是舞蹈设计中的一个重要手段。

7) 常见的形状



内容提要

- 感知过程
- 视觉的生理机制
- 视皮层信息处理
- 格式塔视觉理论
- 拓扑视觉理论
- 马尔的视觉计算理论
- 计算机视觉
- 注意

知觉理论

刺激物说

刺激物说与假设考验说相反，主张知觉只具有直接性质，否认已有知识经验的作用。其著名代表人Gibson(1950, 1966, 1979)认为，自然界的刺激是完整的，可以提供非常丰富的信息，人完全可以利用这些信息，直接产生与作用于感官的刺激相对应的知觉经验，根本不需要在过去经验基础上形成假设并进行考验。

吉布森生态学理论

Gibsonian Theory

美国心理学家吉布森(J. J. Gibson)因其对知觉的研究而闻名于学术界。1950年他提出生态知觉理论。他认为知觉是直接的，没有任何推理步骤、中介变量或联想。生态学理论（刺激物说）与构造理论（假设考验说）相反，主张知觉只具有直接性质，否认已有知识经验的作用。

吉布森生态学理论

Gibsonian Theory

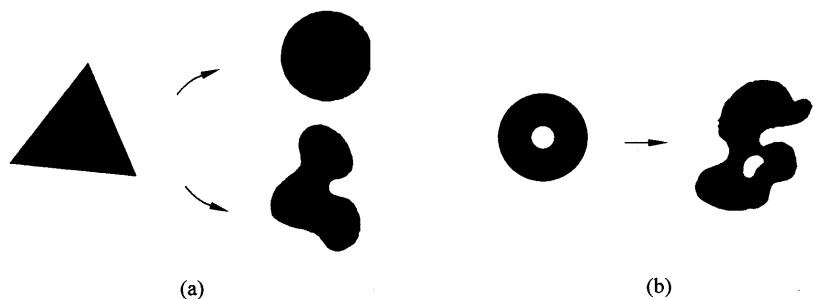
吉布森(J. J. Gibson)的知觉理论之所以冠之以“生态知觉理论”，原因在于它强调与生物适应最有关系的环境事实。对吉布森而言，感觉是因演进而对环境的适应，而且环境中有些重要现象，如重力、昼夜循环和天地对比等，在进化史上都是不变的。

拓扑性质初期知觉

1982年，陈霖在《科学》杂志上就知觉过程从哪里开始的根本问题，原创性地提出了“拓扑性质初期知觉”的假说。这是他在视知觉研究领域的独创性贡献，向半个世纪以来占统治地位的初期特征分析理论提出了挑战。与传统的初期特征分析理论根本不同，拓扑性质初期知觉理论从大范围性质到局部性质的不变性知觉的角度，为理解知觉信息基本表达的问题提出了一个理论框架。

拓扑性质初期知觉

拓扑学研究的是在拓扑变换下图形保持不变的性质和关系，这种性质和关系就称为拓扑性质。所谓拓扑变换是一对一的连接变换，它可以形象地想象成橡皮薄膜的任意变形，只要不把薄膜剪开或不把薄膜的任意两点粘合起来。



拓扑性质初期知觉

2005年，陈霖在“Visual Cognition”第四期上发表长达88页的“重大主题论文”，对拓扑视觉理论概括为：“知觉组织的拓扑学研究基于一个核心思想和包括两个方面。核心思想是，知觉组织应该从变换(transformation)和变换中的不变性(invariance)知觉的角度来理解。两个方面是，第一方面强调形状知觉中的拓扑结构，这就是，知觉组织的大范围性质能够用拓扑不变性来描述；第二方面进一步强调早期拓扑性质知觉，这就是，拓扑性质知觉优先于局部特征性质的知觉。

拓扑性质初期知觉意义

- 美国科学院院士R. Desimone (2005) 认为：陈霖的大范围到局部的知觉拓扑模型是和神经生物学体系一致的，也是和心理学行为实验的物体基本表达的证据一致的。
- 拓扑性质知觉模型将会激励神经生物学家对视觉系统特征分析采取新的研究方向，即强调物体特征是变换下不变性质的研究方向。

知觉组织的理论架构

Wilson和Boaz(2000)提出了知觉组织的一个理论架构，认为知觉组织由四个部分组成：

- 层次表示
- 基元的探测
- 图像各部分间相似性和差异性的探测
- 图像各部分的分组机制。

知觉组织的理论架构

- 层次表示

它描述了知觉组织许多层次上“部分”到“整体”的关系。即：一个层次上的“整体”常是下一层的“部分”。然而，有证据表明，层次表示生成的过程并不严格遵循从局部到全局或从全局到局部的方式执行

知觉组织的理论架构

基元的探测：感受野匹配

- 视觉皮层的简单细胞和复杂细胞的感受野近似于一个过滤器。一个简单细胞的兴奋反应表明存在一个具有一定位置、朝向、大小和相位相似于它的感受野属性的局部图像图案。感受野匹配也可能出现在除了主视觉皮层外的其它区域，可能包含除了局部亮度或色彩轮廓外的其它图像结构的探测，例如相位不连续结构和简单的极端对称图案。
- 感受野匹配为后续的知觉组织机制提供基元。

知觉组织的理论架构

相似性/差异性探测机制

- **变形匹配** 把一个图像区域用作与另一个图像区域比较的变形模板。如果应用一个合理的变形下，使得区域匹配很好，那么图像区域间的相似性被探测。而且，产生最佳匹配的某个变形提供了图像区域间差异性的信息。
- **属性匹配** 另一个比较分组的有名的方法是测度这些分组的属性，然后通过被测度属性上的差异表示各分组间的差异。这些属性可能是简单的测量，如一个分组中基元的颜色、位置、朝向或大小的均值和方差，或者可能是更复杂的测量，如不变形状矩等。

