

# 智能科学

---

## -Chapter 5-

### 感知与注意

史忠植

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

---

# 内容提要

---

- 一、概述
- 二、知觉理论
- 三、知觉组织
- 四、Marr视觉计算理论
- 五、知觉有效编码理论
- 六、结束语

# 概述

- 感知是客观外界直接作用于官能的，人而产生的，在发展中经历三种基本形式：
- 感觉
- 知觉
- 表象



# 认识的辩证过程

---

- 关于人的认识的发展过程，列宁曾作过这样的概括：“从生动的直观到抽象的思维，并从抽象的思维到实践，这就是认识真理、认识客观实在的辩证的途径。”毛泽东也指出：“一切比较完全的知识都是由两个阶段构成的，第一阶段是感性知识，第二阶段是理性知识，理性知识是感性知识的高级发展阶段。”

# 认识的辩证过程

---

- 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，

# 从低级感知到高级感知

---

- At the lower level, the world is broken down into primitive features
- We perceive the world as a unified whole
- These holistic perceptions are the result of higher perceptual functions operating on these feature primitives
- Perceiving the world is connected with our knowledge of the world
- *A system that links the products of lower perception to memory*

# 感觉和知觉

---

- 感觉**Sensation**: Conscious outcome of sense organs and projection regions. (“I detect something”, not necessarily conscious)
- 知觉**Perception**: means by which information acquired from the environment via the sense organs is transformed (organized) into experiences of objects, events, sounds, tastes, etc. (“I know, recognize, appreciate what I am sensing”)

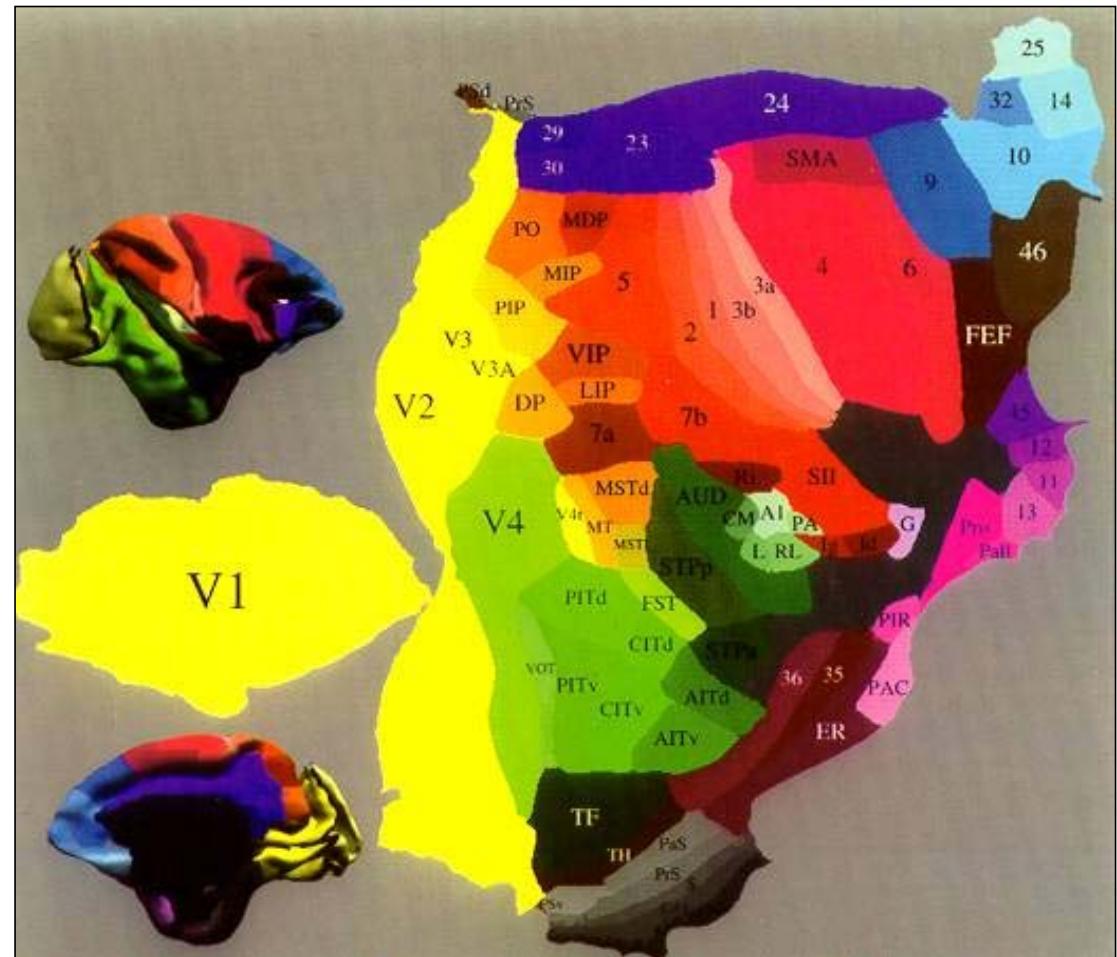
# 知觉组织

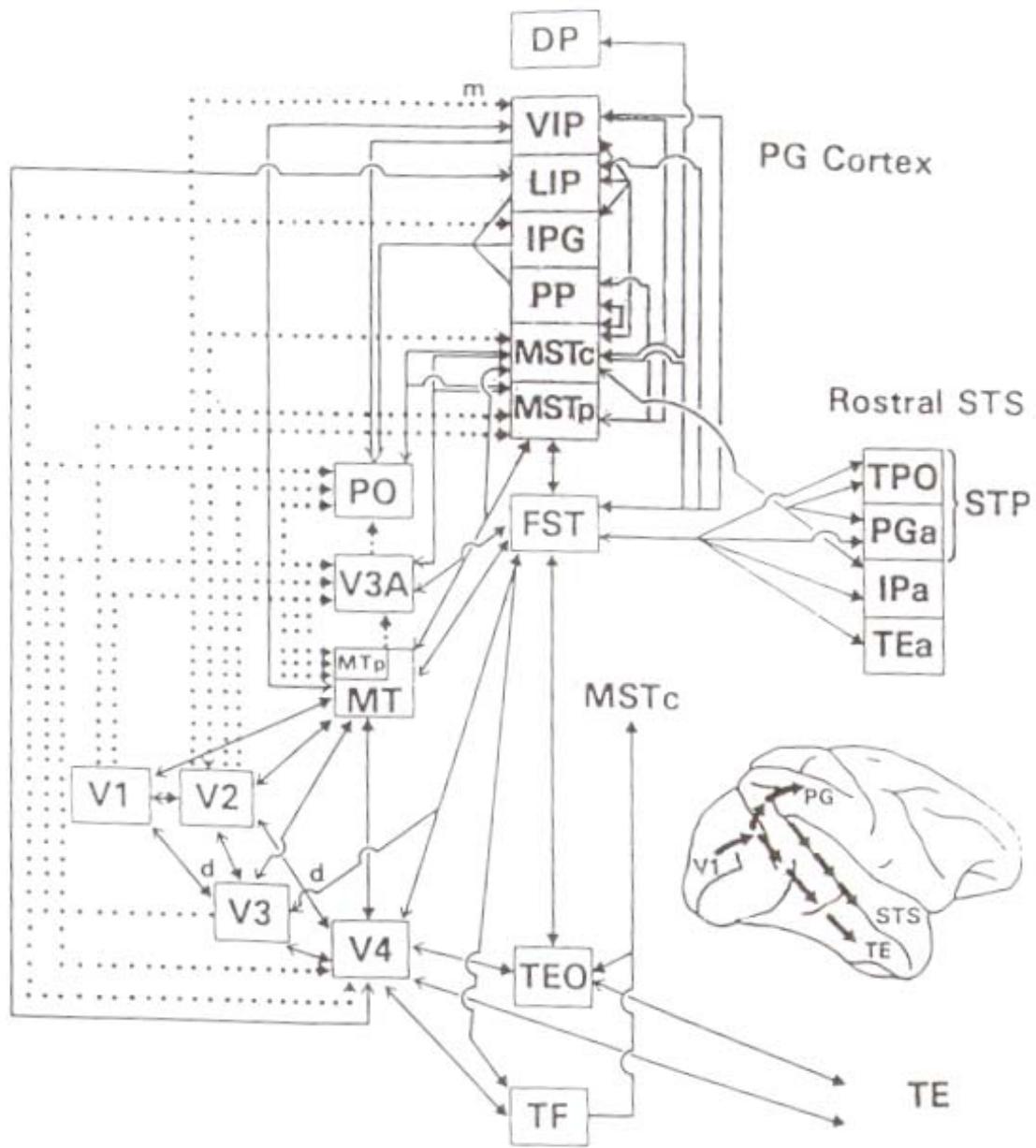
---

- Two complementary issues:
  - 组织*Organizing* coherent objects and events out of segregated sensory/perceptual inputs (“binding”)
  - 分析“*Parsing*” the perceptual world; understanding which inputs belong together and which come from separate objects

# 知觉组织

- *Perceptual organization and parsing accomplished by a variety of visual subsystems*





**General principle:**  
inferior lesions produce perceptual impairments; superior lesions produce syndromes dominated by spatial impairment

# 知觉组织

---

- **Size and Shape Constancies:** reflect mechanisms whereby perception of a distal object stays the same despite proximal changes in the perception
- **Depth Cues:** monocular or binocular sources of information that convey information about relative distance of objects from viewer

## Theories of Perception

- 间接知觉“**Indirect**” perception: perception is the result of interactive processes from without and from within – use of internal representations based on “top-down” and “bottom-up” processing
- 直接知觉“**Direct**” perception (Gibson): information from the visual world is sufficient to permit perception without the involvement of internal representations - focus in “bottom-up” processing

# 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)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。根据他的生态知觉理论，知觉是和外部世界保持接触的过程，是刺激的直接作用。他把这种直接的刺激作用解释为感官对之作出反应的物理能量的类型和变量。知觉是环境直接作用的产物这一观点，是和传统的知觉理论相背离的。吉布森把表示一个人周围刺激的“刺激生态学”概念用公式来表示，它们包括倾斜的和反射的表面的关系，以及人们在走路、坐着和躺下时都感受到的引力，他坚信知觉不变，因此当环境提供给活跃的有机体连续而稳定的信息流时，有机体能够对此作出反应。

# 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。

# 直接观点

## ““Direct” View (Gibson)

- “**Optic array**” contains all necessary visual information
- Layout of objects in space given by **texture gradients, optic flow patterns, and affordances (implied meaning of objects)**
- Perception involves “**picking up**” information through “resonance”
- Has had historical impact in restoring interest in the perceptual environment
- Has been criticized as being underspecified, and neglects role of knowledge in stimulus exploration

# 间接观点

---

- **Context effects in perception:** what is seen depends upon surround
- **Effects of emotion**
- **Visual illusions:** idea is that general knowledge about objects is applied inappropriately to the perception of two-dimensional figures (e.g., Muller-Lyer)

# 构造理论

---

- 布鲁纳（J.S. Bruner）提出构造理论，认为所有感知都受到人们的经验和期望的影响。构造论者关于知觉的看法是把记忆的作用赋予极大的重要性。他们认为先前经验的记忆痕迹，加到此时此地被刺激诱导出来的感觉中去，因此就构造出一个知觉象。而且，构造论者主张有组织的知觉基础是从一个人的记忆中选择、分析并添加刺激信息的过程，而不是格式塔论者所主张的大脑组织的天生定律所引起的自然操作作用。

# 构造理论

- 已有的知识经验对知觉的影响是多方面的，最引人注目的是体现为上下文的作用。当前一些心理学家认为，总的看来，过去的知识经验主要是以假设、期望或因式的形式在知觉中起作用的。人在知觉时，接收感觉输入，在已经验的基础上，形成关于当前的刺激是什么 或者激活一定的知识单元而形成对某种客体的期望 。知觉是在这些假设、期望等的引导和规划下进行的。依照格雷哥里 (Gregory) 的看法，知觉是一种包含假设考验的构造过程。人通过接收信息、形成和考验假设，再接收或搜寻信息，再考验假设，直至验证某个假设，从而对感觉刺激作出正确的解释 ，这被称作知觉的假设考验说。照这个学说看来，感觉刺激的物理特征 、刺激的上下文和有关的概念都可激活长时记忆中的有关知识而形成各种假设。知觉因而是以假设为纽带的现实刺激信息和记忆信息相结合的再造。在通常情况下，人们在知觉时意识不到假设的参与，但在某些特殊条件下，如在弱的照明下看东西，有时是可以体验到这种假设考验的

# 构造理论



- 假设考验说赋予知觉过程以主动性和智慧性的色彩，是目前在认知心理学中有相当影响的知觉理论
- 知觉的假设考验说是一种建立在过去经验作用基础上的知觉理论。支持这个理论的还有其他的重要论据。例如，外部刺激与知觉经验并没有一对一的关系，同一刺激可引起不同的知觉，不同的刺激却又可以引起相同的知觉。

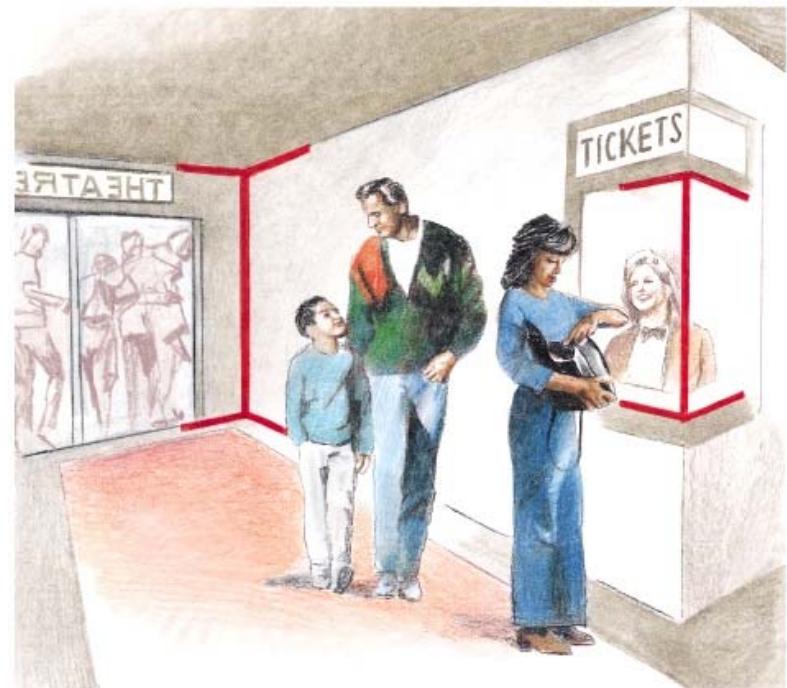
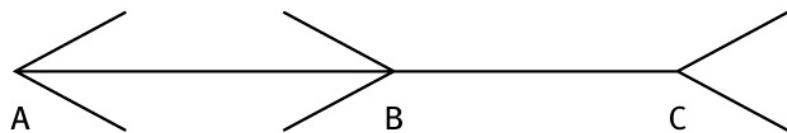
# 格式塔理论



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个知觉的学说，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

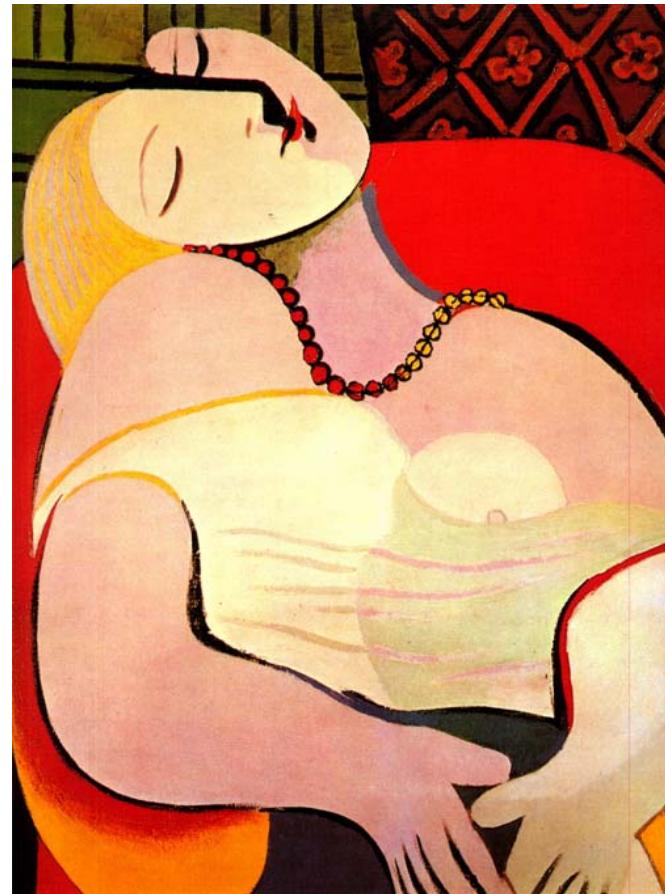
- 知觉的形成是基于先验知识的刺激组织



# 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  
painting

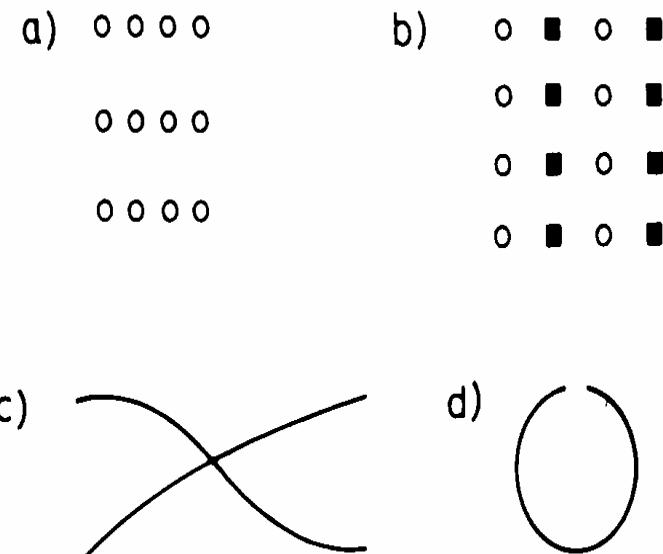


# 知觉组织法则

- Wertheimer 等格式塔心理学家说明了人类知觉组织能归结为几个重要的法则：

## 1) 邻近性。

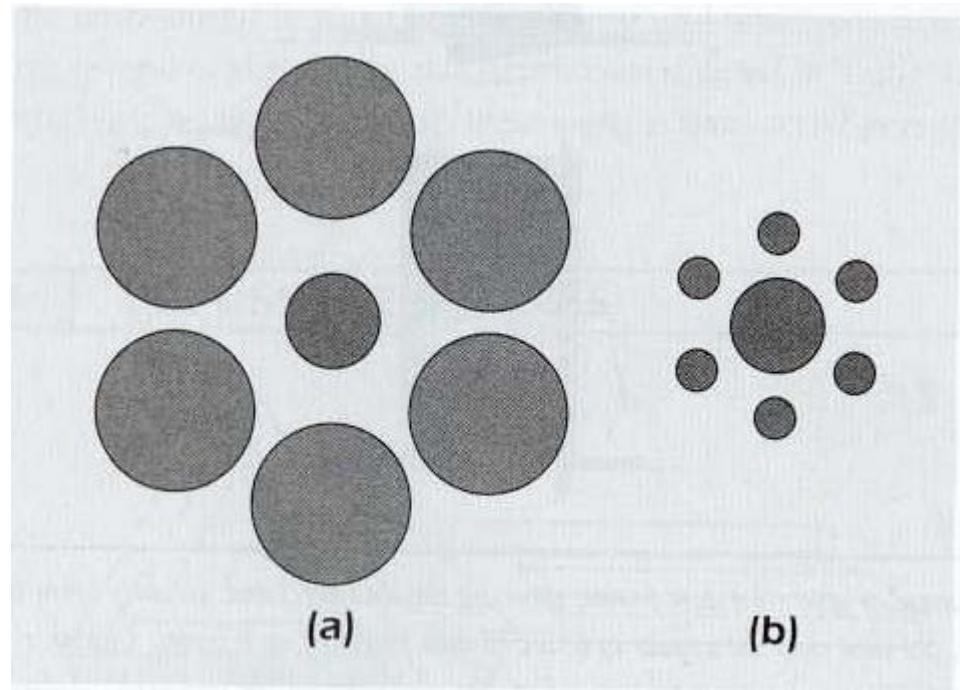
彼此紧密接近的刺激物比相隔较远的刺激物有较大的组合倾向。接近可能是空间的，也可能是时间的。按不规则的时间间隔发生的一系列轻拍响声中，在时间上接近的响声倾向于组合在一起。由于接近而组合成的刺激不必都是同一种感觉形式的；例如，夏天下雨时，雷电交加，我们就把它们知觉为一个整体，即知觉为同一事件的组成部