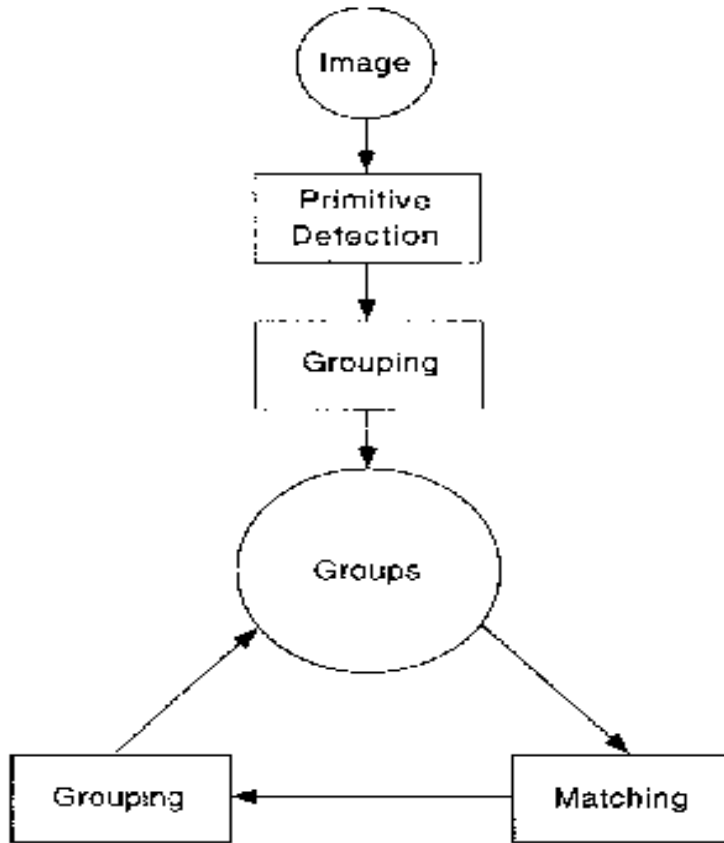
知觉组织的理论架构

分组机制

根据图像各部分间的相似性和差异性，这些部分可组织成整体，或者被重新分割成不同的部分的集合，符合广义唯一性原理。

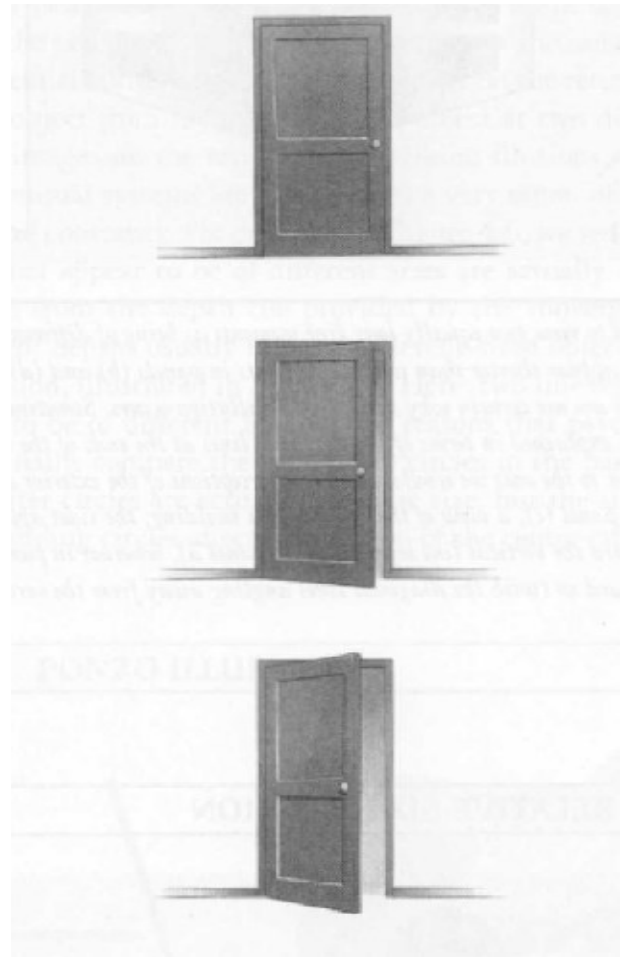
- 传递分组
- 非传递分组
- 多级分组。

知觉组织的理论架构



视觉不变性

如果你认识了一个人，
那么无论他穿什么衣服，
离你或远或近，
甚至没有面对着你，
你是否都能把他认出来？

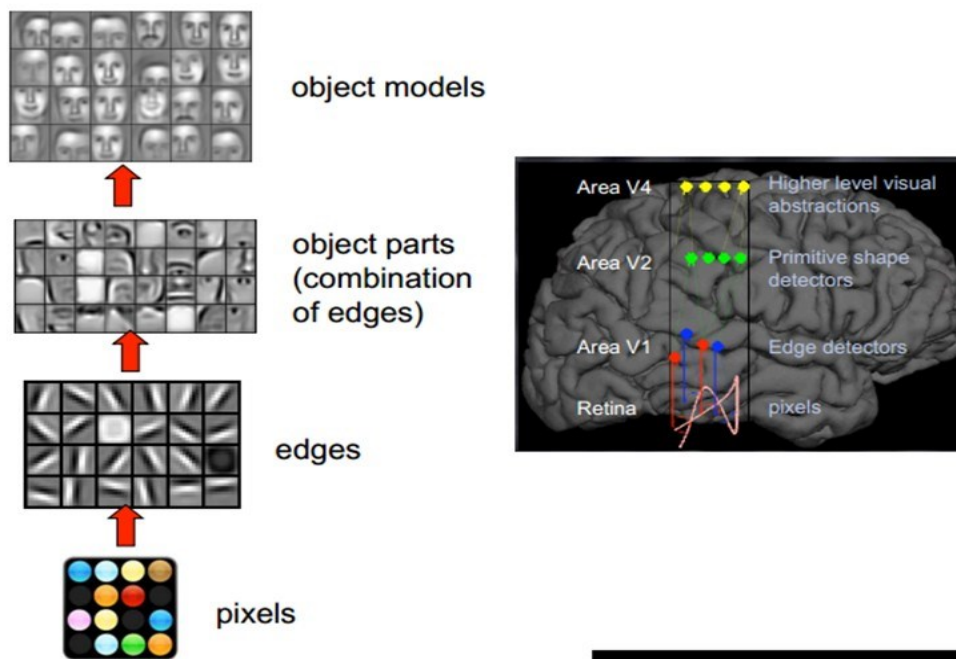


多重视觉区 (Zeki)

- V1 and V2: 响应颜色和形状
- V3 and V3A: 形状和运动
- V4: 颜色
- V5: 运动

人脑视觉机理

从原始信号摄入开始（瞳孔摄入像素 **Pixels**），接着做初步处理（大脑皮层某些细胞发现边缘和方向），进行抽象（大脑判定，眼前物体的形状），然后与物体模型比较（大脑进一步判定人脸的情况）。



关键问题

- **绑定问题The “binding problem”:**

如果对给定物体的感知分布在整个视觉区域，大脑如何结合特征产生统一的感知？

- **意识与无意识知觉:**

对物体或物体性质的感知可能分为有意识和无意识的感知。

内容提要

- 感知过程
- 视觉的生理机制
- 视皮层信息处理
- 格式塔视觉理论
- 拓扑视觉理论
- 马尔的视觉计算理论
- 计算机视觉
- 注意

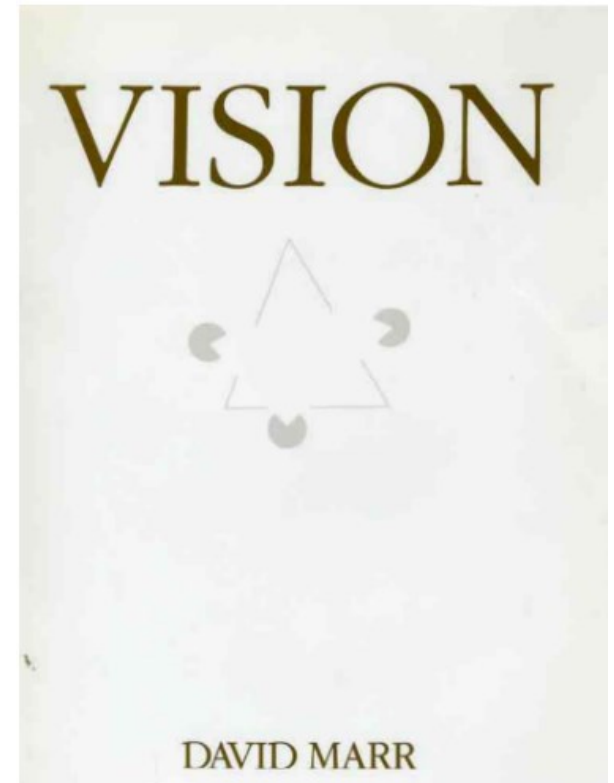
视觉计算理论发展

计算机视觉经历了以下4个主要发展历程：

- 1982年 马尔提出视觉计算理论.
- 以 Aloimonos, Bajcsy等为代表的学者在上世纪八十年代末、九十年代初提出了“目的视觉”、“主动视觉”、“定性视觉”等理论。
- 上世纪90年代，以Faugeras 和Hartley等提出的分层重建理论
- 基于学习的视觉是近年来计算机视觉的研究热点。

马尔的视觉计算理论

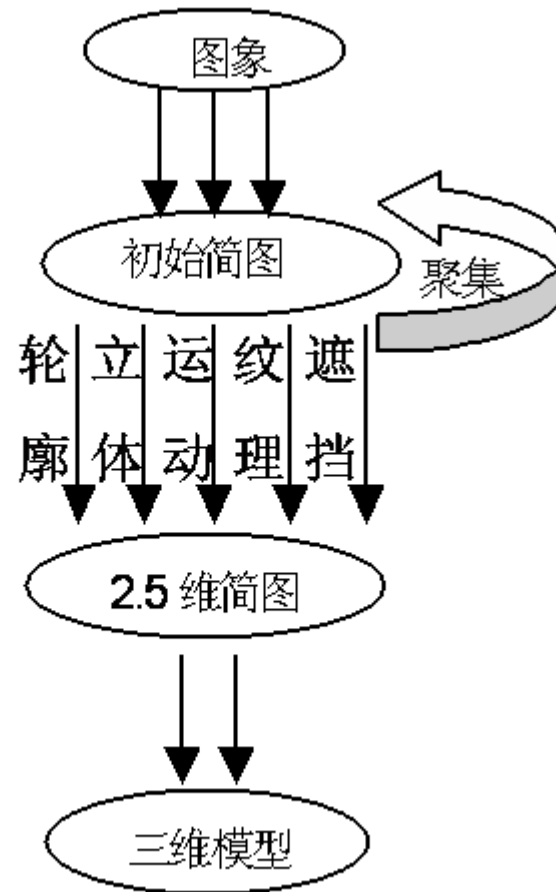
- 信息处理有三个研究层次
 - 计算理论
 - 对什么信息进行计算和为什么要进行这些计算
 - 算法
 - 如何进行所要求的计算，也就是要设计特定的算法
 - 实现算法的机制或硬件
 - 研究完成某一特定算法的计算机构。



视觉系统的表象

- 表象
 - 是指一种能把某种实体或某几类信息表达清楚的形式化系统，以及说明该系统如何行使其职能的若干规则
 - 例如，阿拉伯数制，二进制数制
 - 叁拾柒等于 $37=3 \times 10^1 + 7 \times 10^0$
- 从图象推得形状信息的过程分成三个表象阶段：
 - 初始简图 (primal sketch)
 - 2.5维简图 (2.5 dimensional sketch)
 - 三维模型 (3D model)

视觉系统的三个表象层次



Marr's 视觉理论

初始简图 灰色描述

2.5维简图

3D三维模型

实物

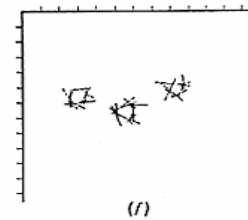
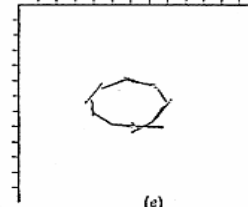
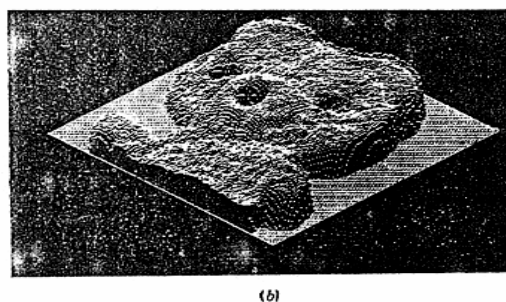
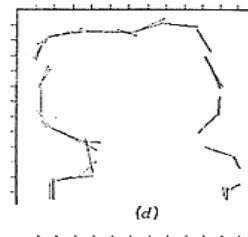
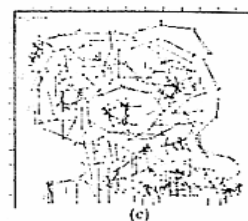
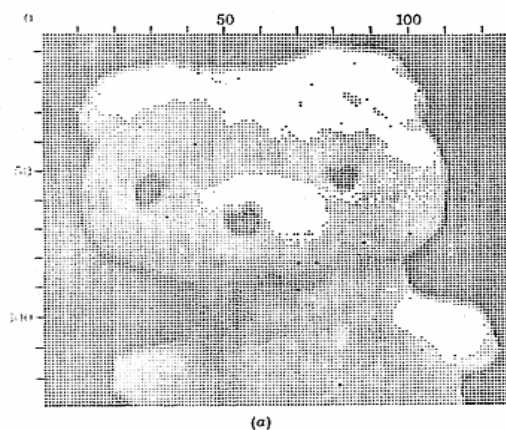