# 知觉组织法则

- 2) 相似性

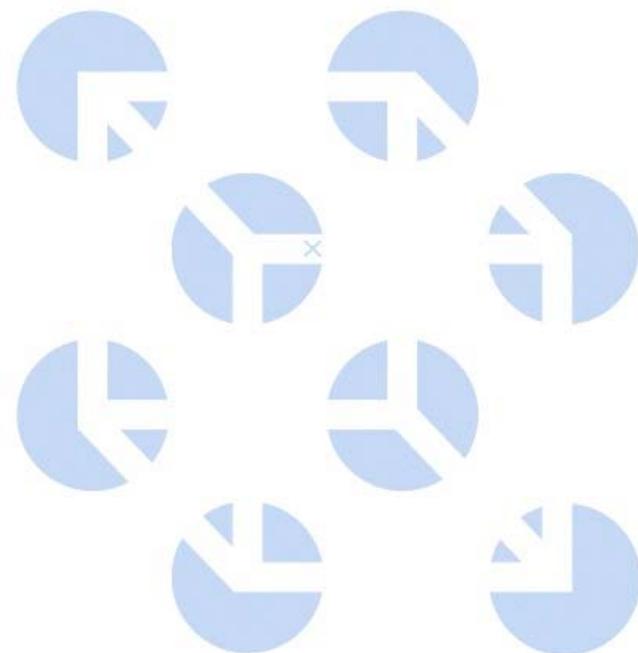
彼此相似的刺激物比不相似的刺激物有较大的组合倾向。相似意味着强度、颜色、大小、形状等这样一些物理属性上的类似。俗话说：“物以类聚，人以群分”，也就包含这种原则。”



# 知觉组织法则

## 3) 好的连续性

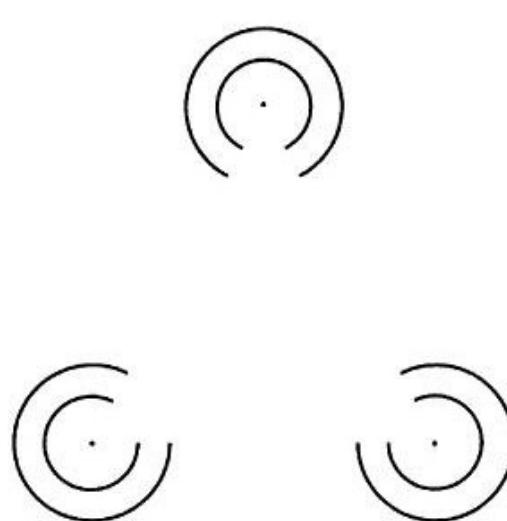
一种倾向，即一些成分和其它成分以这样一种方式连接在一起，以便有可能使一条直线、一条曲线或者一个动作沿着已经确立的方向继续下去。



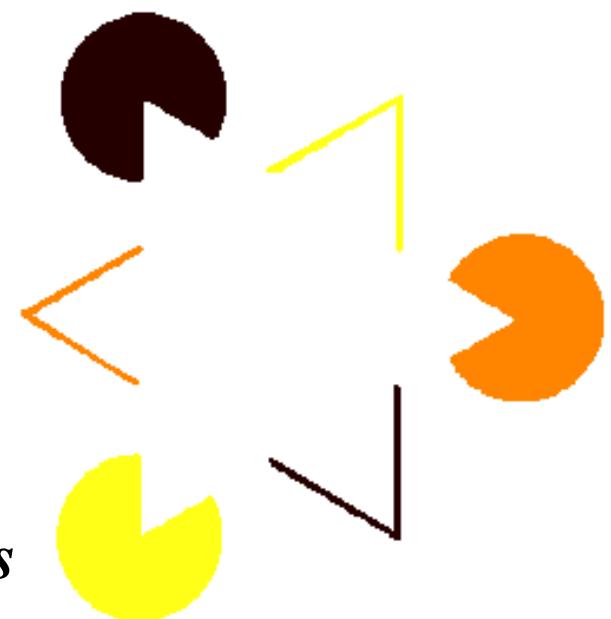
# 知觉组织法则

## 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。



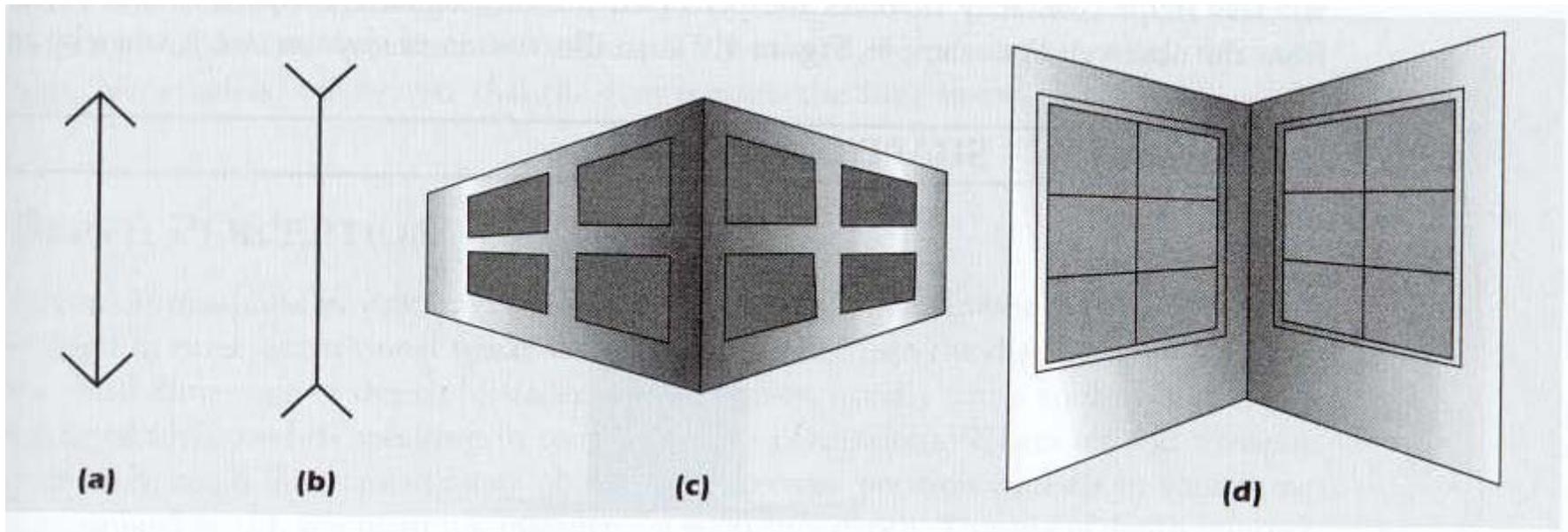
*Kanizsa Figures*



# 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。



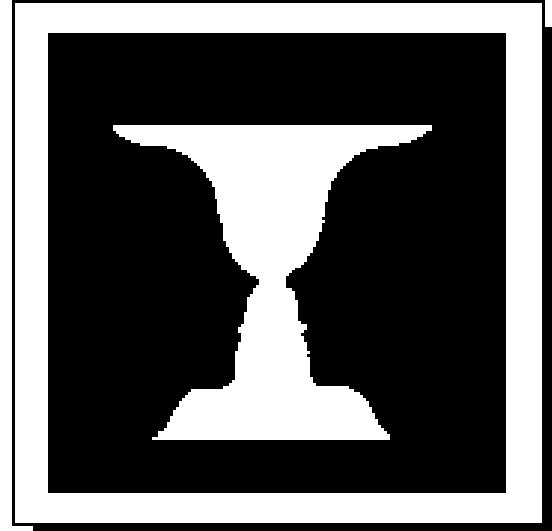
# 知觉组织法则

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

2007-7-20

史忠植 感知过程



# 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和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** 其思想是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，那么图像区域间的的一定的相似性被探测。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

# 知觉组织的理论架构



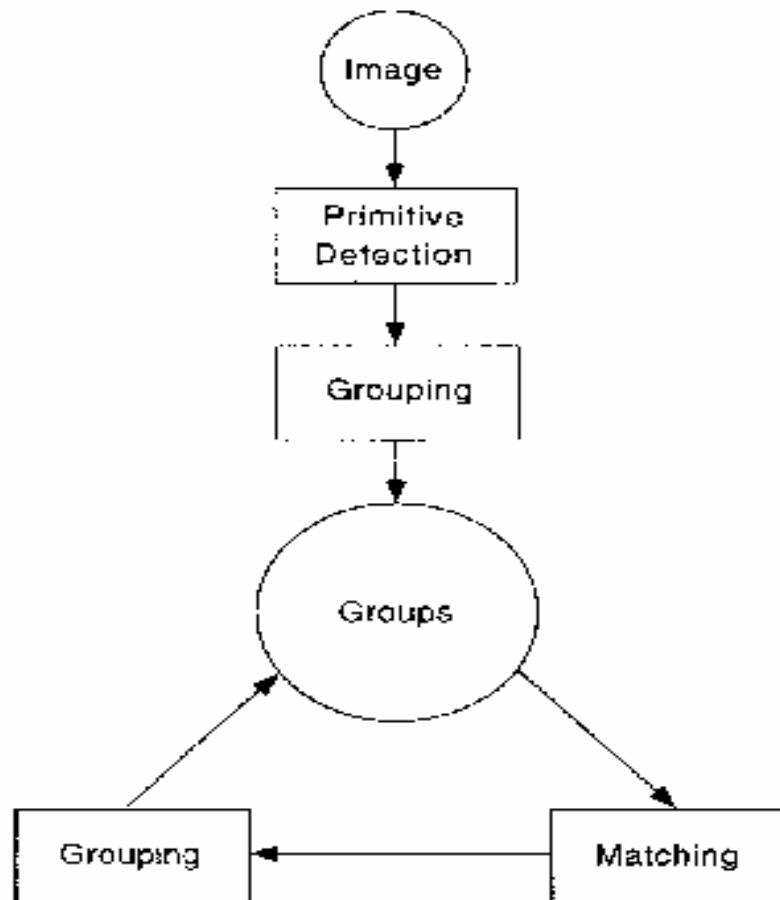
## 分组机制

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。

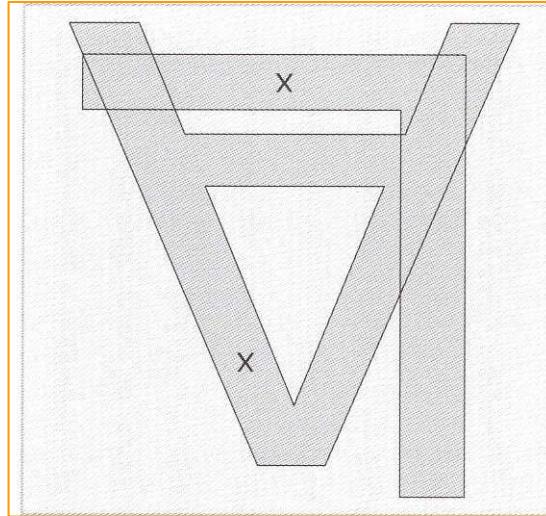
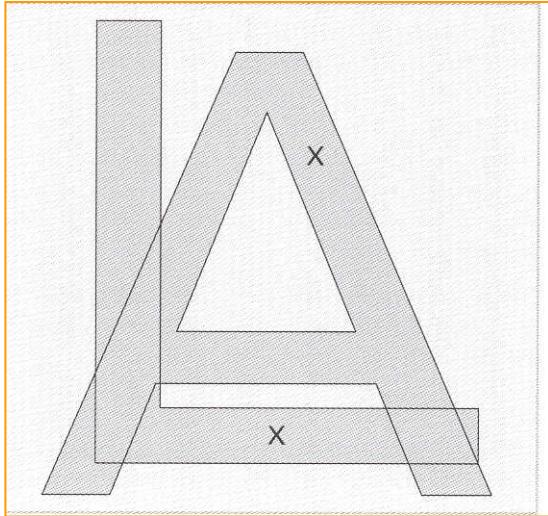
广义唯一性原理

传递分组、非传递分组和多级分组。

# 知觉组织的理论架构



# 分割中自顶向下的影响



Vecera & Farah, 1997

Is the “x” on the same or different object?

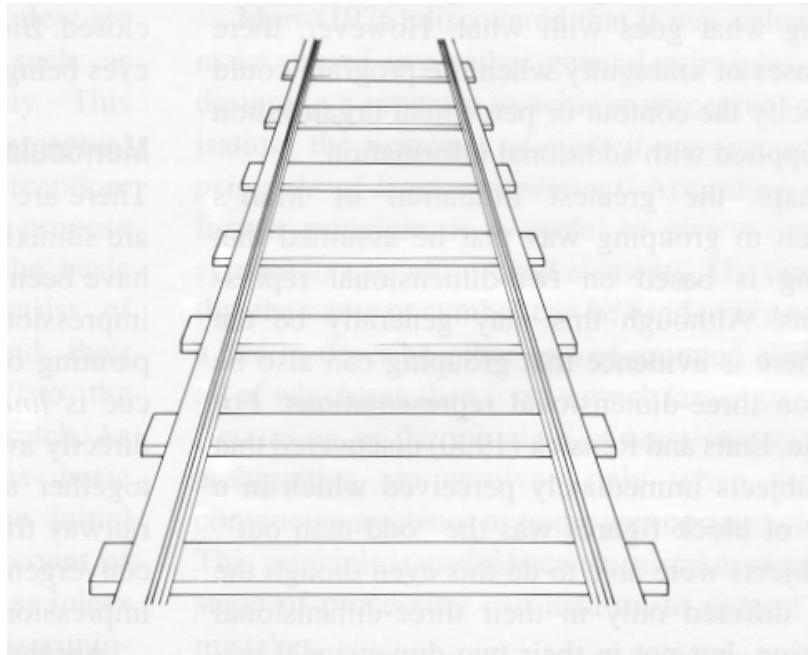
- Answering this question takes longer with the display on the right
- Suggests Gestalt processes are influenced by

# 深度知觉

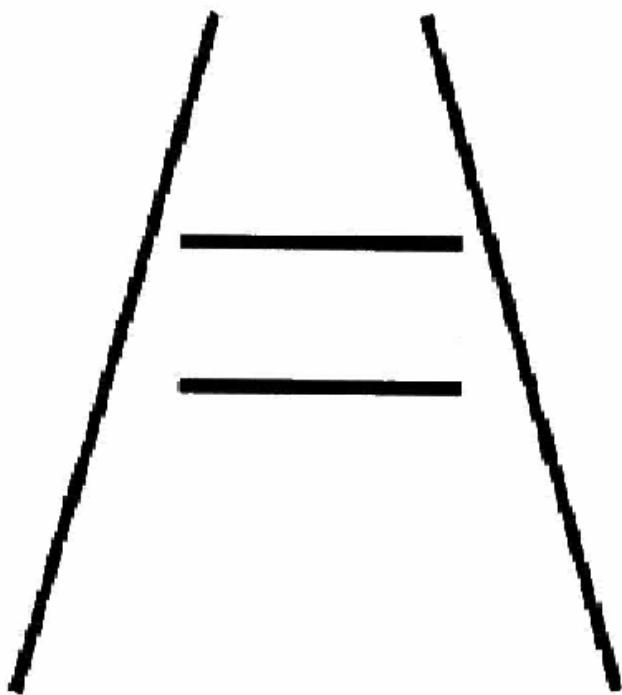
- 决定绝对和相对深度
- 单眼线索
  - 图形线索 (e. g., 线性远景)
  - 空中远景 (差别对比)
  - 纹理梯度
  - 插入中间
  - 底纹
  - 相对尺寸
  - 运动视差



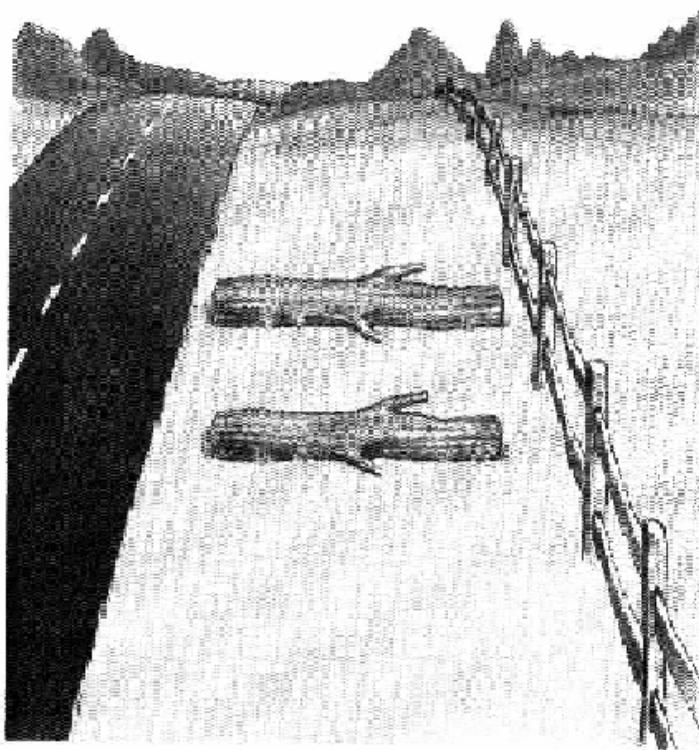
# 深度线索：收敛线



# 导致“Ponzo 错觉！”



(a)



(b)

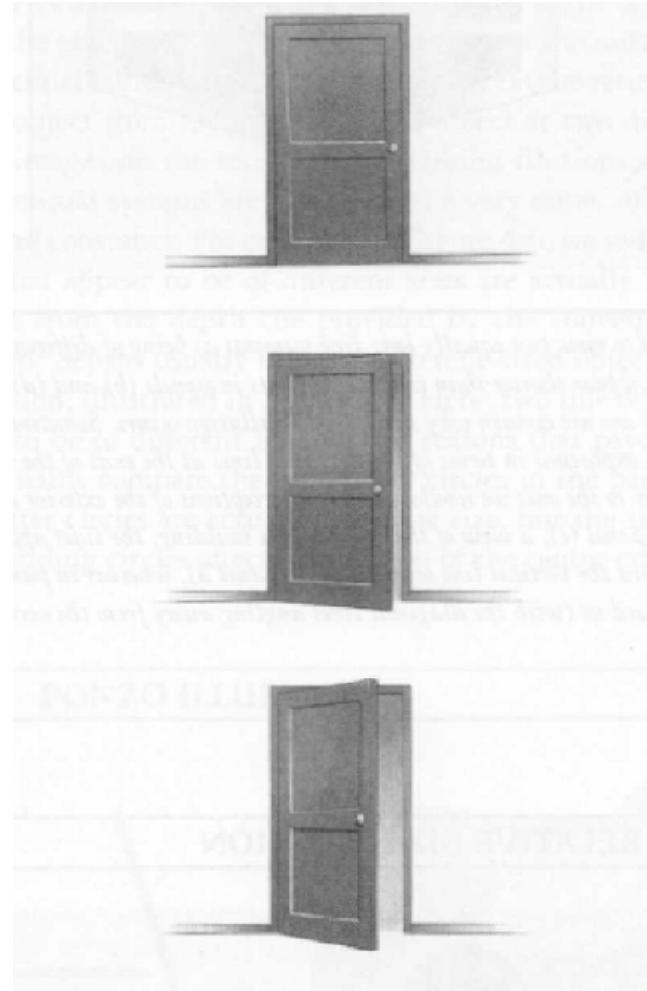
# 相对尺寸



- Depends on “shape constancy”

# 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？



# 插入中间

- Objects that are nearer should occlude objects that are farther away



# 浓淡远近法



- Relative clarity of farther versus nearer objects

# 纹理梯度



- Closer objects show more detail, articulation than those farther away

# 运动视差

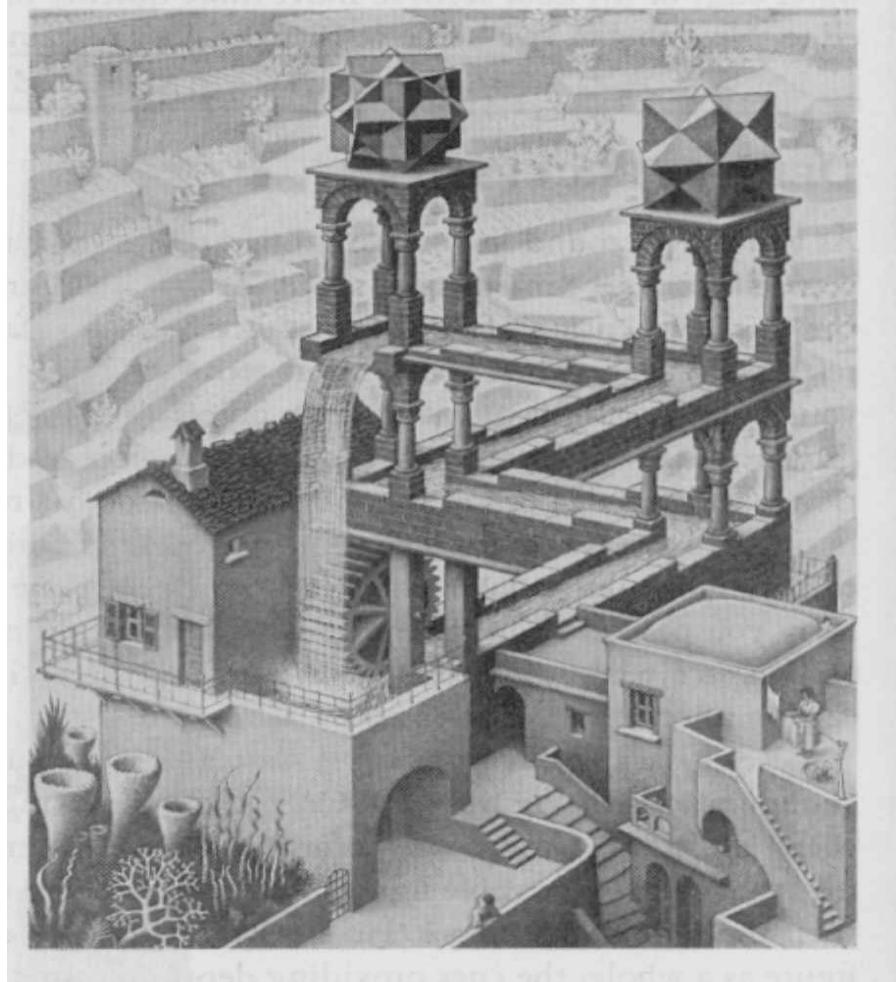


Relative motion

Direction of passenger's motion →

- Closer objects appear to move faster than those farther away

# 单目深度产生错觉



M.C. Escher:  
Waterfall

# 深度知觉



- **Binocular Cues to Depth**

- **Binocular disparity**: different images to each eye as a function of object closeness; basis of stereoscopic vision
- **Binocular convergence**: rotation of eyes depending on object closeness; brain receives efferent motor signals and interprets distance accordingly

# 深度线索组合



- 添加Additively?
- 优先Preferentially?
- 倍增Multiplicatively?

# 多重视觉区 (Zeki)

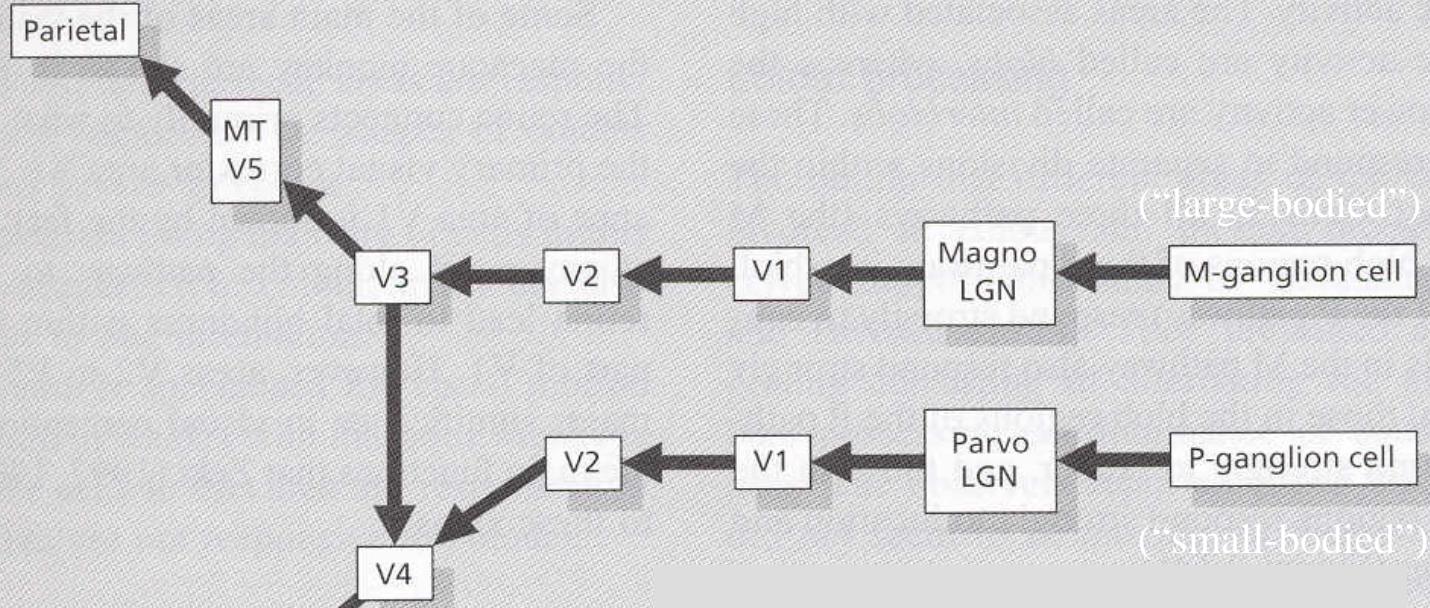


- V1 and V2: responsive to color and form
- V3 and V3A: form and motion
- V4: color
- V5: motion

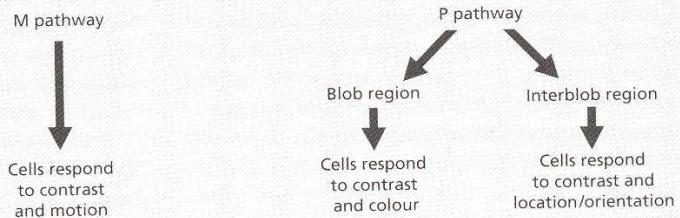
# 多重视觉区 (Zeki)

FIGURE 2.13

"Where" pathway



"What" pathway



# 关键问题

---

- 绑定问题**The “binding problem”**: if perception of given object is distributed throughout visual areas, how does the brain combine features to produce unitary percept?
- 意识与无意识知觉**Conscious vs. unconscious perception**: perception of object or object qualities may proceed pre-consciously

# 一般答案

---

- Perception = Memory:
  - Patterned co-occurrence in network binds features together probabilistically (e.g., learning in parallel networks)
- Perception before Memory:
  - “Unbound” percept matched with stored representations (perception)

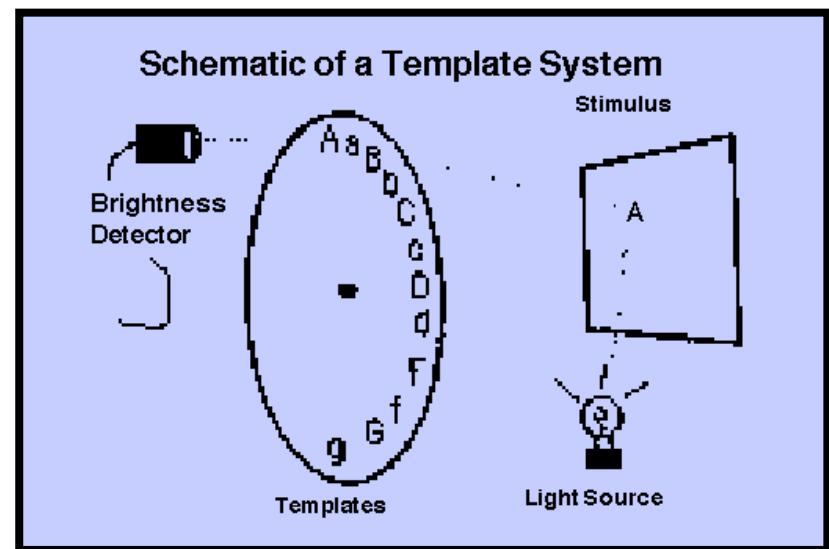
# 物体识别

---

- When does one object end and another begin?
- Viewpoint independence
- How do we know that dogs are dogs and not cats?

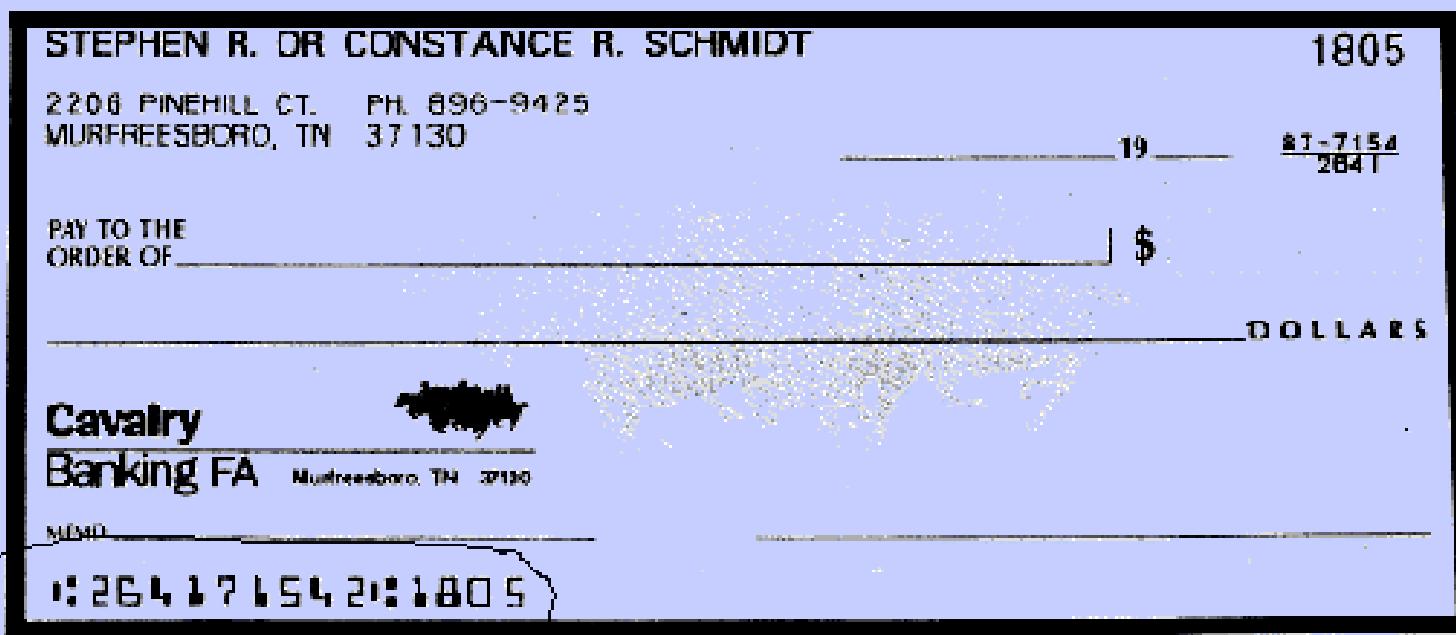
# 模式识别的模板理论

- **Template theories:** a miniature copy or template of each known pattern is stored in long-term memory
  - straightforward template theory
  - normalized template theory
- **Problems**
  - not adaptable
  - impose large storage requirements



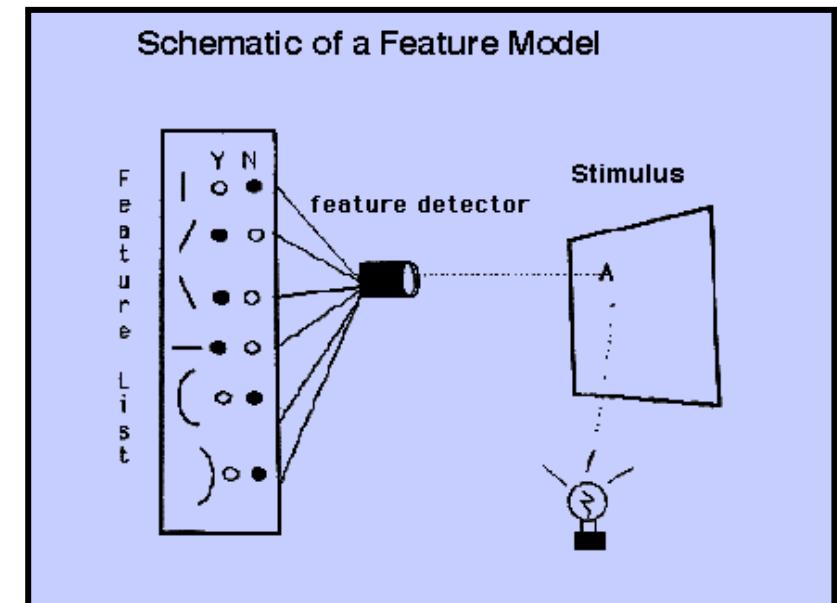
# 运用模板

## Template systems in action



# 模式识别的特征理论

- **Feature theories:** patterns consist of a set of specific features or attributes
- **Advantages:**
  - elementary features can combine to form multiple objects



# 模式识别的原型理论



- Individual instances are not stored; what is stored is an “exemplar” or representative element of a category
- Recognition based on “distance” between perceived item and prototype
- Nature of computation still relatively unknown

# 匹配

---

A A

Object Template  
MATCH!

A A

Object Template  
NO MATCH!