初始简图 (primal sketch)

- 初始简图--是一种基元表象，它由两部分组成
 - 第一部分，用边缘段 (edge segments)、线 (bar)、斑点 (blob) 和端点 (terminations) 这些基元构成对图象中亮度变化的描述，这被称为未处理的初始简图。
 - 第二部分利用虚拟线来完全而明确地表示几何关系，并且通过用各种方式对原始的基元进行选择。聚合和概括等过程来构成的更大、更为抽象的标记 (tokens)。最后得到的描述层次可以复盖某一尺度范围。这被称为完全的初始简图。

初始简图

在灰度图像中，包含两种重要的信息：图像中存在的灰度变化和局部的几何特征。初始简图是一种基元表示法，它可以完全而清楚地表示这些信息。粗略地说，初始简图是以勾划草图的形式来表示图像中的灰度变化。



2.5维简图

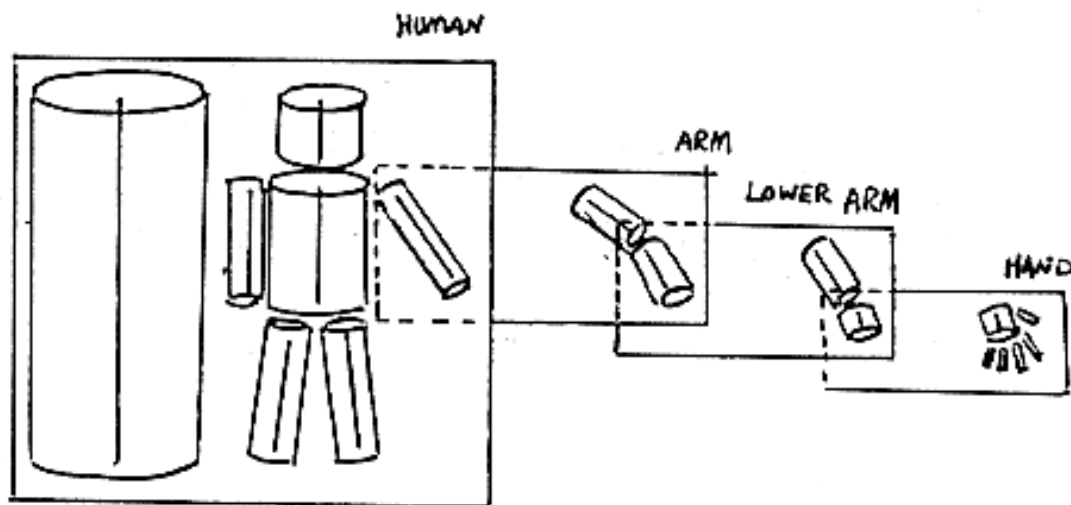
物体表面的局部特性可以用所谓的内在特性来描述。典型的内在特性包括表面方向、观察者到表面的距离，反射和入射光照、表面的纹理和材料特性。内在图像由图像中各点的某项单独的内在特性值，以及关于这项内在特性在什么地方产生不连续的信息所组成。二维半简图可以看成是某些内在图像的混合物。简而言之，二维半简图完全而清楚地表示关于物体表面的信息。

三维模型

- 在三维模型表象中，以一个形状的标准轴线为基础的分解最容易得到。在这些轴线中，每条轴线都和一个粗略的空间关系相联系；这种关系对包含在该空间关系范围内的主要的形状组元轴线提供了一种自然的组合方式。用这种方法定义的模块我们称之为三维模型。

三维模型

- 建立在以物体为中心的坐标系中，这样就不随观察的方位而改变，
- 有两种以物体为中心的坐标系。
 - 一种是用单一坐标系来确定物体上所有点的位置
 - 第二种是分布的坐标系。对物体上的每一个部分或单独的形状特性都有一个局部的坐标系。



由三维模型分级组合构成的人体表象系统

Marr视觉计算理论的特点

- 完全不考虑视觉中的选择性和整体性 (holistic)
- 把低层视觉研究的目标确定为按照各种物理模型和附加约束条件, 根据图象中各点的亮度或其它测量结果恢复景物中表面的各种有关特性, 如表面方向、距离、反射率等, 即重构景物的三维结构
- 在低层视觉信息处理中不需要关于图象中包含内容的先验知识

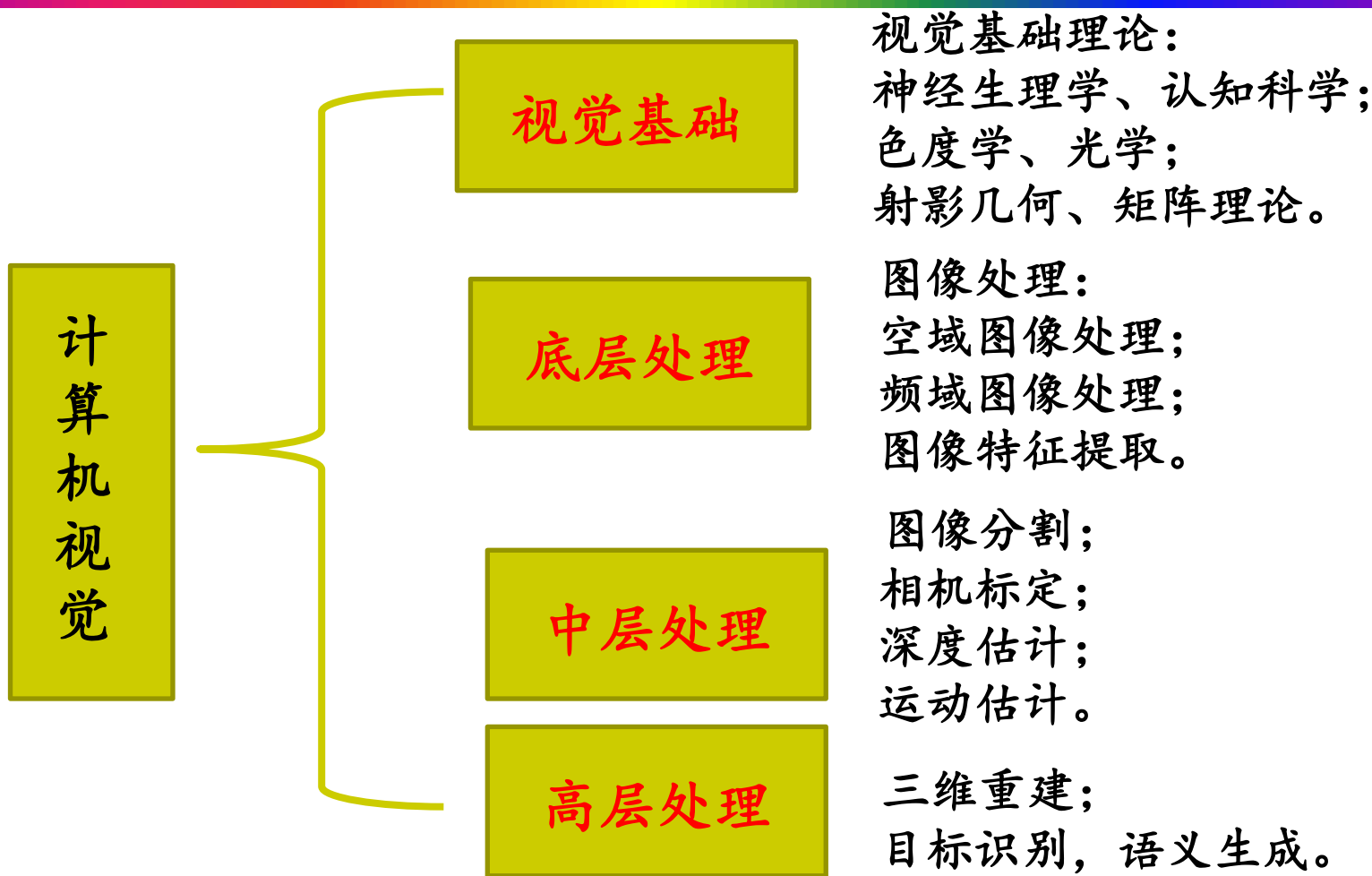
Marr视觉计算理论的问题

- 图象中某点的测量值与相应表面点性质之间的关系是不确定和多义性的。若要根据图象数据确定表面点的特性值必须增加附加的约束条件
- 是以逐点计算方式进行处理带来的难以承受的巨大计算量
- 从整体方式处理问题。