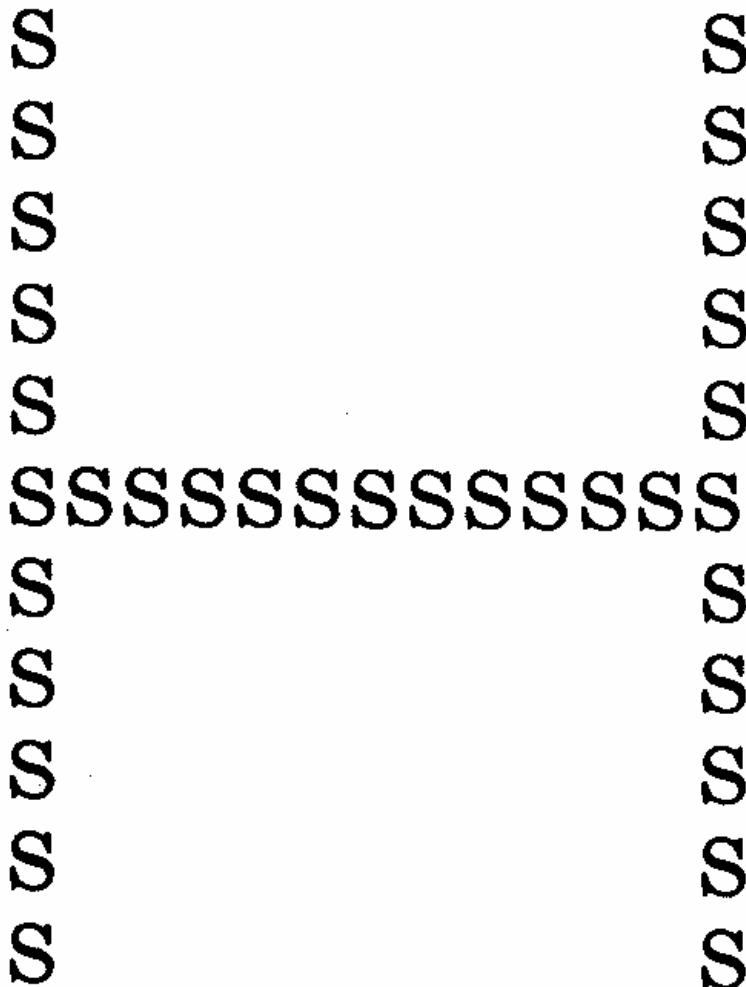
A A

Object Prototype  
MATCH!

A A

Object Prototype  
MATCH!

# 全局处理过程

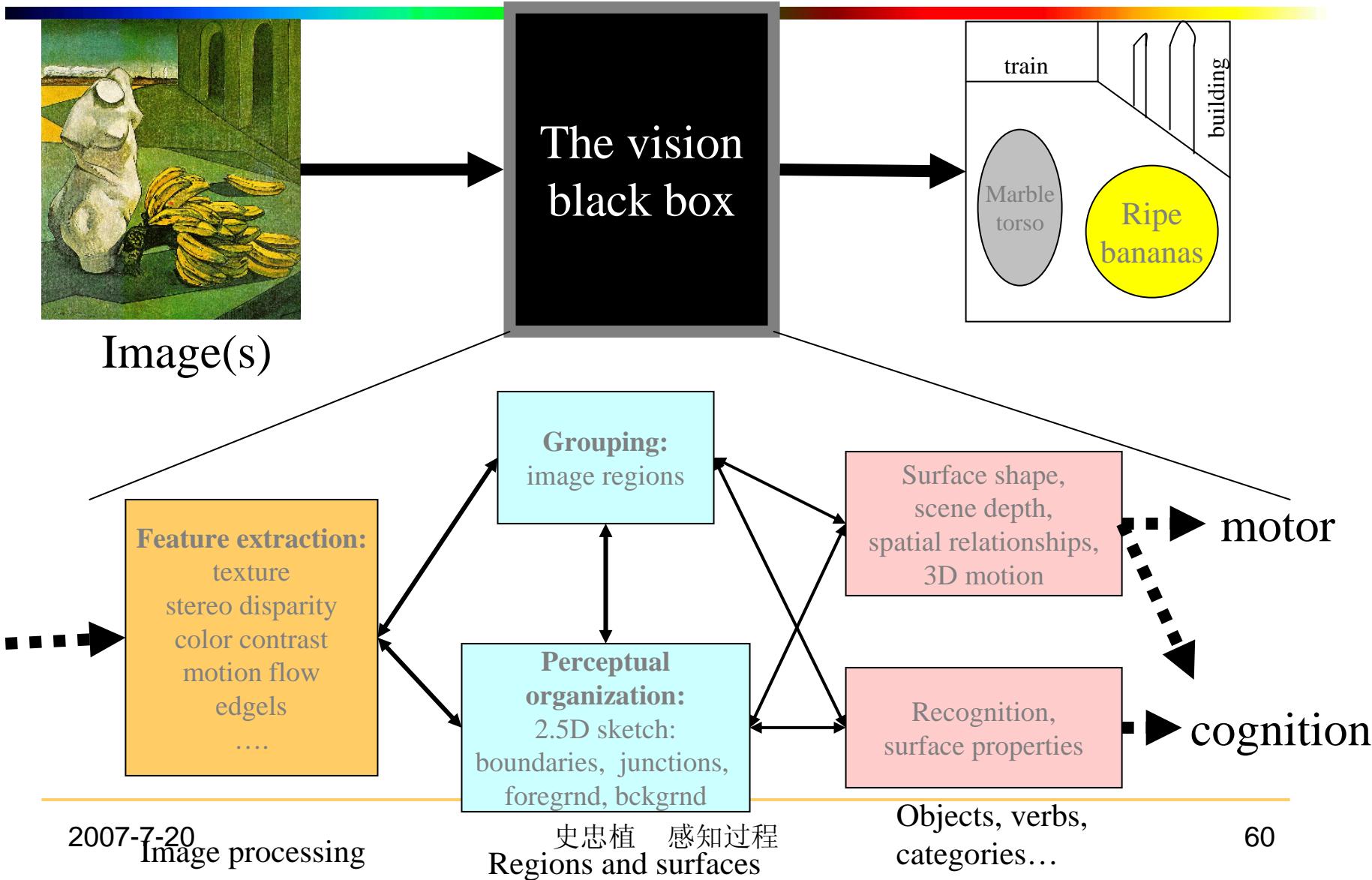


Global Processing Precedence?

Navon (1977): Is the large letter an H or an S? Are the small letters H's or S's?

Decision speed for small letters greatly affected by whether large letter was the same or different; decision speed for large letters unaffected by small letters

# 视觉处理



# 马尔关于视觉的计算理论

---

视觉是个信息处理任务

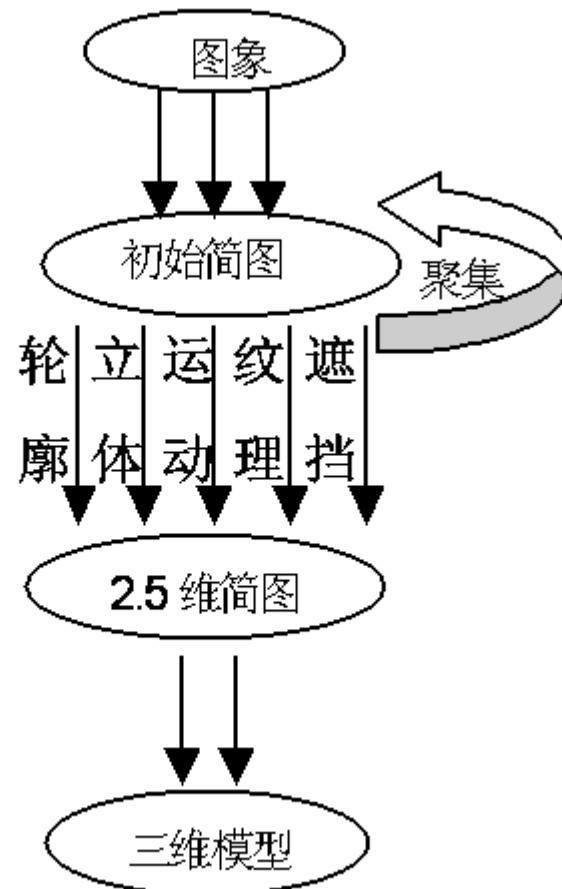
- 视觉是一种处理过程, 经过这个过程从图象中发现外部世界中有什么东西和它们在什么地方
- 信息处理有三个研究层次
  - 第一个层次是信息处理的计论理论
    - 研究的是对什么信息进行计算和为什么要进行这些计算
  - 第二个层次是算法
    - 如何进行所要求的计算, 也就是要设计特定的算法;
  - 第三个层次是实现算法的机制或硬件
    - 研究完成某一特定算法的计算机构。

# 视觉系统的表象 (representation)

## ● 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 $3 \times 10^1 + 7 \times 10^0$ 变成37
- 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

---

Grey level description



Primal sketch.



2.5 Dimensional sketch.



3 Dimensional model  
sketch

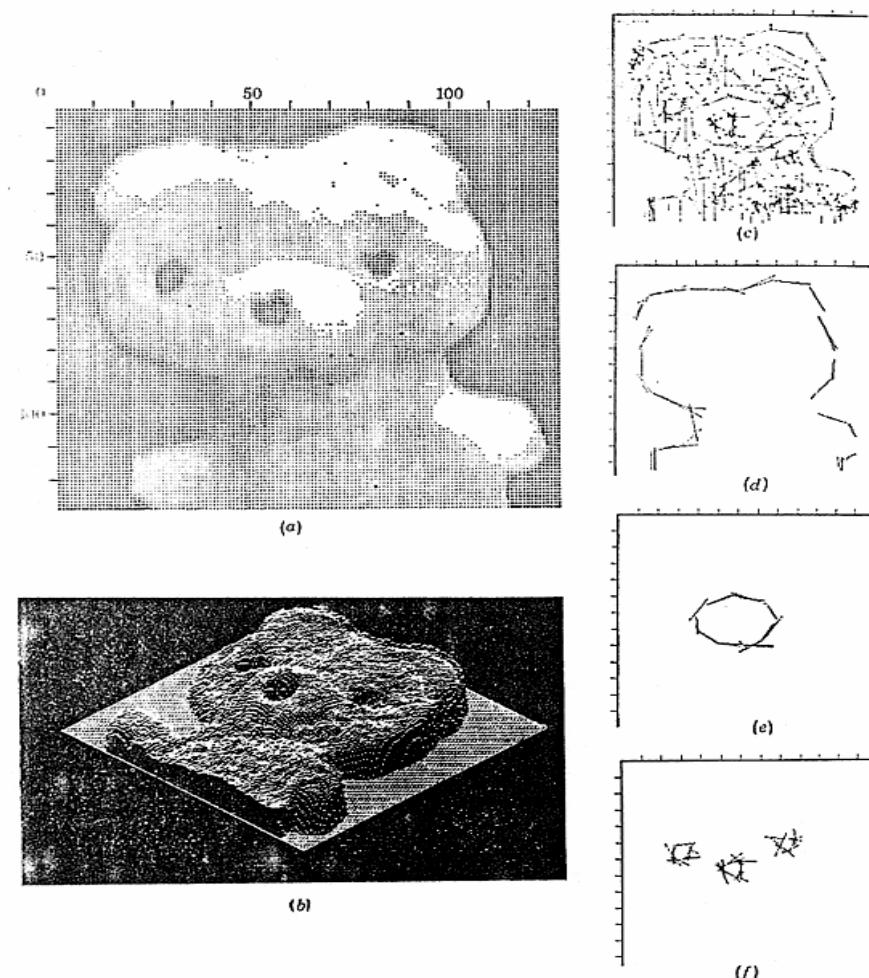


# 初始简图 (primal sketch)

- 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。初始简图所包含的大部分信息集中在与实际的边缘以及边缘的终止点有关的急剧的灰度变化上。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。

# 2. 5维简图

信息源

立体视觉

方向选择性

从运动恢复结构

光 源

遮挡轮廓

其它遮挡线索

表面方向轮廓

表面纹理

表面轮廓

影 调

信息类型

视差，因而可得到  $\delta_\gamma, \Delta_\gamma$  和  $S$

$\Delta\gamma$

$\gamma, \delta\gamma, \Delta\gamma$  和  $S$

$\gamma$  和  $S$

$\Delta\gamma$

$\Delta\gamma$

$\Delta s$

可能有  $\gamma$

$\Delta\gamma$  和  $\gamma$

$\delta s$  和  $\Delta s$

## 2. 5维简图

$\gamma$   
在 $s$ 中:  $\gamma$

$\Delta\gamma$ —相对深度（按垂直投影），就是观察者到表面点的距离

$s$ — 的连续或小的变化

$\delta s$ —  $s$ 的不连续点

$\Delta s$ — 局部表面方向

— 的连续或小的变化

— 的不连续点

# 三维模型

---

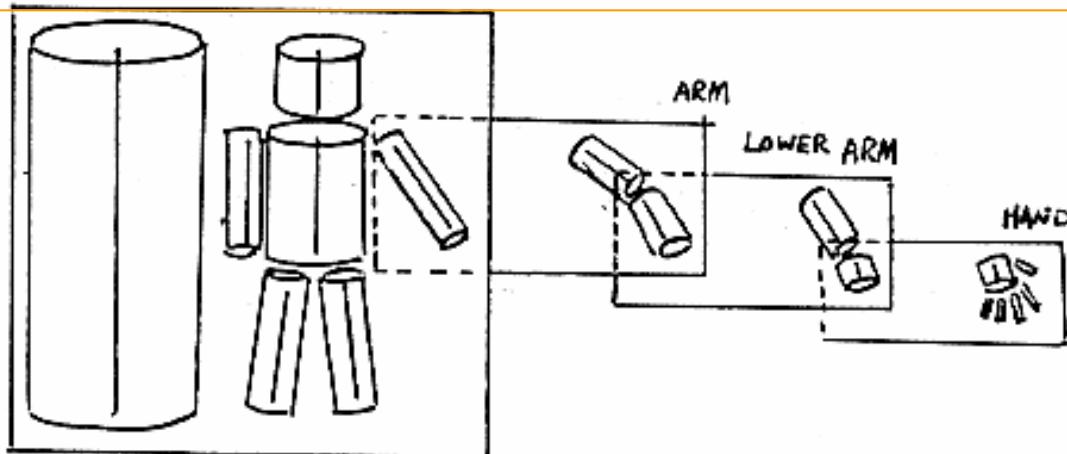
- 在三维模型表象中，以一个形状的标准轴线为基础的分解最容易得到。在这些轴线中，每条轴线都和一个粗略的空间关系相联系；这种关系对包含在该空间关系范围内的主要的形状组元轴线提供了一种自然的组合方式。用这种方法定义的模块我们称之为三维模型。

---

由三维模型分级组合构成的人体表象系统

# 三维模型

- 2.5维简图采用以观察者为中心的坐标系
- 适合于识别的形状表象应该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~~<sup>有一个</sup>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

# 三维模型

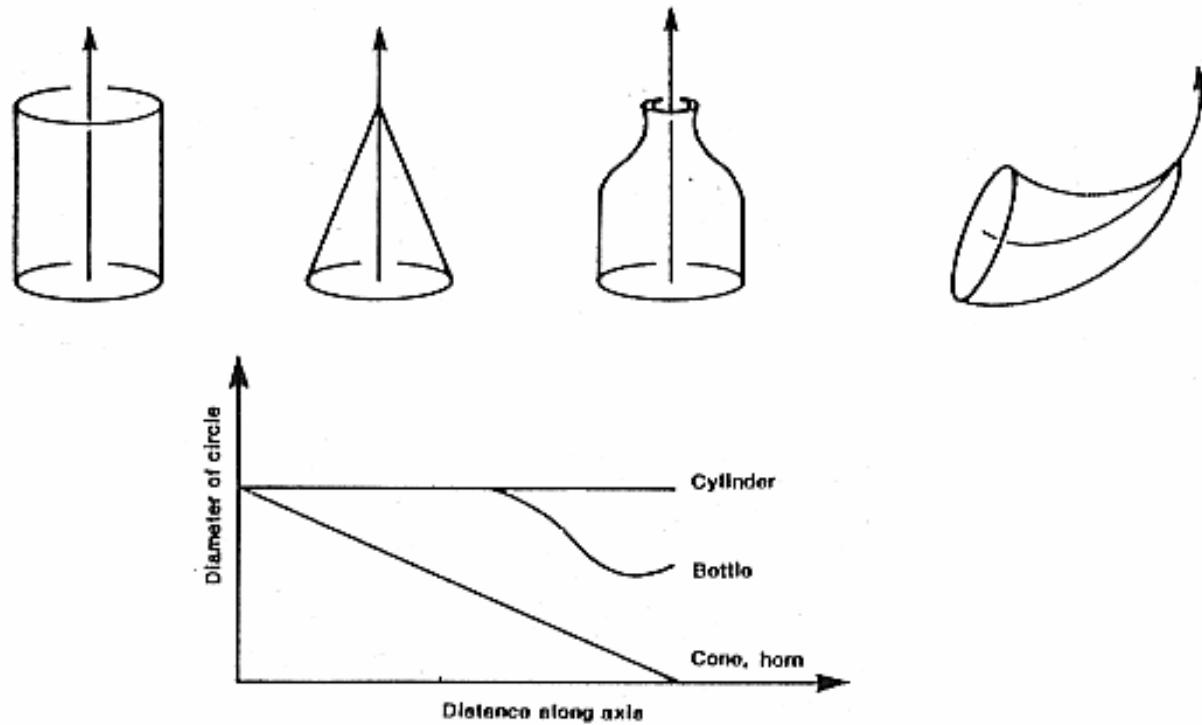


图2.19 截面大小不固定及轴线不是直线时的广义锥

# 关于Marr的视觉计算理论的讨论

- 关于Marr的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

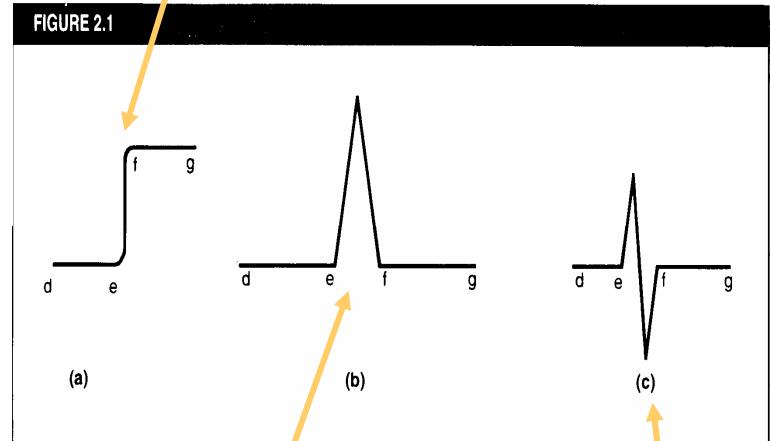
- 面临的困难

- 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

# Marr-Hildreth 算法

- Attempts to account for development of primal sketch
- Idea of “blurred representations”
- **“Zero-crossings”** identify edges within a visual image
- Only **‘reliable’** zero-crossings are kept
- **Four types of tokens:** edge-segments, bars, terminations, and blobs

Intensity change



Rate of change (1st D)

Zero-crossing (2nd D)

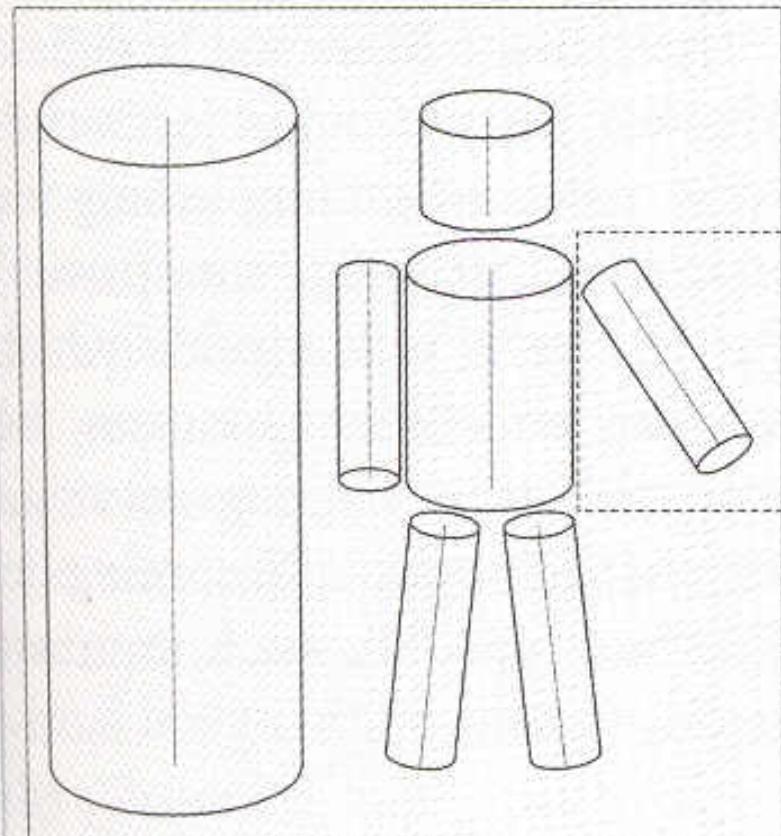
# Marr & Nishihara (1978)

---

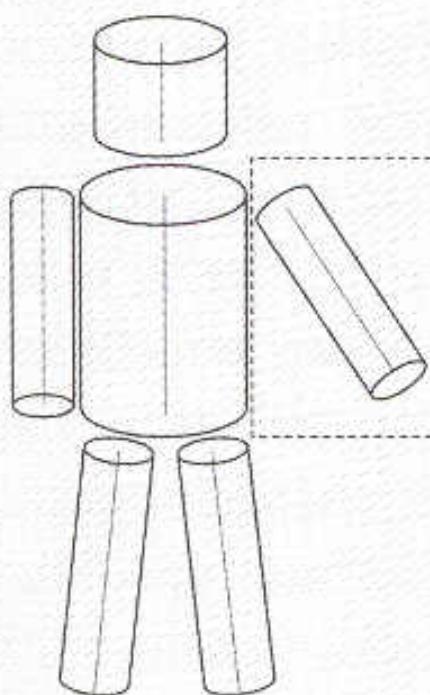
- Development of 3-D sketch based on processing of more elementary shape primitives (basic primitive is a cylinder with a major axis)
- Hierarchical organization of primitives
- Concavities important in segmenting parts

**FIGURE 4.4**

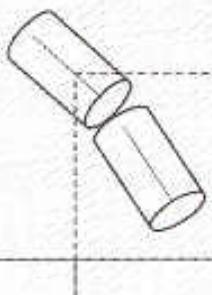
Human



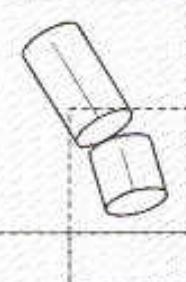
(b)



Arm



Forearm

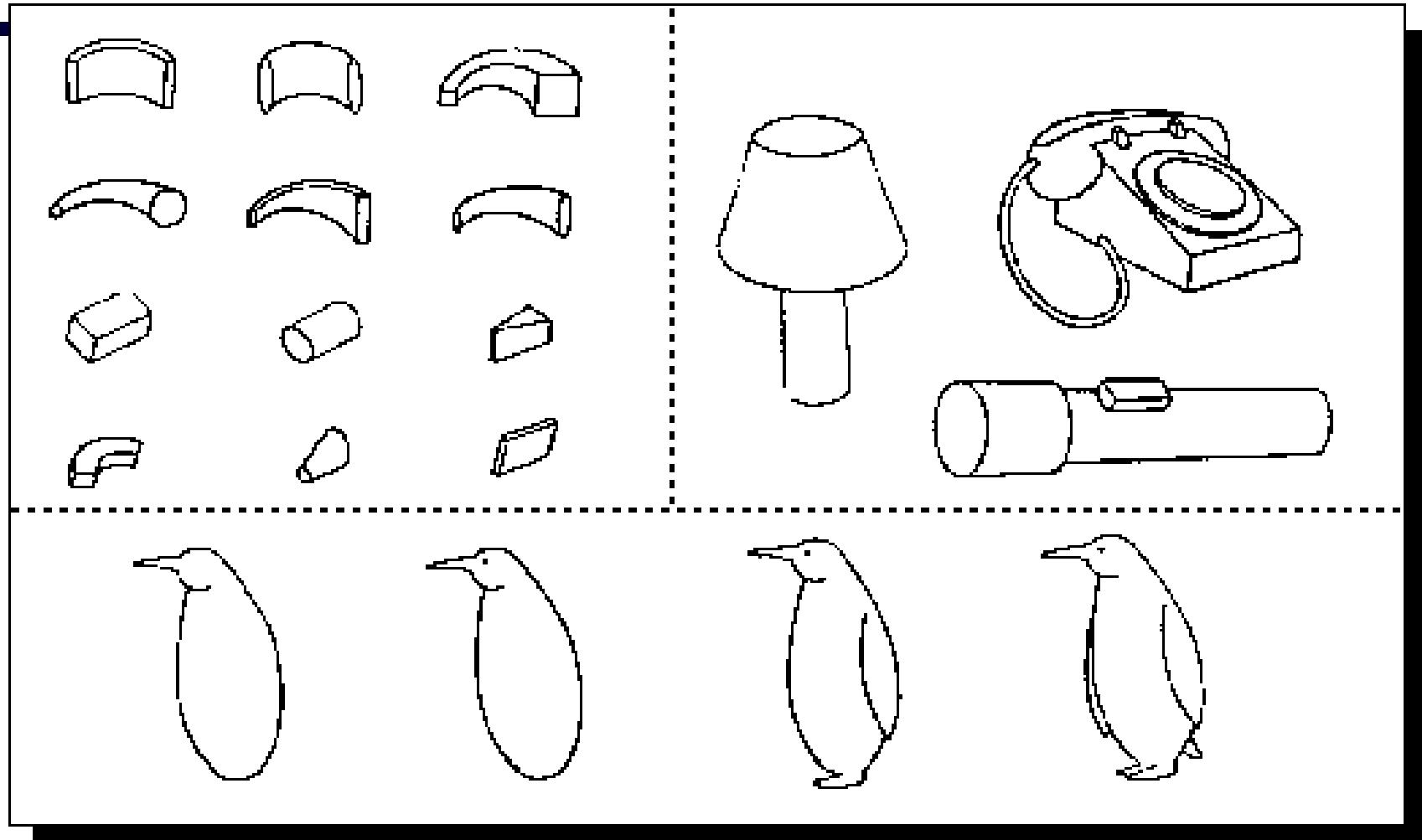


Hand



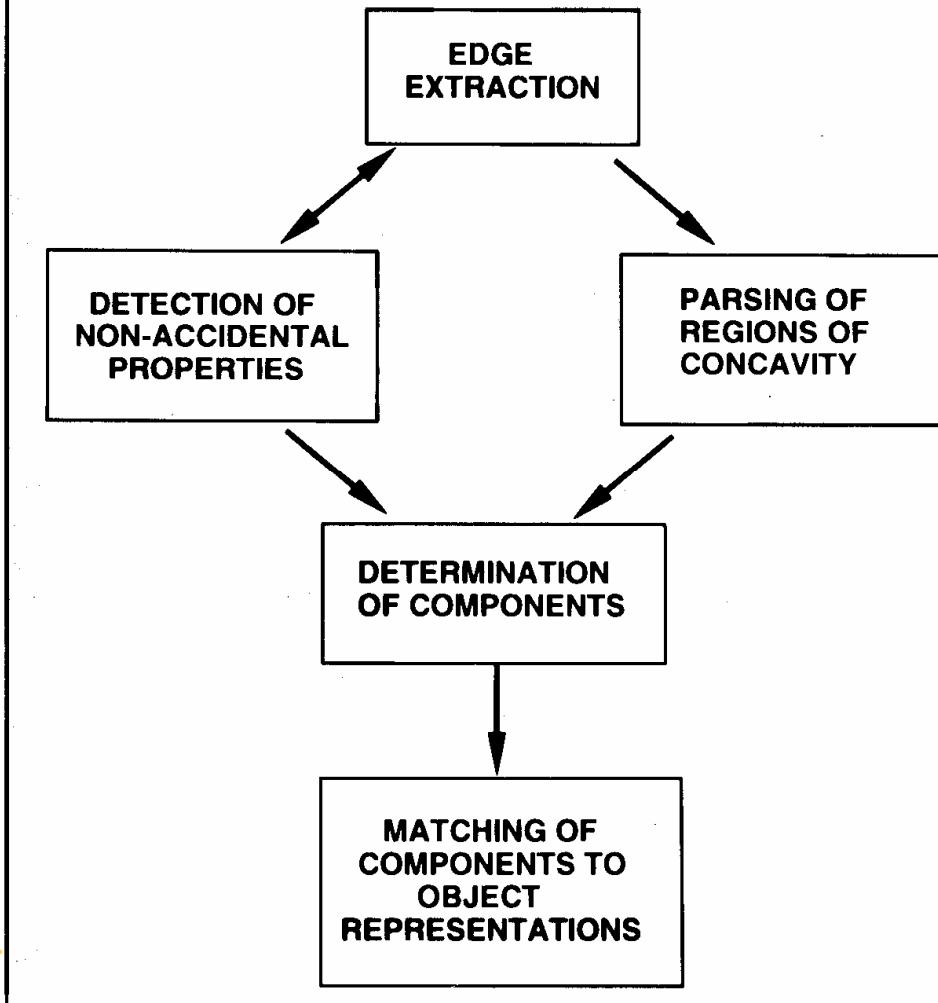
# Biederman's RBC Theory

- Objects built from primitives called “geons” ( $n = 36$ )
- Must decide appropriate way a visual object is to be segmented into geons
- Two key components of decision:
  - locating concavity
  - deciding which edge information remains invariant across different viewing angles (invariant properties like curvature, parallelism, etc.)



# Overview of Biederman's Model

FIGURE 3.7

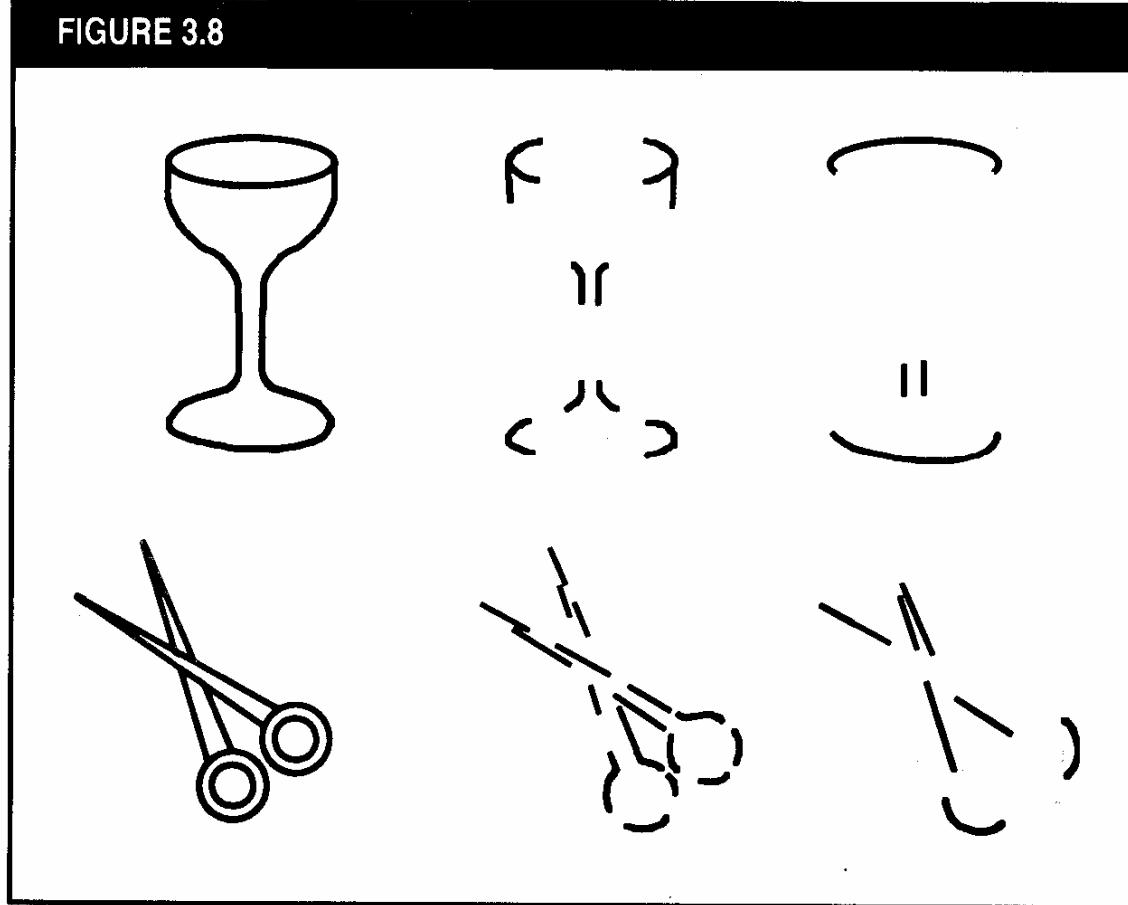


A

B

C

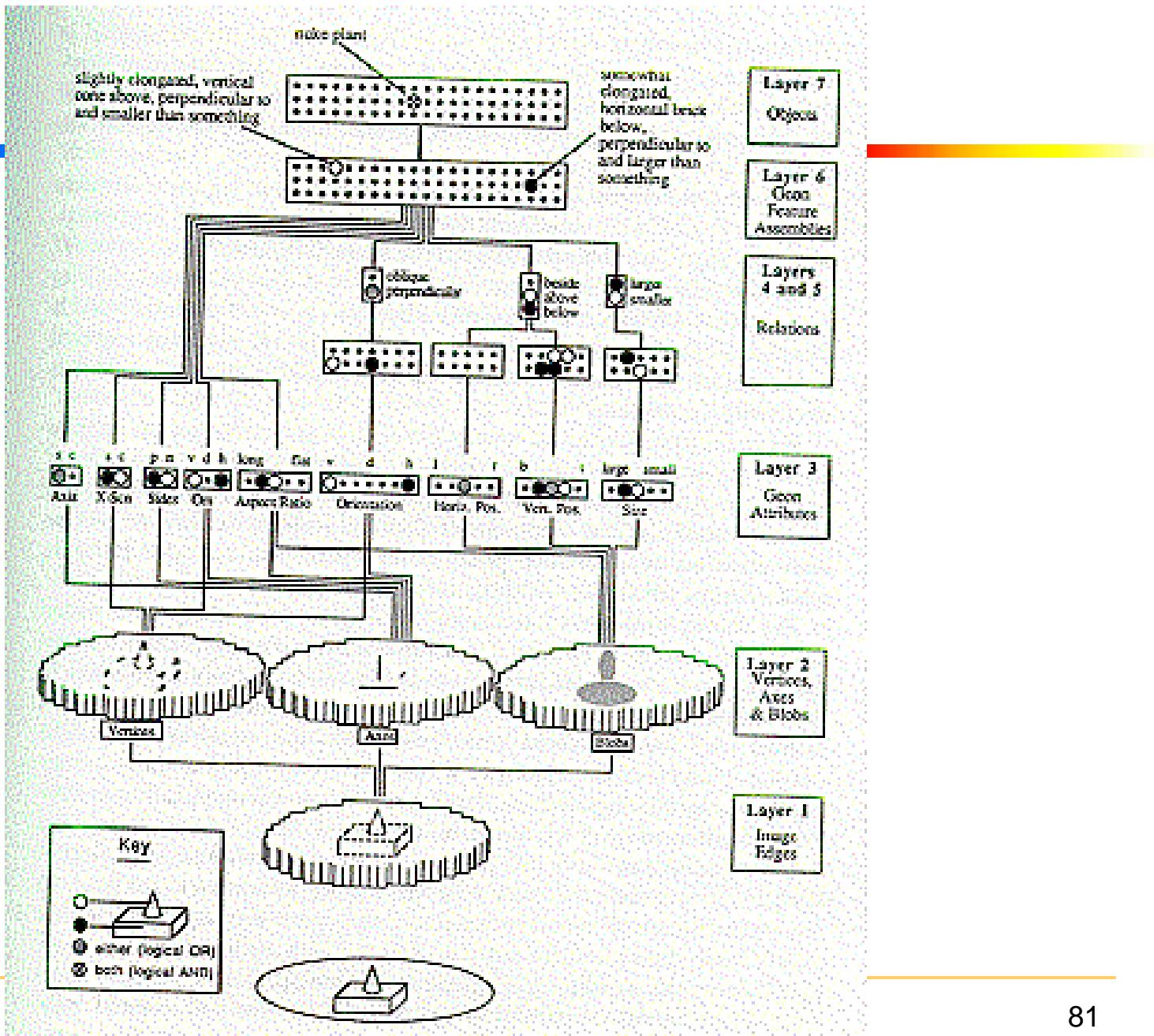
FIGURE 3.8



A: complete stimuli

B: stimuli preserving concavity information

C: stimuli lacking concavity information



# Key Similarities

---

- Edge coding
- Grouping into higher order features
- Matching to stored structural knowledge
- Access to semantic knowledge
- Naming