内容提要

- 感知过程
- 视觉的生理机制
- 视皮层信息处理
- 格式塔视觉理论
- 拓扑视觉理论
- 马尔的视觉计算理论
- 计算机视觉
- 注意

计算机视觉



计算机视觉的意义

- ▶ 智能机器：能够模拟人类的功能，感知外部世界并有效解决人所能解决问题的系统。
- ▶ 在人类的感知器官中，视觉获取的信息量最大，大约80%，因此对于发展智能机器而言，赋予机器以人类视觉功能是十分重要的。
- ▶ 计算机视觉：研究用计算机来模拟生物外显或宏观视觉功能的技术学科。
- ▶ 计算机视觉的任务是用图像创建或恢复现实世界模型，然后认知现实世界。
- ▶ 具体来说，让计算机具有对周围世界的空间物体进行传感、抽象、判断的能力，从而达到识别、理解的目的。

图像处理

图像处理(image processing), 用计算机对图像进行分析, 以达到所需结果的技术。又称影像处理。图像处理一般指数字图像处理。数字图像是指用工业相机、摄像机、扫描仪等设备经过拍摄得到的一个大的二维数组, 该数组的元素称为像素, 其值称为灰度值。图像处理技术一般包括

- 图像压缩
- 增强和复原
- 匹配、描述和识别

图像分割

图像分割就是把图像分成若干个特定的、具有独特性质的区域并提出感兴趣目标的技术和过程。它是由图像处理到图像分析的关键步骤。从数学角度来看，图像分割是将数字图像划分成互不相交的区域的过程。图像分割的过程也是一个标记过程，即把属于同一区域的像素赋予相同的编号。现有的图像分割方法主要分以下几类：

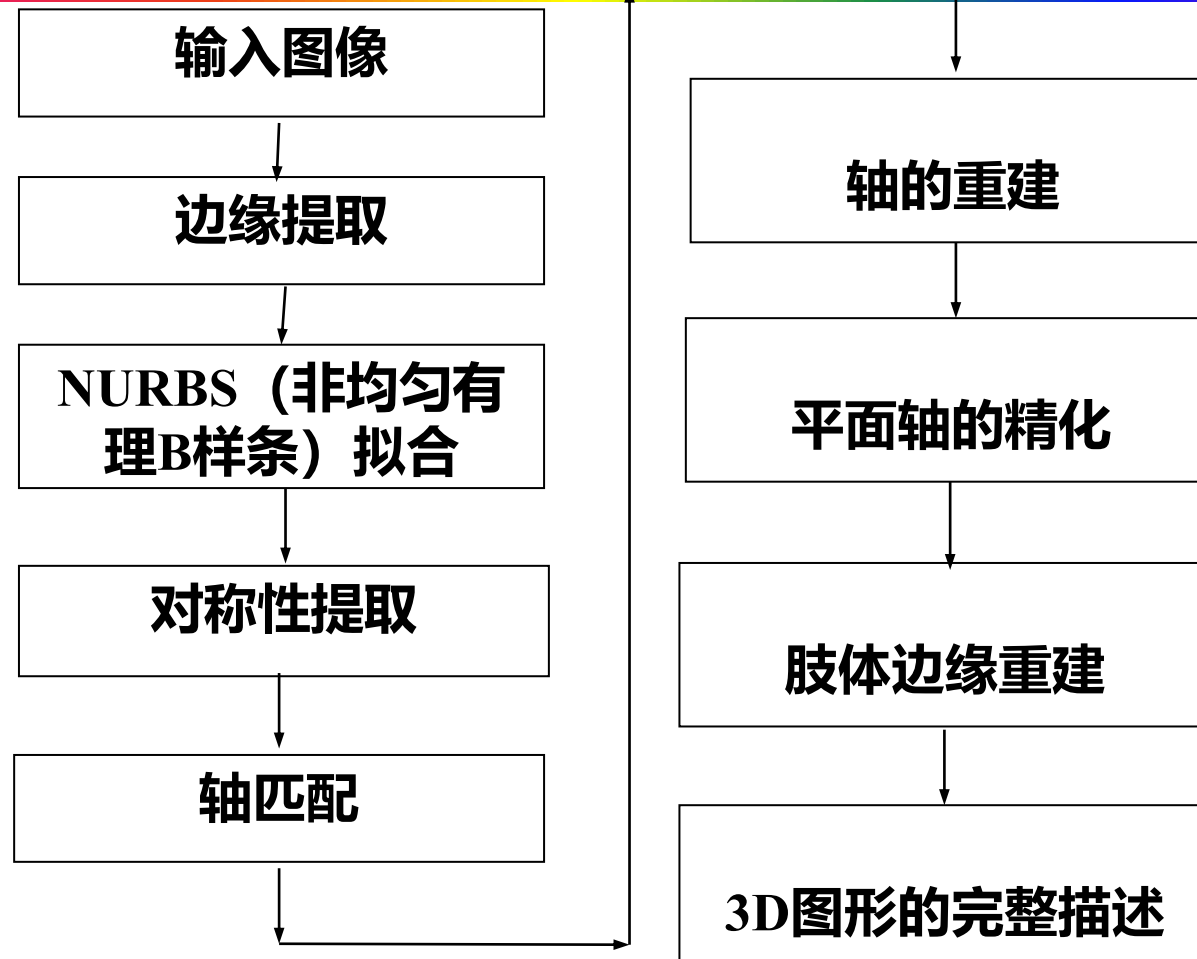
- 基于阈值的分割方法
- 基于区域的分割方法
- 基于边缘的分割方法
- 基于特定理论的分割方法等。

三维重建

✓ **一般圆柱体的三维重建：在圆柱体的横截面和真实的边缘被遮挡的情况下，从图像中重建一般圆柱体**



三维重构系统框架

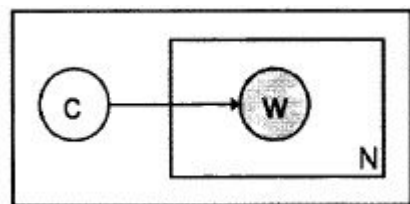


物体识别

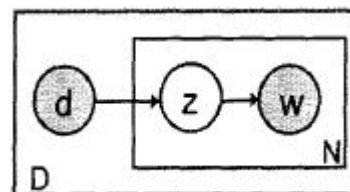
- 物体识别 (Object Recognition) 是计算机视觉领域中的一项基础研究，它的任务是识别出图像中有什么物体，并报告出这个物体在图像表示的场景中的位置和方向。解决低层特征有限的表达能力与用户丰富的语义表达能力之间的差异。
- 目前物体识别方法可以归为两类：基于模型的或者基于上下文识别的方法。
- 对于物体识别方法的评价标准有4个：健壮性 (robustness)、正确性 (correctness)、效率 (efficiency) 和范围 (scope)。

基于统计的方法

一幅图像可以看作是一篇“文档”，而图像中提取出的特征认为是“词语”。生成性的学习方法通过先验知识去拟合并解释图像中的信号。有两种主要的生成性方法，一种是NB（朴素贝叶斯），另外一种是pLSA（概率潜语义分析）与LDA（线性判别分析）。



朴素贝叶斯



概率潜语义分析

基于PLSA的图像自动标注

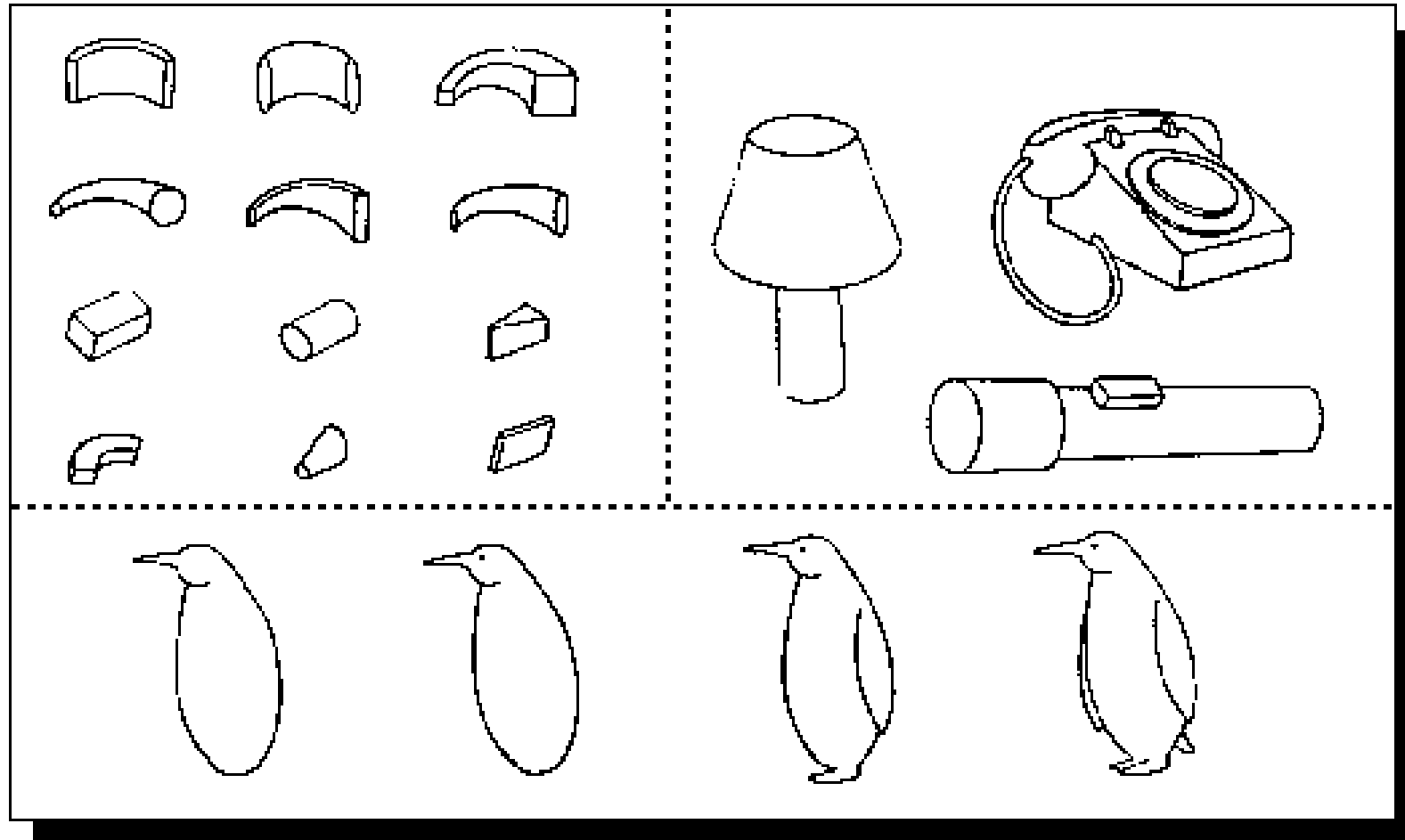
- 首先学习训练集中各个图像的主题分布 $P(z|d)$ ；
- 然后依据这个参数进一步学习视觉词和文本词在给定各个语义主题下的分布 $P(v|z)$ 和 $P(w|z)$ 。由PLSA的条件独立假设，这两个分布**独立于具体的训练图像**，对于训练集之外的图像也是有效的。
- 于是，给定一幅未知图像 d_{new} ，可以依据自动获取的视觉词表示 $v(d_{new})$ 和训练得到的参数 $P(v|z)$ ，使用folding-in算法计算该图像的主题分布 $P(z|d_{new})$ ，从而**计算概率** $P(w|d_{new})$ 并经过排序得到对应的标注关键词集合。