# 物体识别中的公共元素

Common Elements in Object Recognition Theories



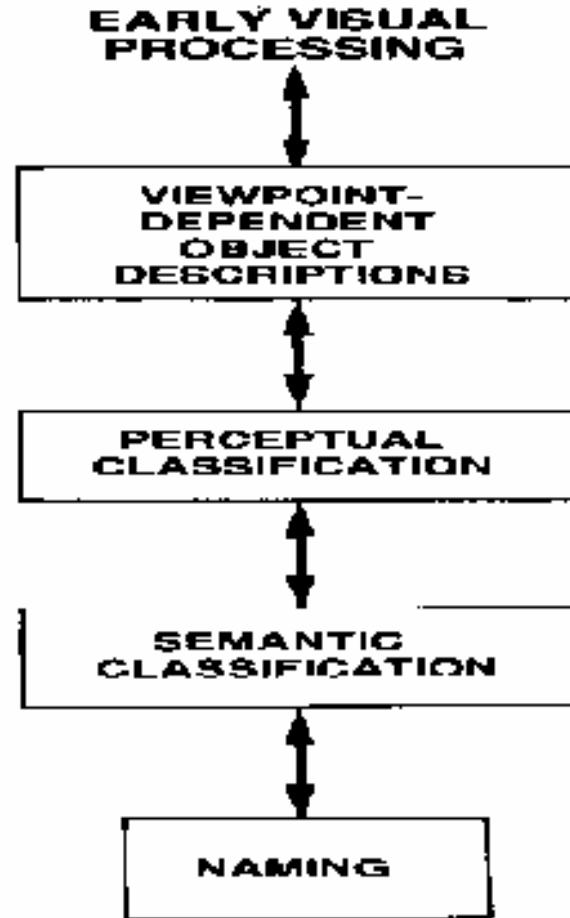
- Edge coding
- Grouping or encoding into higher-order features
- Matching to a stored “structural representation”
- Access to semantic knowledge

# 结构描述 Structural Descriptions

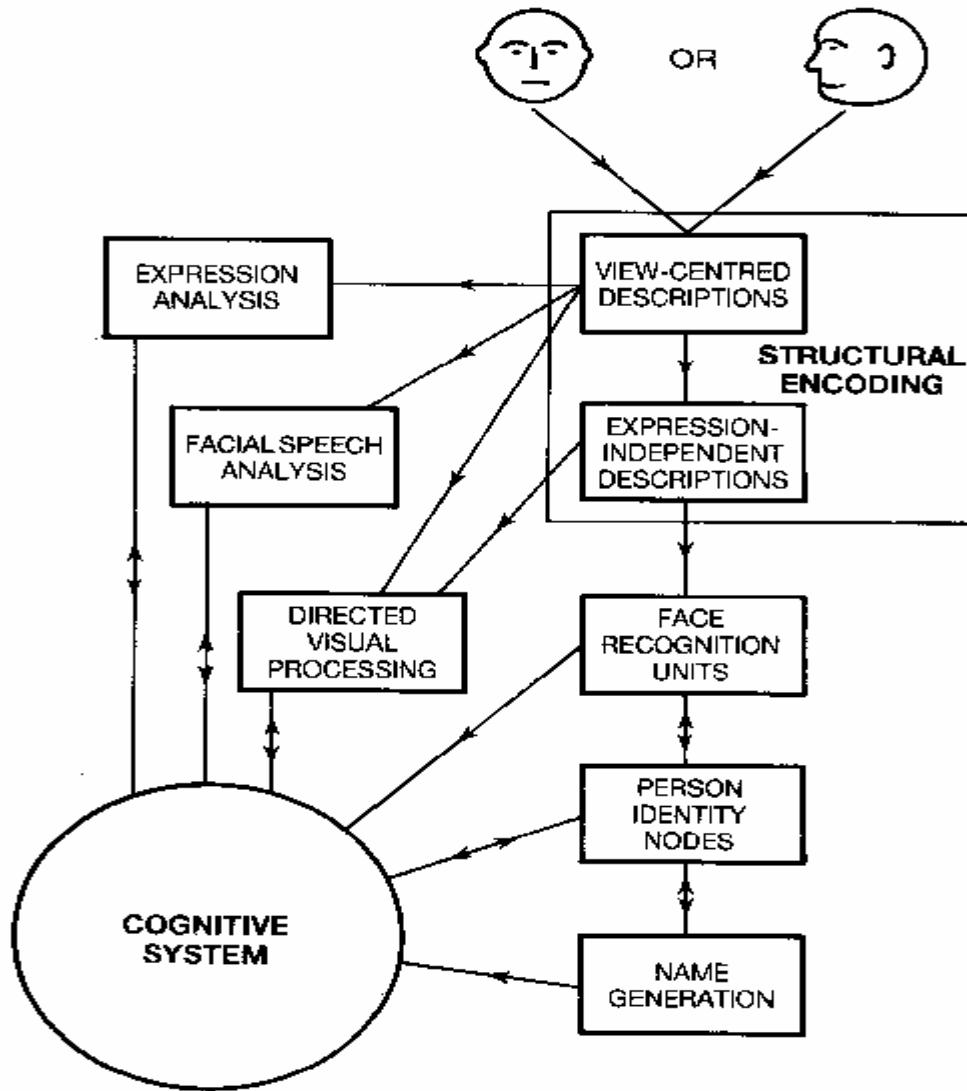


- **Structural descriptions** consist of propositions which “describe the nature of the components of a configuration and make explicit the spatial arrangement of these parts” (Bruce & Green, 1990)
- **Perceptual representation systems** for faces, objects, etc. Evidence for separate systems: category-specific recognition defects

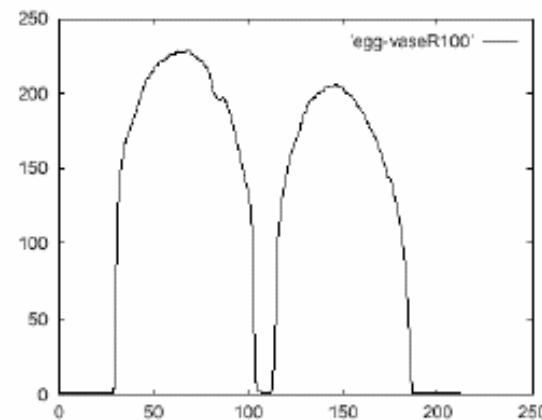
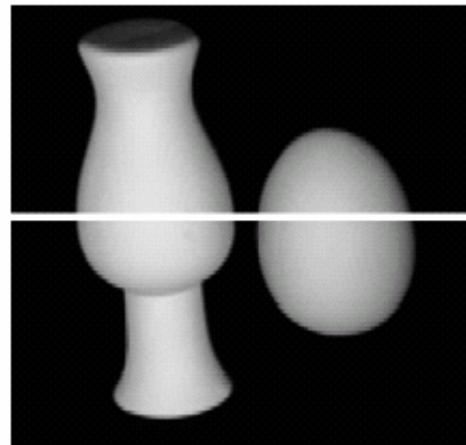
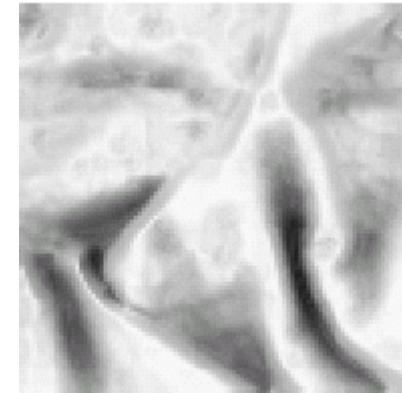
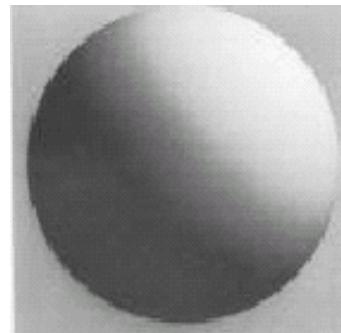
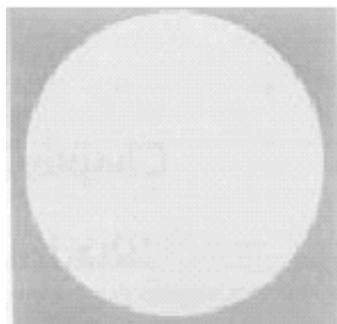
# 视觉处理



# 视觉处理



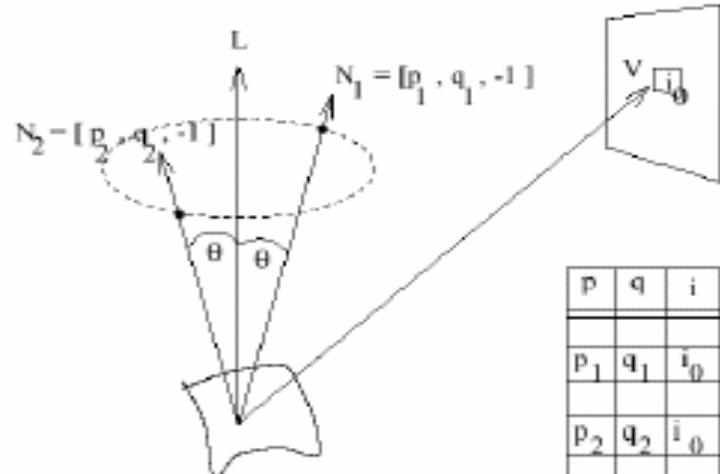
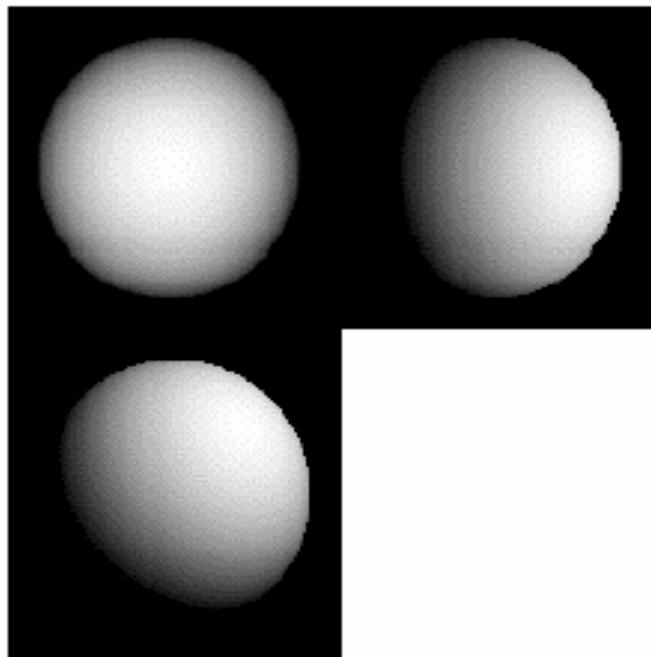
# 影调提供形状的信息



# 底纹

## Shading

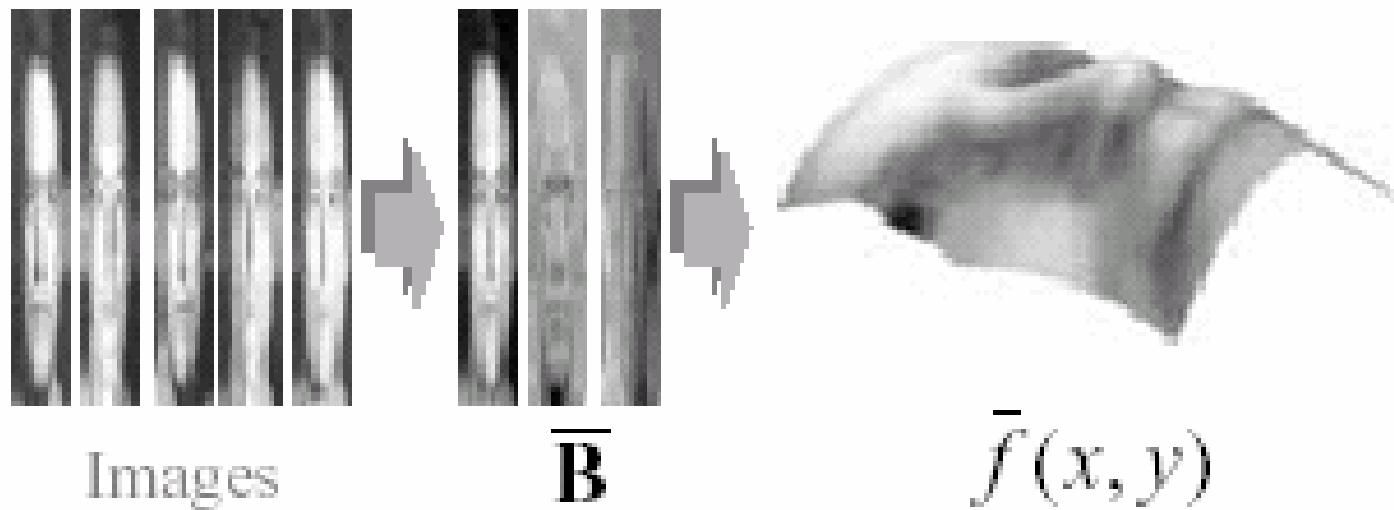
- Appearance varies under different illumination.



# 由不同光照下的图象重构3D形状

## GBR and Reconstruction

**Proposition:** From as few as three images of a Lambertian surface, taken under different lighting conditions, we can reconstruct the surface up to a Generalized Bas-Relief transformation – without knowledge of the light sources. (See also Fan & Wolff 97.)

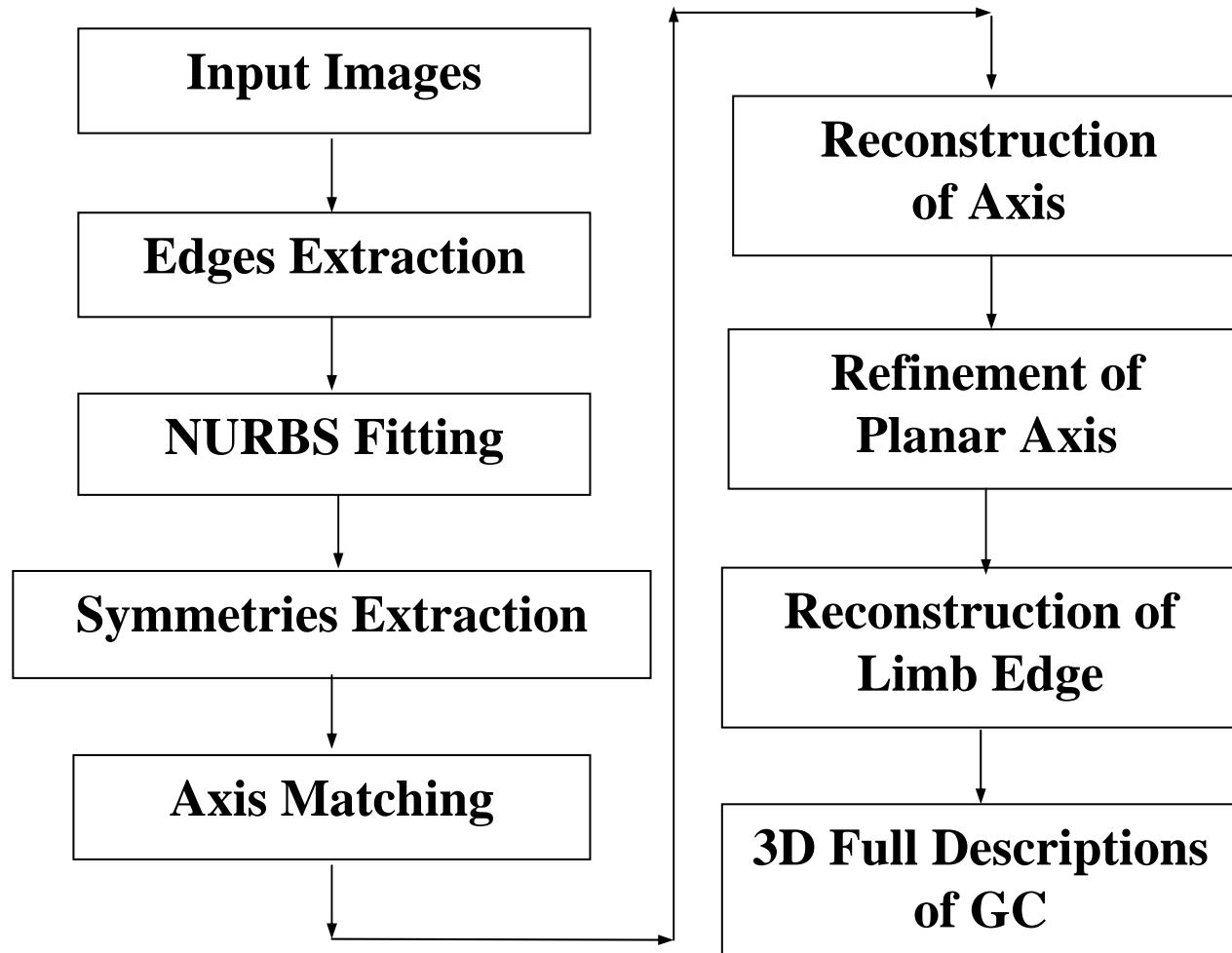


# 三维重建

- ✓ 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



# 系统框架



# 参数化跟踪

- ✓ 人体轮廓的参数化跟踪：在局部被遮挡的情况下，在序列图像中有效地跟踪人体的轮廓



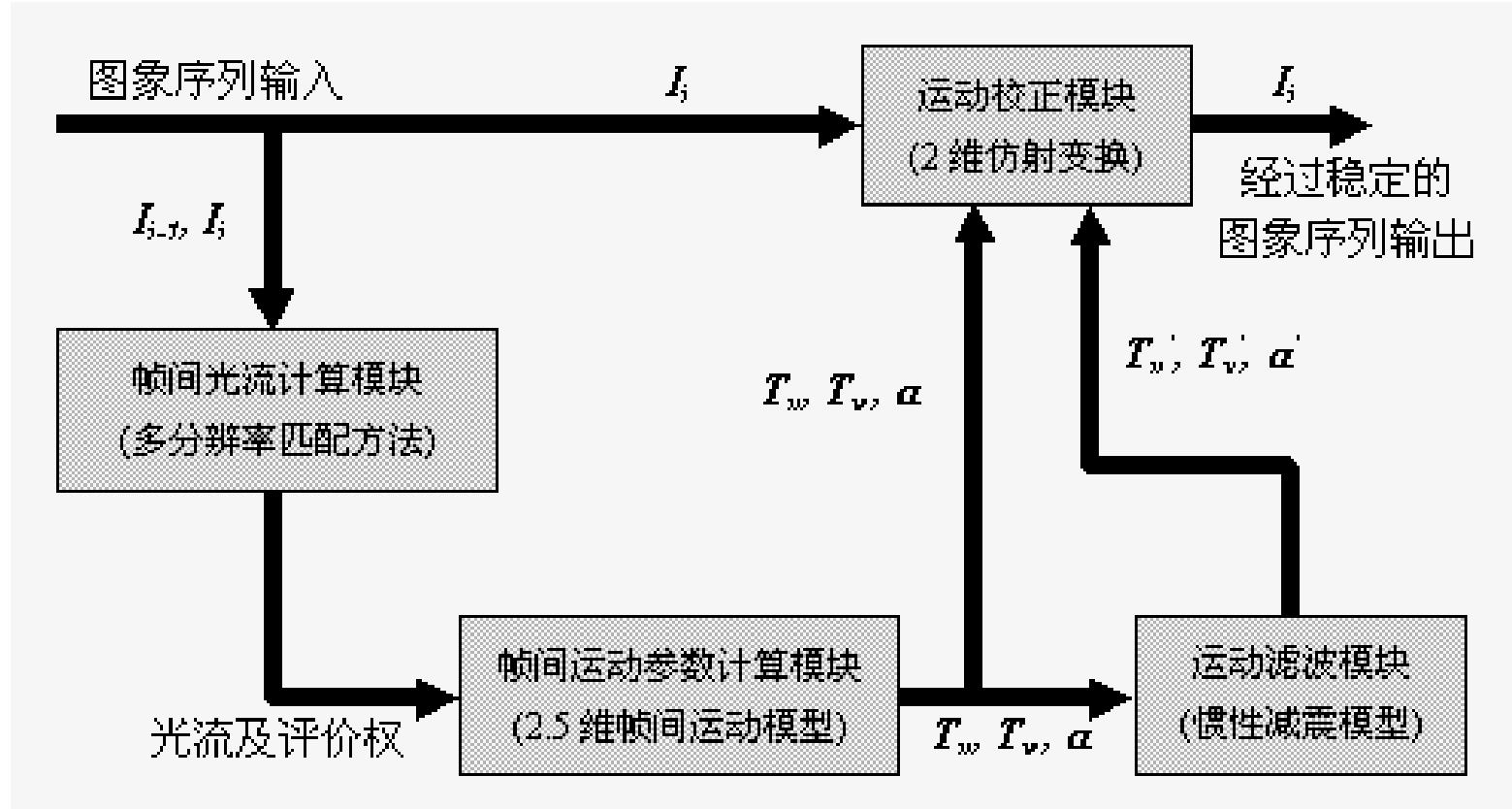
2007-7-20 原始序列

史忠植 感知过程



被跟踪结果

# 图象序列稳定系统总体结构

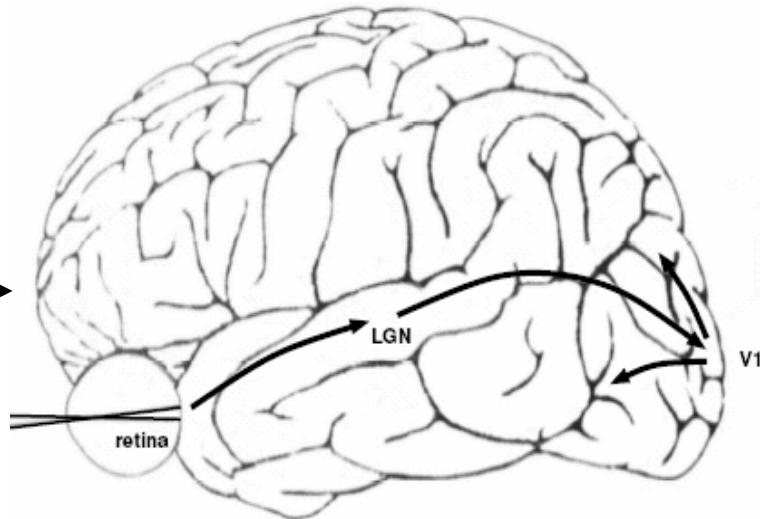


# 有效编码

研究动机



怎么联系？



# 稀疏编码发展历程

- 视觉感知的目标就是产生一个外部输入信号的有效表示 (Attneave 1954)；
- 有效编码假设：初级视皮层神经细胞的主要功能就是去除输入刺激的统计相关性 (Barlow 1961)；
- 稀疏编码理论：初级视皮层神经细胞对自然图像刺激进行稀疏响应 (Olhausen & Field, Nature, 1996)。

# 有效编码



## ❖ 机理测试

从生物机理上，在自然图像刺激条件下检测神经细胞的响应特性

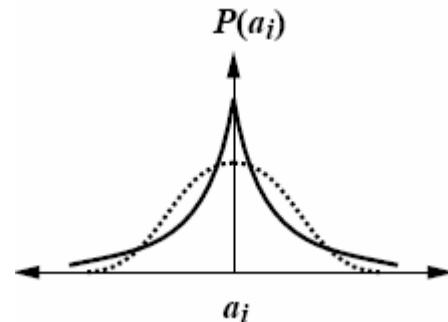
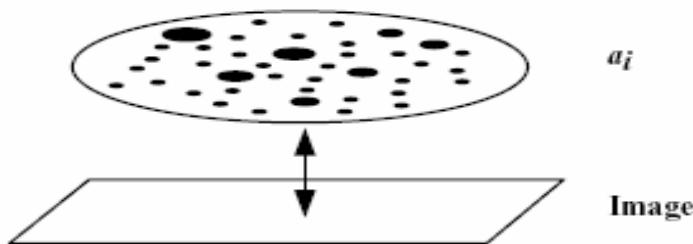
- Vinje 2000年在science上类似成果表明：短尾猿视皮层（V1区）用稀疏编码有效表示自然场景
- Nirenberg 等2001年在nature上发表的研究结果表明，在冗余性测度和自然刺激条件下，一组视网膜神经节对外界刺激独立编码。

## ❖ 模型仿真

利用自然图像的统计特性，建立模型模拟早期视觉系统的神经细胞编码机制

# 稀疏编码模型

**Sparse representation**



在作用于同一刺激的神经元群中仅有少数的神经元被激活；

某一神经元响应值的分布具有稀疏特性：即大部分值都分布在0附近，而激活的概率很少；

# 稀疏编码模型（续）

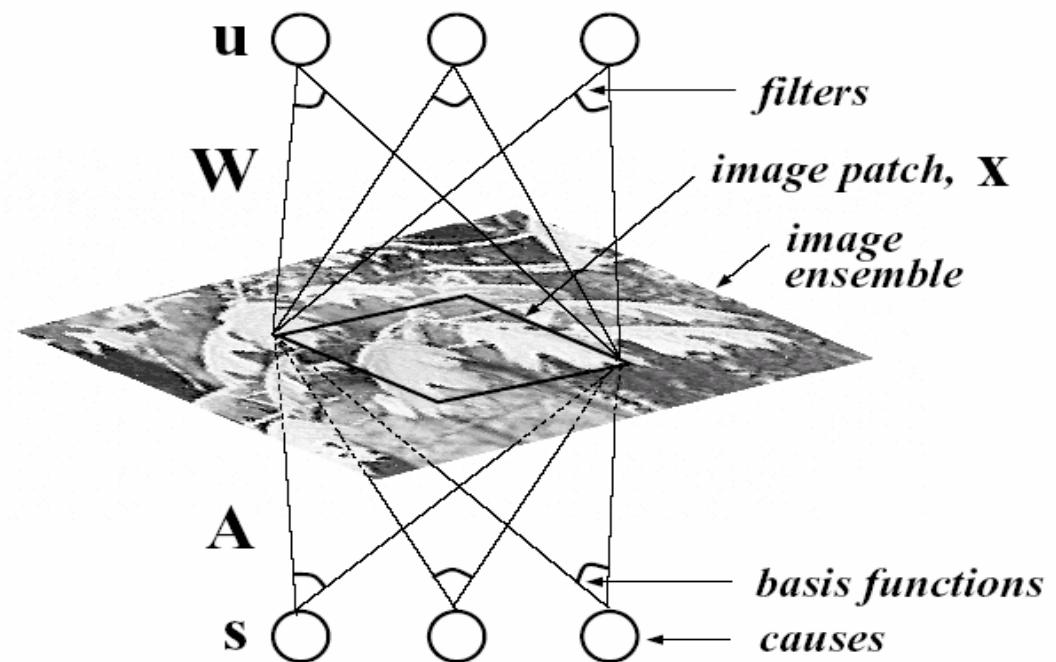
- 神经生物学角度：测试神经元细胞在自然图像刺激下响应的统计特征
  - ✓ *Sparse coding and decorrelation in primary visual cortex during natural vision. Science, 287:1273–1276, Feb 2000.*
  - ✓ *Retinal ganglion cells act largely as independent encoders. Nature, 411:698{701, June 2001.*
- 计算模型角度：利用自然图像的统计特性，构造初级视皮层神经元的信息处理模型
  - ✓ 基于稀疏性准则的模型  
*Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381:607–609, 1996.*
  - ✓ 基于独立性条件的模型  
*The 'independent components' of natural scenes are edge filters. Vision Research, 37 23):3327–3338, 1997.*

# 有效编码研究机制 – 模型仿真

基于外部环境刺激的统计特性，根据特定的优化准则建立响应模型，模拟初级视皮层中神经细胞的响应特性（神经生理学）。

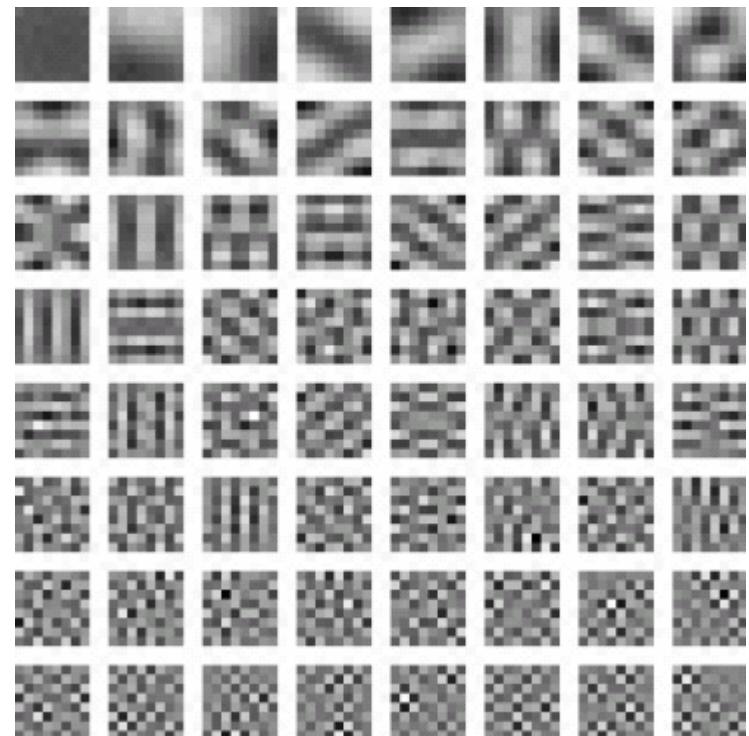
$$X = AS$$

$$S = WX$$



# 基于二阶统计的线性有效编码模型

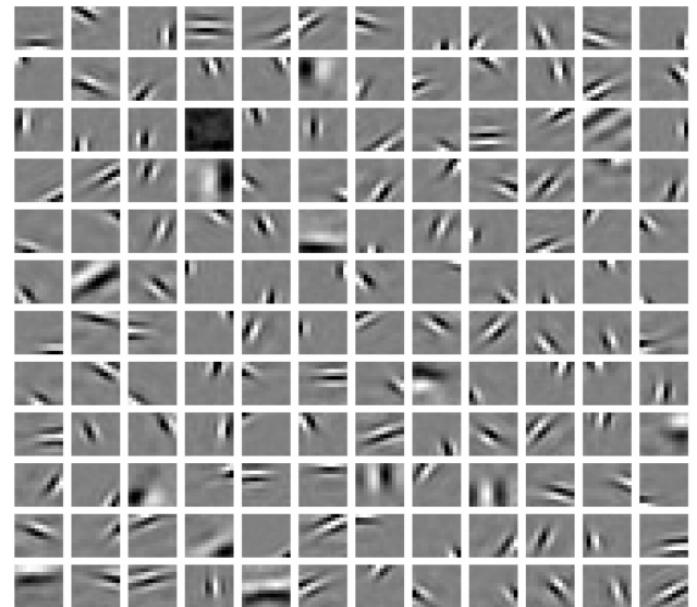
$Cov(S) = Diagonal\ Matrix$



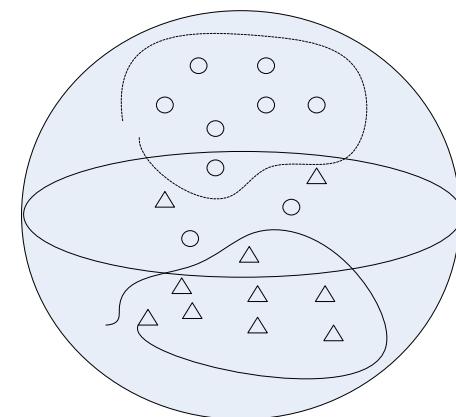
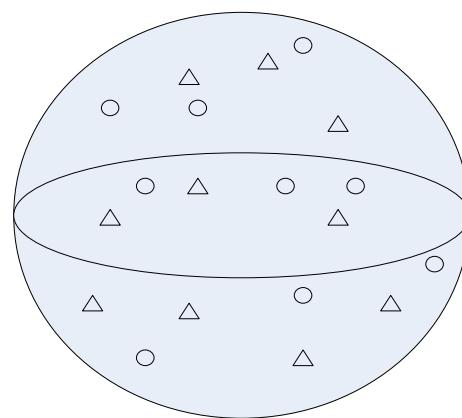
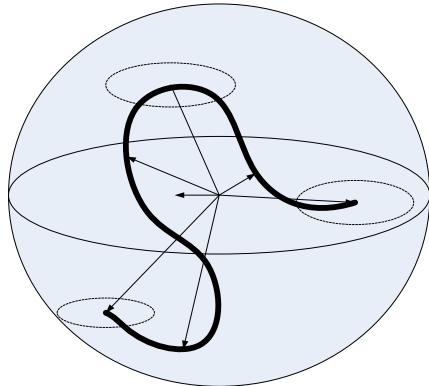
# 基于高阶统计的线性有效编码模型

$$P(S) = P(S_1) * P(S_2) * \dots * P(S_n)$$

$$E(a, \Phi) = \sum_{x,y} \left[ I(x,y) - \sum a_i \Phi_i(x,y) \right] + \lambda_s \sum_i S\left(\frac{a_i}{\sigma_i}\right)$$



# 区分空间编码



a

N-dimentional state space  
of image pixel values

b

# 面向分类任务的有效编码模型——基本 模型



- 线性叠加模型

$$I(x, y) = As = \sum_{i=1}^n s_i * a_i(x, y)$$

- 稀疏编码模型

$$E(s, A) = \sum_{x, y} \left[ I(x, y) - \sum s_i a_i(x, y) \right]^2 + \lambda_s \sum_i S\left(\frac{s_i}{\sigma_i}\right)$$

# 面向分类任务的有效编码模型—定义

- 训练数据和编码系数

$$X_1 = \{I_1^1, I_2^1, \dots, I_{N1}^1\} \quad X_2 = \{I_1^2, I_2^2, \dots, I_{N2}^2\} \quad I_i^j = [s_1, s_2, \dots, s_n]$$

- 距离函数

$$D(I_1, I_2) = \sqrt{\sum_{i=1}^n (s_i^1 - s_i^2)^2}$$

- 类间距离

$$D_w = D(I_i^j, \tilde{m}_j)$$

- 类内距离

$$D_B = D(I_i^j, \tilde{m}_{\hat{j}})$$

# 面向分类任务的有效编码模型—定义

- 类别距离

$$DR = \frac{D_B}{D_W} \quad Dis(s) = \ln(DR^2) = \ln\left(\sum_{i=1}^n (s_i^j - \tilde{m}_i^j)^2\right) - \ln\left(\sum_{i=1}^n (s_i^j - \hat{m}_i^j)^2\right)$$

- 目标函数

$$E(s, A) = Error(s, A) + \lambda_s Sparseness(s, A) + \lambda_d Dis(s)$$

# 面向分类任务的有效编码模型—学习过程

- 优化目标  
 $\text{Min}(E)$
- 两层优化过程:

内层

$$\frac{dE}{ds_i} = -2b_i + \frac{\lambda_s}{\sigma} S'(\frac{s_i}{\sigma}) + \lambda_d \left( \frac{2 * (s_i^j - \tilde{m}_i^j)}{\sum_{i=1}^n (s_i^j - \tilde{m}_i^j)^2} - \frac{2 * (s_i^j - \hat{m}_i^j)}{\sum_{i=1}^n (s_i^j - \hat{m}_i^j)^2} \right)$$

外层

$$\Delta a_i(x, y) = \eta \left\langle s_i \left[ I(x, y) - \sum_j s_j a_j(x, y) \right] \right\rangle$$