基于物体部件的识别

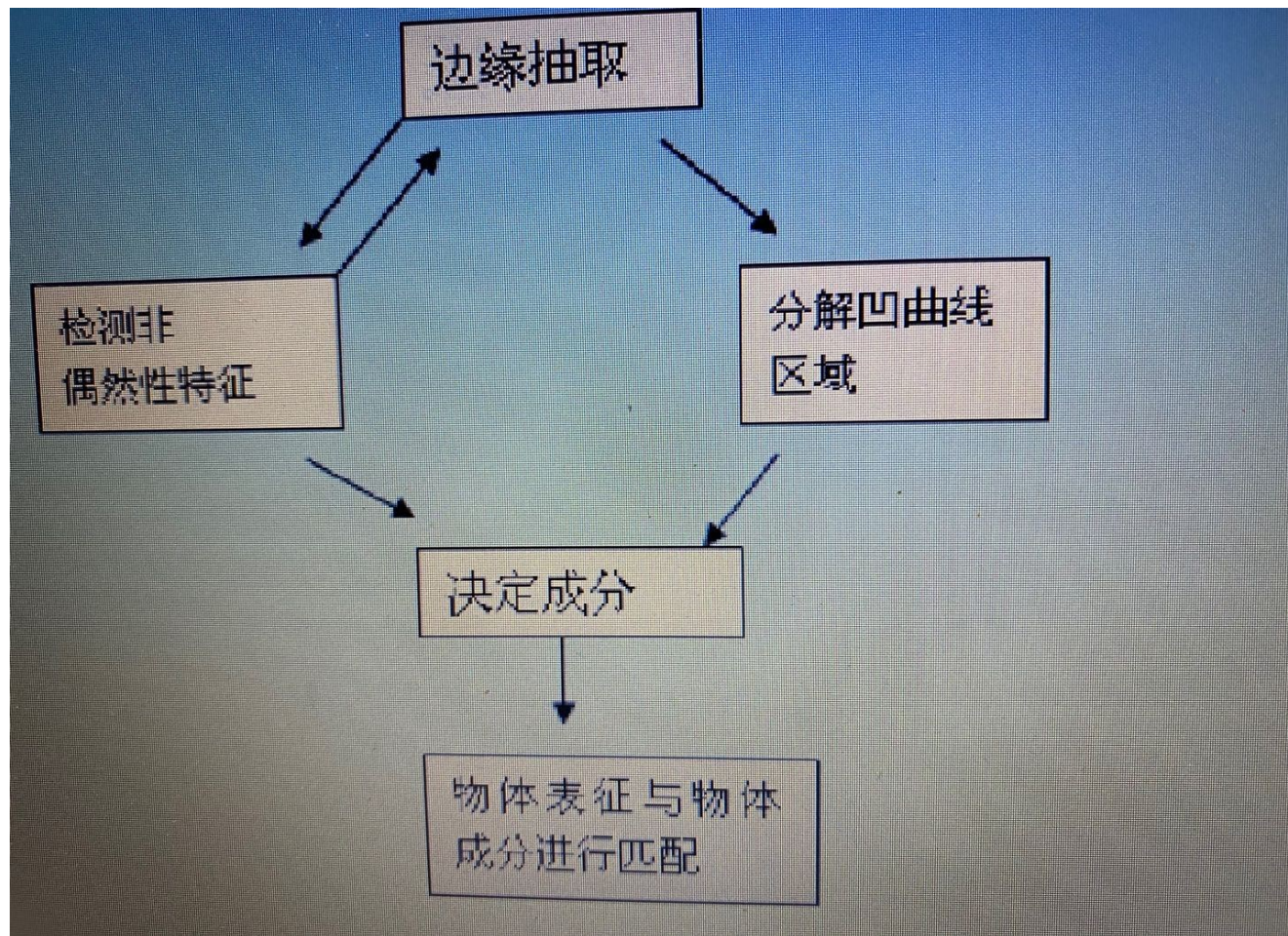
1987年，Biederman提出了一种模式识别的理论—成分识别理论（Recognition-by-components theory，简称RBC理论）。该模型基于这样一种观点，通过把复杂对象的结构拆分为称做简单的部件形状，就可以进行模式识别。

- 对象是由称为“geons”的原语构建的
- 必须确定将视觉对象分割成geons的适当方式
- 决策的两个关键组成部分：
 - 定位凹陷
 - 决定哪些边缘信息在不同视角下保持不变（不变属性，如曲率、平行度等）

基于物体部件的识别



Biederman成分识别理论



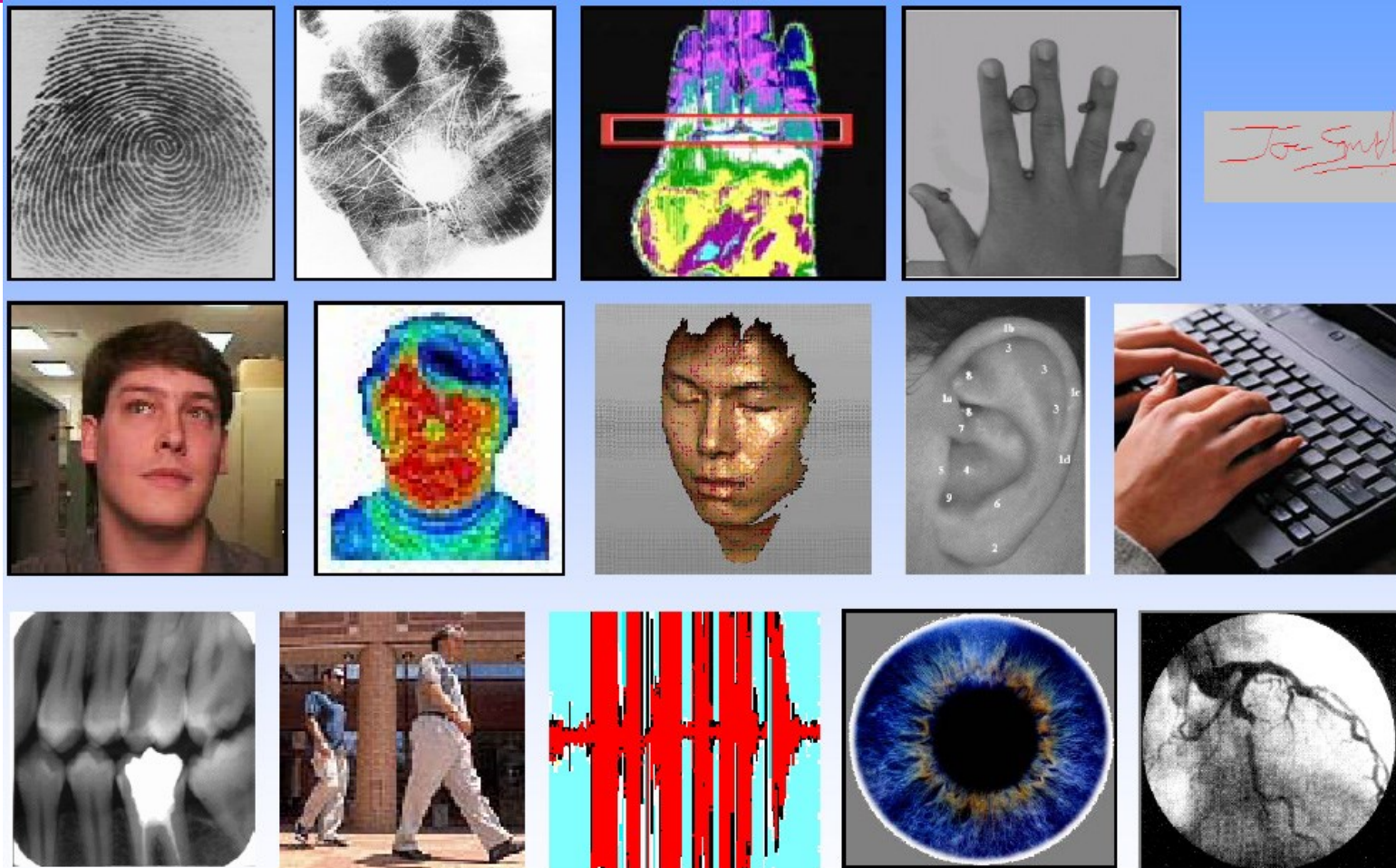
Biederman成分识别理论

- 边缘编码
- 按高阶特征分组
- 与存储的结构知识匹配
- 语义知识的获取
- 命名

视觉感知智能案例研究

- 人脸
- 指纹
- 虹膜
- 步态
- 行为分析

生物特征识别技术 (Biometrics)



生物特征识别技术比较

Biometrics	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	H	L	M	H	L	H	L
Fingerprint	M	H	H	M	H	M	L
Hand Geometry	M	M	M	H	M	M	M
Iris	H	H	H	M	H	L	H
Retinal Scan	H	H	M	L	H	L	H
Signature	L	L	L	H	L	H	L
Voice	M	L	L	M	L	H	L
Facial Thermogram	H	H	L	H	M	H	H

内容提要

- 感知过程
- 视觉的生理机制
- 视皮层信息处理
- 格式塔视觉理论
- 拓扑视觉理论
- 马尔的视觉计算理论
- 计算机视觉
- 注意

什么是注意

- 心理活动对一定的对象的选择性和集中性。
- 通常认为，注意不是一种独立的心理过程，它并没有反映事物及其属性，而被视作心理活动的一种积极的状态。
- 认知心理学主要研究选择性，将注意看作一种内部机制——实现对刺激进行选择的控制和调节行为两大功能。也称选择性注意。



注意的研究

- Moray (1969) 认为注意的研究应包括以下6个方面：
 - (1) 选择性 (selectiveness) ；
 - (2) 集中性 (concentration) ；
 - (3) 搜索 (search) ；
 - (4) 激活 (activation) ；
 - (5) 定势 (set) ；
 - (6) 警觉 (vigilance)

过滤器模型

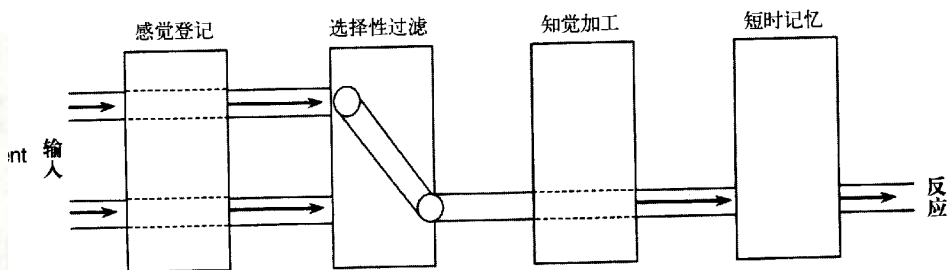
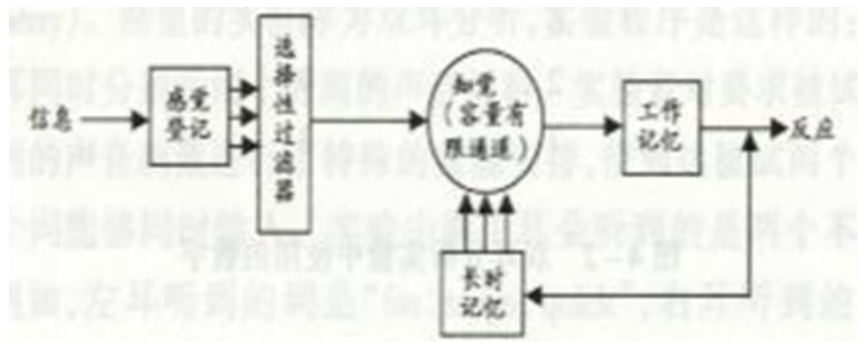
1958年，罗德贝特（D Broadbent）提出过滤器模型，为了避免系统超载，需要过滤器加调节，选择一些信息进入高级分析阶段，而其余信息可能暂存于某种记忆之中，然后迅速衰退。



基本观点

- (1) 瓶颈理论：选择是为了避免系统超载
- (2) 是一种单通道模型：按照“全或无”的方式选择
- (3) 是一种早期模型：依据物理属性（新异刺激、强刺激）在高级分析之前进行选择

过滤器模型



批评

- (1) 无法解释“鸡尾酒会效应”：非追随耳的内容也能够进入高级分析阶段
- (2) 无法解释双耳分听中的意义效应

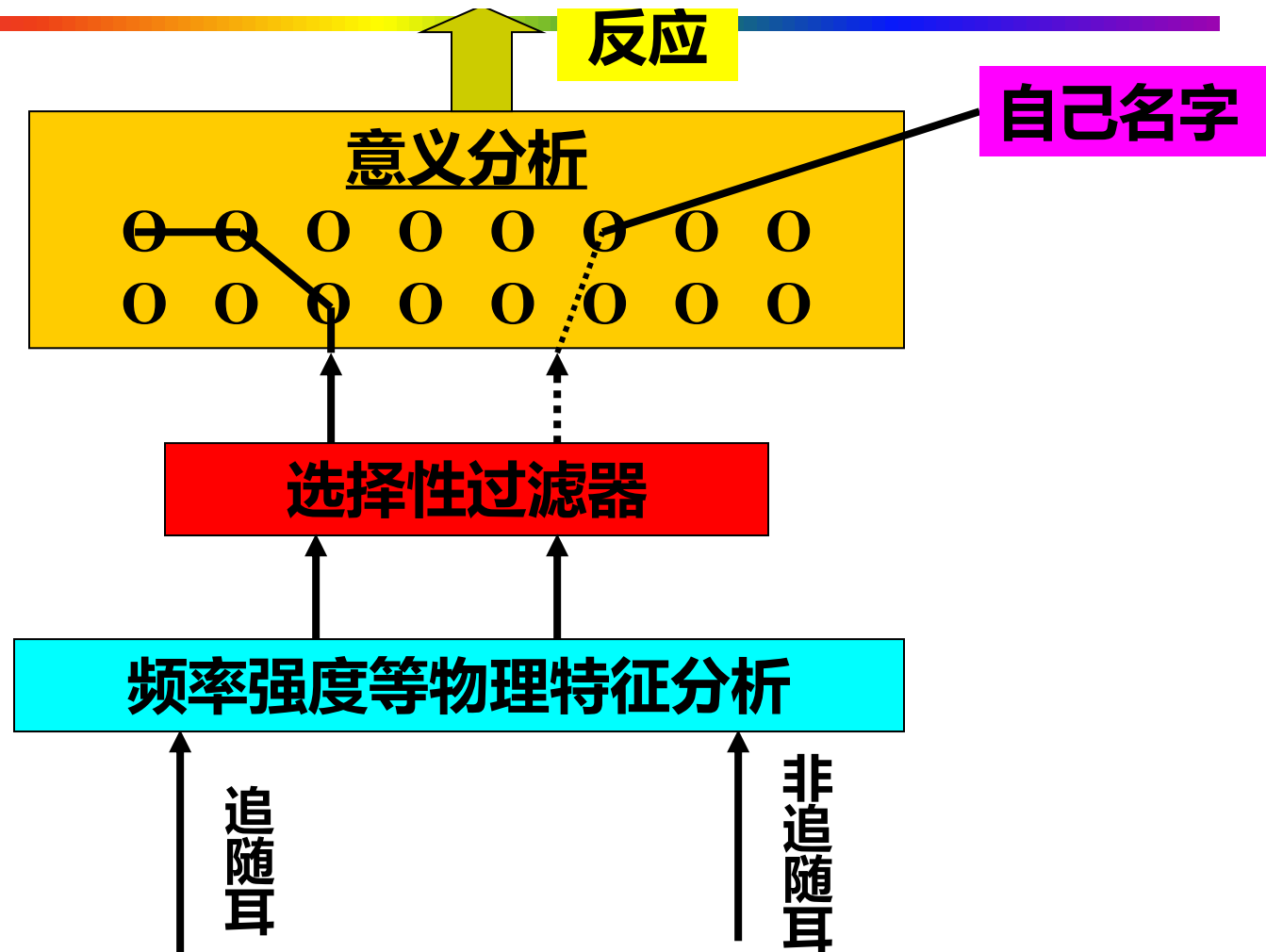
衰减模型 (Attenuation model)

- 由特瑞斯曼 (A Treisman 1960, 1964) 提出, 她认为过滤器不是按“全或无”的方式来工作, 而是既允许追随耳的信息通过, 也允许非追随耳的信息通过, 只是非追随耳的信号受到衰减, 强度减弱了, 但一些信息仍可得到高级加工
- 已经储存在大脑中的信息如字词在高级分析水平有不同的兴奋阈限. 没有衰减的信号可以顺利地激活有关的字词 (如追随耳的信号在通过过滤器时没有受到衰减); 非追随耳的信号由于受到衰减而强度变弱, 常常不能激活相应的字词, 但是特别有意义的项目如自己的名字, 则有较低的阈限, 可受到激活而被识别



衰减模型 (Attenuation model)

Treisman 的衰减模型

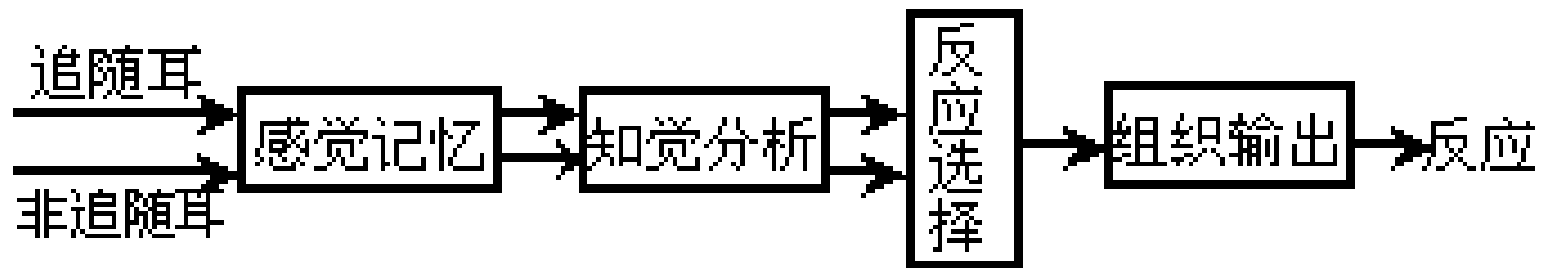


衰减模型 (Attenuation model)

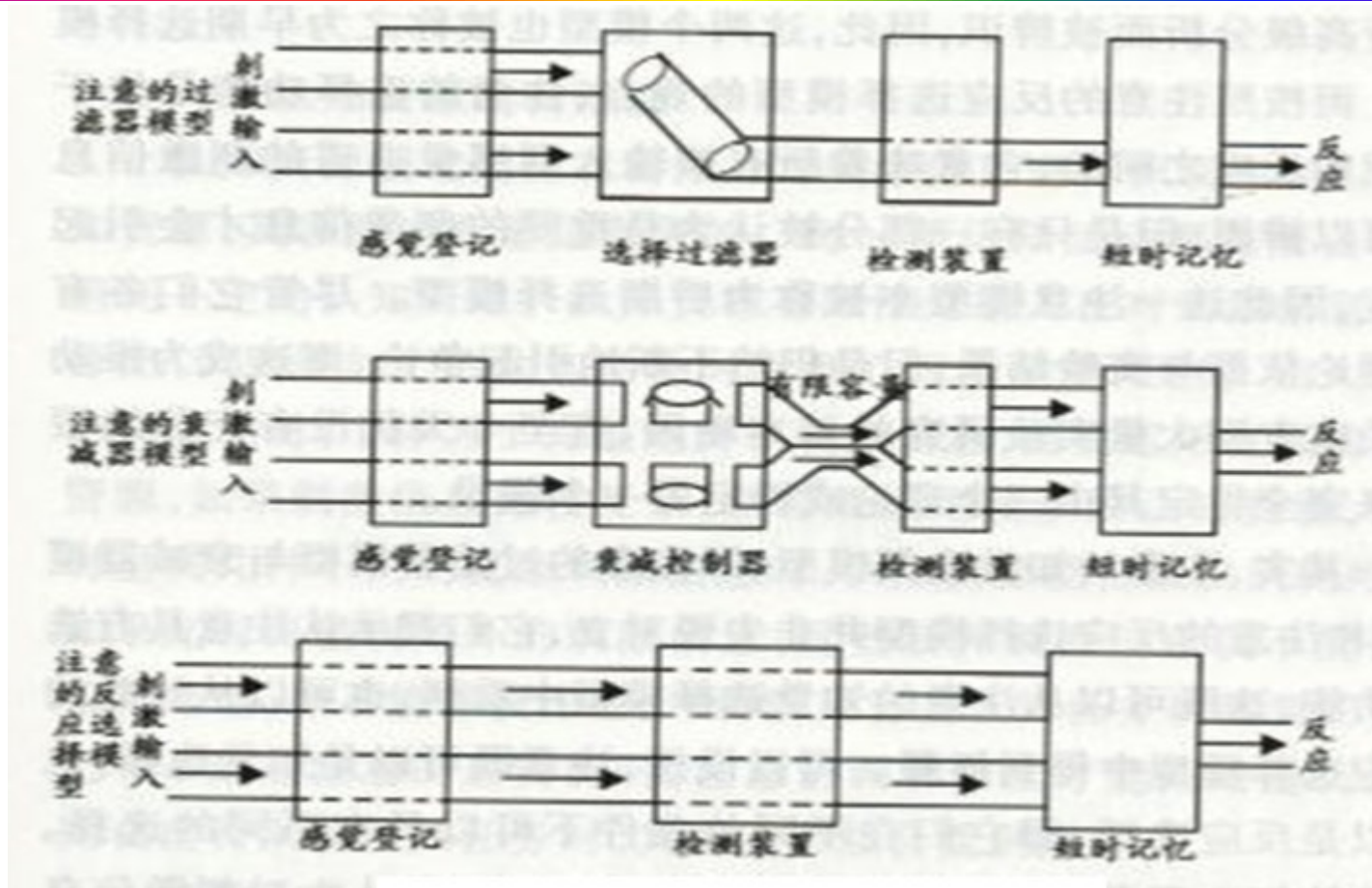
- ~~两个模型的不同之处：衰减模型将过滤器模型的“全或无”工作方式改为衰减，从而由单通道模型改成双通道或多通道模型。~~
- 共同点：第一，这两个模型的根本点是相同的，即都认为高级分析水平的容量有限或通道容量有限，必须通过过滤器调节；第二，这种过滤器的位置在这两个模型中是相同的，都处在初级分析和高级的意义之间；第三，这种过滤器的作用基本相同，即选择一部分信息进入高级的知觉分析水平，使之得到识别，注意选择都是知觉水平的。因此多倾向于将这两个模型合并，称之为Broadbent-Treisman 过滤器-衰减模型，并将它看作注意的知觉选择模型。

注意的反应选择模型

由 Deutsch 和 Deutsch 于1963年提出，他们认为，几个输入通道的信息均可进入高级分析水平，得到全部的知觉加工，注意不在于选择知觉刺激，而在于选择对刺激的反应。他们设想中枢的分析结构可以识别一切输入，但输出是按重要性来安排的，对重要的刺激才会作出反应，对不重要的刺激则不作出反应。该理论认为注意是对反应的选择，因而称之为反应选择模型



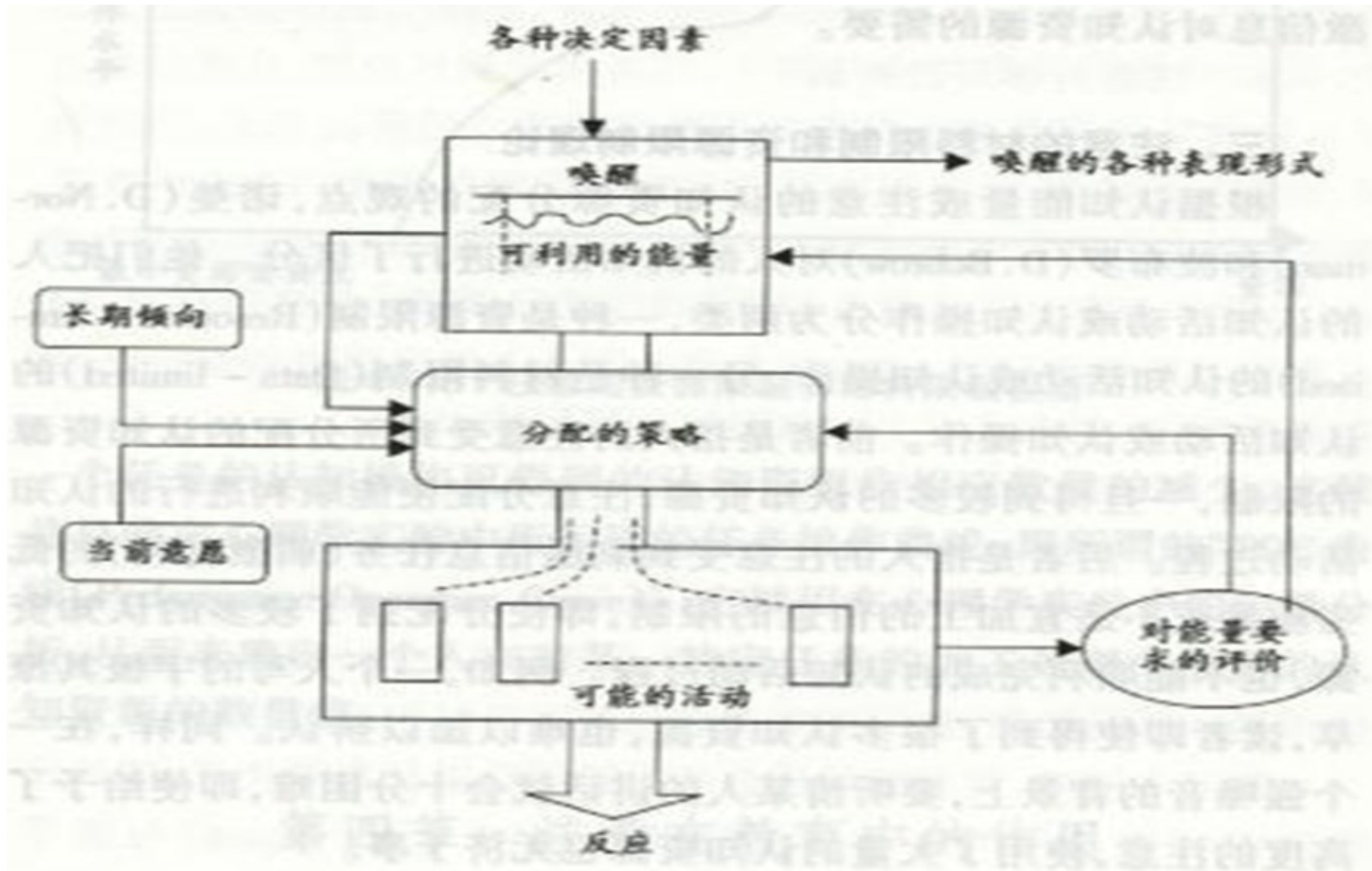
三种注意模型比较



认知资源能量分配

- 1973年卡尼曼(D. Kahneman)提出资源分配模型，认为认知资源有限性是相对的，不是一个固定的数量值，它与唤醒连结在一起，在某段时间内，唤醒水平将决定注意的认知资源数量
- 决定注意的关键是注意认知资源的适宜分配
- 认知资源的分配机制是灵活的，它能根据人的实际需要来调节与控制，优先加工自己认为更为重要的任务
- 因此注意分配受制于个体的唤醒水平、当时的意愿、对完成任务所需认知能力的评估以及个体的某些心理倾向

认知资源分配模型



认知资源分配模型

- 诺曼 (D Norman) 和波布罗 (D Bobrow) 把人的认知活动或认知操作分为两类，资源限制 (Resource-limited) 和材料限制 (Data-limited)
- 资源限制是指由于资源有限，因而一旦它们全部被使用，则没有更多的资源去执行其他任务。资源限制任务是指：如果把执行这项任务的作业分配给较多的资源，那么这项作业便得到改善
- 材料限制任务是指执行这些任务的作业所受到的限制，不是由指派的资源有限造成的，而是由资料的质量问题造成的

认知资源分配模型

分配策略：

第一，反映不随意注意规则的持久倾向（如，给任何新奇的刺激、给突然运动中的任何对象、给其中提到自己名字的任何交谈提供能量）。

第二，瞬时意向；

第三，对要求的评估，有一个规则，当两种活动要求多于现有的可用能量时，完成一种活动；

第四，唤醒效应。能量需要的评价

影响分配的因素

- (1) 可供分配的能量受制于唤醒因素
- (2) 个人的长期倾向
- (3) 当时的任务要求（指导语）和目的
- (4) 完成任务所需能量的评价

认知资源分配模型

1、优点

- (1) 回避了过滤器在哪个阶段上的争论，持更灵活的观点
- (2) 能解释多种注意现象

2、缺点

- (1) 总资源的多少无法确定，一项任务需要多少资源及如何分配难以量化，不能得到整体上的实验证明。
- (2) 忽视了注意内部过程、加工阶段的研究。

聚光灯模型

20世纪80年代初期，特雷斯曼(A. Treisman)提出的特征整合模型把注意和知觉加工的内部过程紧密地结合起来，并用“聚光灯”形象地比喻注意的空间选择性，受到认知心理学家、神经生理学家以及神经计算学家的高度评价。根据这一模型，视觉处理过程被分为两个相互联系阶段，即预注意和集中注意阶段。前者对视觉刺激的颜色、朝向和运动等简单特征进行快速、自动的并行加工，各种特征在大脑内被分别编码，产生相应的“特征地图”。特征中的各个特征构成预注意的表象。预注意加工是一个“自下而上”的信息处理过程，并不需要集中注意。特征地图中的各个特征在位置上是不确定的，要获得物体知觉就需要依靠集中注意，通过“聚光灯”对“位置地图”进行扫描，把属于被搜索目标的各个特征有机地整合在一起，实现特征的动态组装。

聚光灯模型

- 电生理实验表明，集中注意可以引起与被注意事件相关的神经元的同步发放。这种同步发放通常表现为40周左右的同步振荡。这一发现为注意的特征整合模型提供了神经生理证据。然而，注意的“聚光灯”如何从一个位置移向另一个位置，这仍然是一个有待解决的问题。
- 注意的“聚光灯”模型是一个具有广泛影响的认知心理学模型。然而，它并没有被神经生理学家普遍接受。有些理论甚至对注意“聚光灯”的客观存在也提出了疑义。比如，注意的竞争理论就认为，注意机制并不是视野中快速扫描的“聚光灯”，而是大脑的多种神经机制为解决物体竞争和行为控制而产生的涌现特性。

聚光灯理论

- “注意的聚光灯理论” (Spotlight Metaphor)

Crick认为注意就像聚光灯一样在信息处理空间移动和搜索，是被注意的区域（或特征）的信息处理得到增强，而其外的信息则被削弱甚至忽略。

- “注意的整合竞争假设” (Integrated Competition Hypothesis)

Duncan把感觉，记忆，控制肢体运动等诸系统作为一个整体，认为不存在单独的仅用于注意的系统，甚至注意的各功能组分（如裁剪，竞争，整合等）也没有显著的定位；选择是多个脑区相互竞争、协调的结果，注意是解决这些竞争的许多神经机制的一种突现特性。

注意机制的应用场景

机器翻译 (Machine Translation)

图像标注 (Image Caption)

关系抽取 (Entailment Extraction)

语音识别 (Speech Recognition)

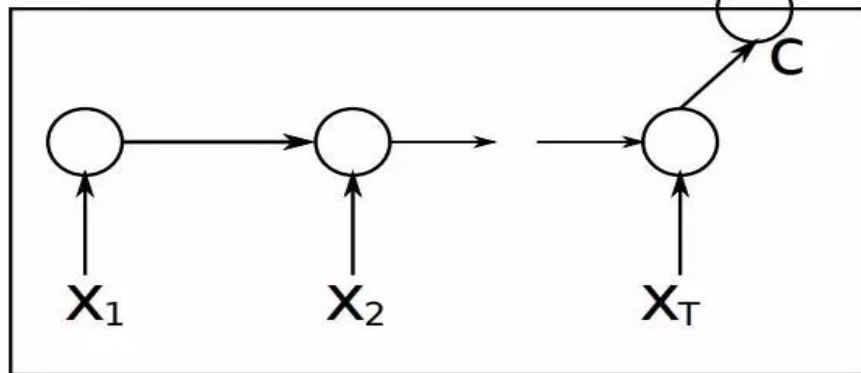
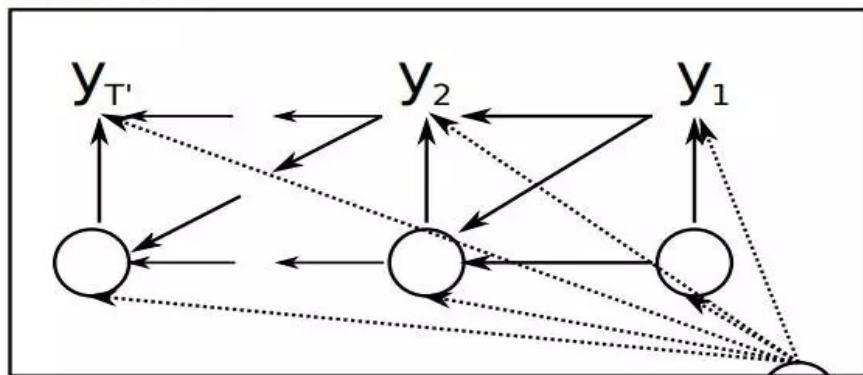
自动摘要生成 (Text Summarization)

深度学习中的注意机制

- Attention Mechanism目前非常流行，广泛应用于机器翻译、语音识别、图像标注（Image Caption）等很多领域，之所以它这么受欢迎，是因为Attention给模型赋予了区分辨别的能力，例如，在机器翻译、语音识别应用中，为句子中的每个词赋予不同的权重，使神经网络模型的学习变得更加灵活（soft），同时Attention本身可以做为一种对齐关系，解释翻译输入/输出句子之间的对齐关系，解释模型到底学到了什么知识，为我们打开深度学习的黑箱，提供了一个窗口。

深度学习中的注意机制

Decoder



Encoder

Encoder把一个变成的输入序列 $x_1, x_2, x_3 \dots x_t$ 编码成一个固定长度隐向量（背景向量，或上下文向量 context） c ， c 有两个作用：1、做为初始向量初始化 Decoder 的模型，做为 decoder 模型预测 y_1 的初始向量。2、做为背景向量，指导 y 序列中每一个 step 的 y 的产出。Decoder 主要基于背景向量 c 和上一步的输出 y_{t-1} 解码得到该时刻 t 的输出 y_t ，直到碰到结束标志（ $\langle \text{EOS} \rangle$ ）为止

<https://blog.csdn.net/u010417>

深度学习中的注意机制

Figure 5. Examples of mistakes where we can use attention to gain intuition into what the model saw.



通过上述Attention Mechanism在图像标注应用的case可以发现，Attention Mechanism与人类对外界事物的观察机制很类似，当人类观察外界事物的时候，一般不会把事物当成一个整体去看，往往倾向于根据需要进行选择性的去获取被观察事物的某些重要部分，

Thank You