# 注意机制

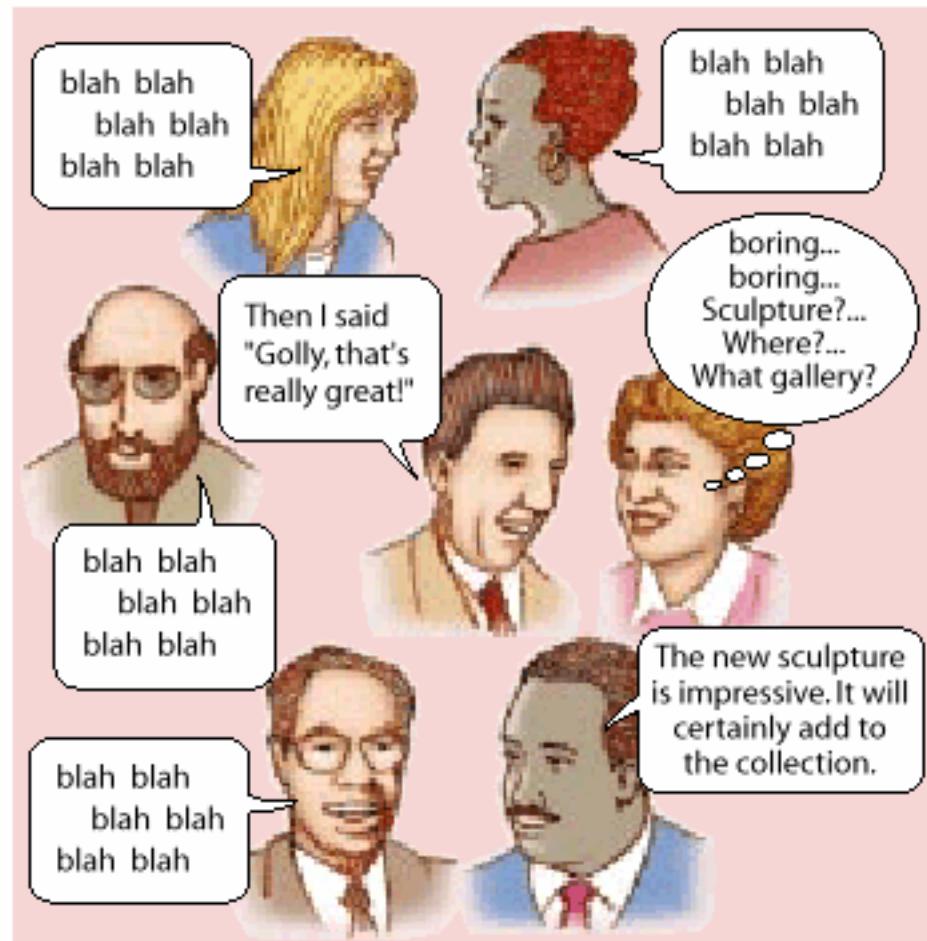


- 注意是心理活动或意识在某一时刻所处状态，表现为对一定对象的指向与集中。在大多数时候人们可以有意识地控制自己的注意方向。注意有两个明显的特点：指向性和集中性。注意的指向性是指人在每一瞬间的心理活动或意识选择了某个对象，而忽略了其余对象。

# 注意机制

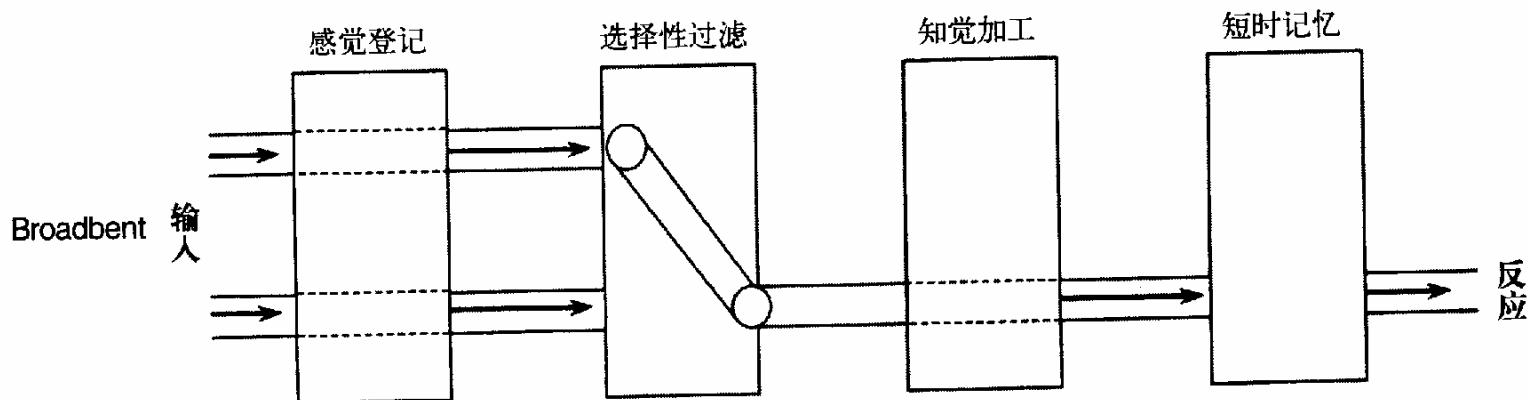
- 视觉注意机制是大脑神经系统信息处理过程中的一种重要策略，在大脑信息处理过程中无处不在。
  - 视觉刺激信息量巨大；
  - 视觉信息并不是同等重要；
  - 视觉神经系统计算能力的有限性。
- 稀疏编码模型必须引入注意机制。
  - 我们研究发现，在稀疏编码模型中被激活的神经元比例很大（大于70%）；
  - 一个视觉刺激中包含着不同的模式，但并不是所有的模式都必须得到表示。

# 鸡尾酒会



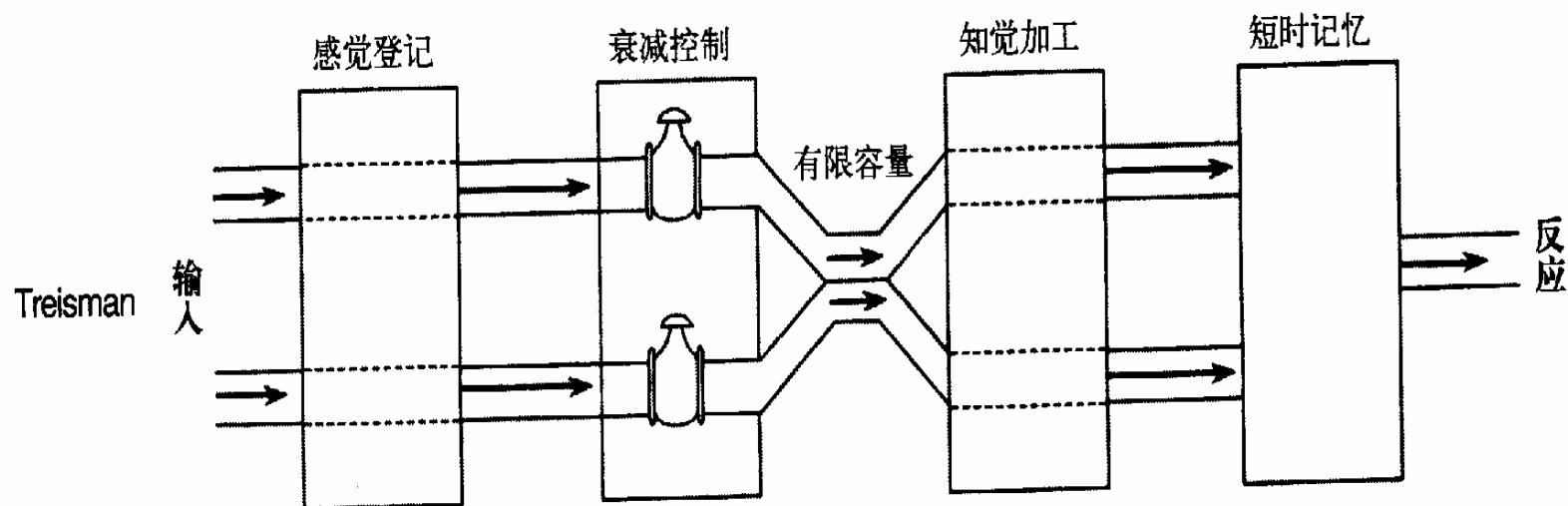
# 过滤器模型

该模型最早由英国著名心理学家布罗德贝特（D. Broadbent）于1958年提出。是关于注意的一个较早的理论模型。



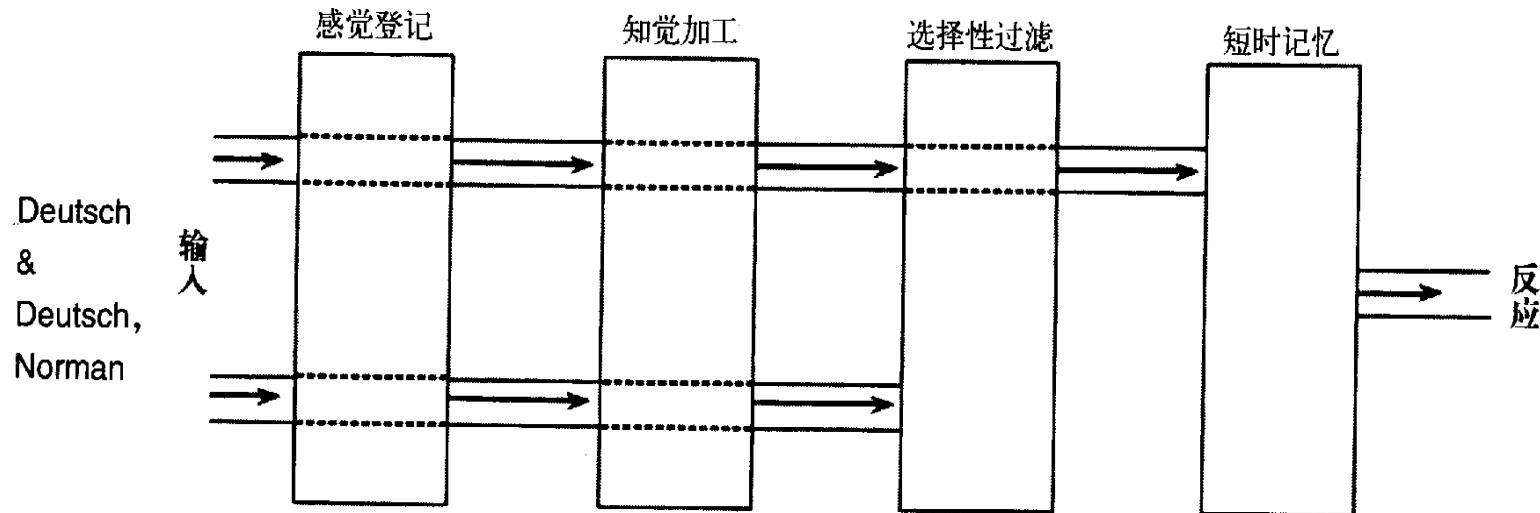
# 衰减模型

衰减模型是美国心理学家特瑞斯曼(A. Treisman)于1960年在修正过滤模型的基础上提出来的。



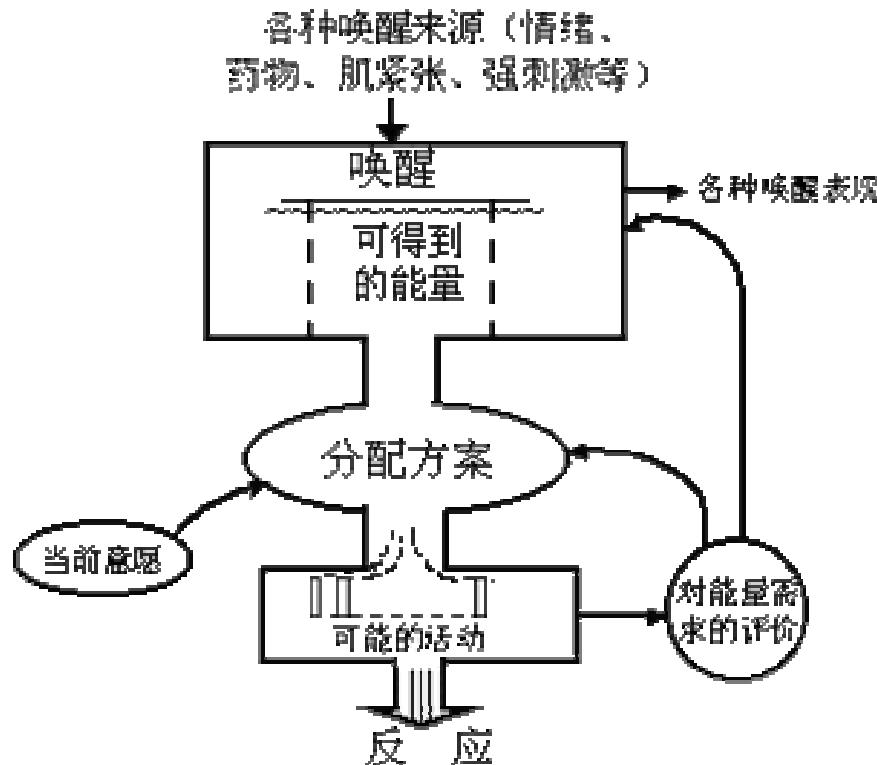
# 反应选择模型

1963年德意志 (J.Deutsch) 提出反应选择模型[112]。该模型认为，注意并不在于选择知觉刺激，而在于选择对刺激的反应。

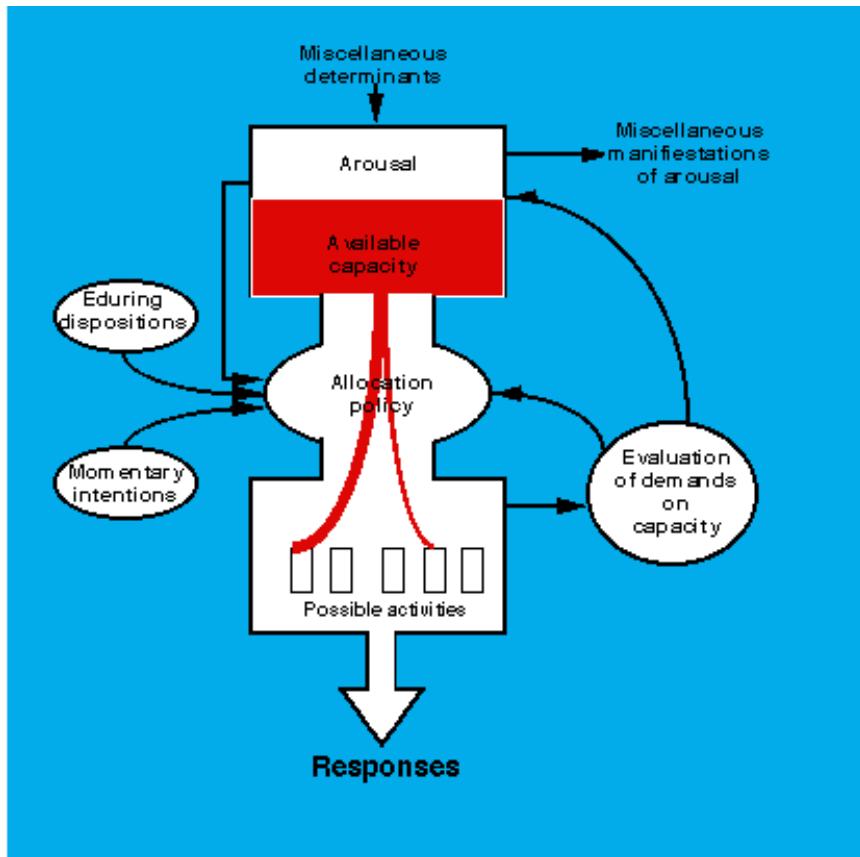


# 资源配置模型

1973年卡尼曼(D. Kahneman)提出资源配置模型。资源配置模型认为，注意是人能用于执行任务的数量有限的资源。



# Resource Allocation Model (Kahneman)



- What Affects Allocation?
  - Resources
    - Arousal
    - Available Capacity
  - Other Effects
    - Enduring Dispositions
    - Momentary Intentions

# 探照灯理论

- “注意的探照灯理论”（**Spotlight Metaphor**）

Crick认为注意就像探照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。

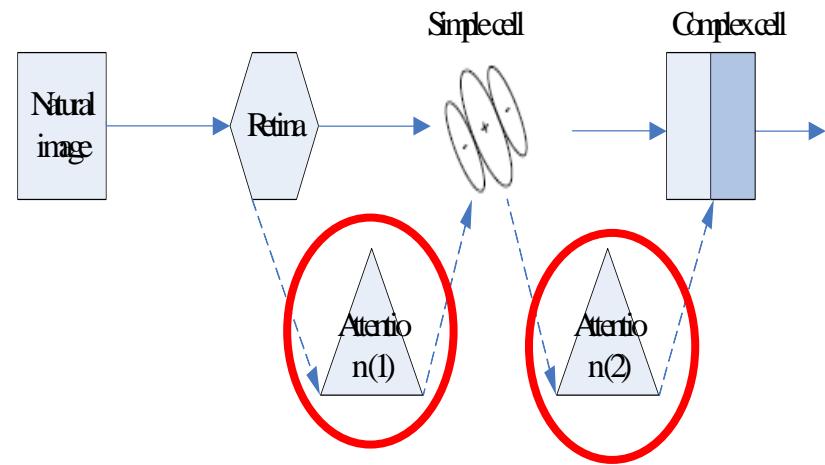
- “注意的整合竞争假设”（**Integrated Competition Hypothesis**）

Ducan把感觉，记忆，控制既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。

# 模型框架

“非均匀采样”注意选择模块

基于“响应显著度”注意选择  
模块

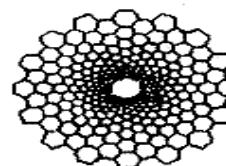


基于注意机制的稀疏编码  
模型框图

# 非均匀采样模块(1)

## 视网膜生理特性

- 视网膜中央凹区感光细胞的密度远远大于周边区域感光细胞的密度；
- 视网膜影像的分辨率在视网膜中央凹区比较高，随着距离的增大分辨率逐渐减少。



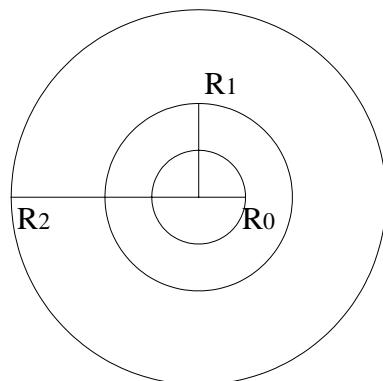
视网膜采样图示

# 非均匀采样模块 (2)

输入刺激图像非均匀地划分为不同采样率的圆形区域：

$R_0$  对应最高采样区域，它进行原采样保持与输入刺激的一致， $R_1$ ，

$R_2$ 的采样率依次降低， $R_2$ 区域代表最低的采样频率。



$$X_{ij} = \begin{cases} R_0^{-1} & \text{if } D(i,j) \leq R_0 \\ R_1^{-2} & \text{if } R_0 \leq D(i,j) \leq R_1 \\ R_2^{-3} & \text{if } R_1 \leq D(i,j) \leq R_2 \end{cases}$$

# 非均匀采样模块 (3)

递归高斯卷积模型模拟非均匀采样算法

$$R_{ij}^{(1)} = X_{ij}$$

$$R_{ij}^{(2)} = \sum_{p=-2}^{p=2} \sum_{q=-2}^{q=2} G_{pq} R_{i-p, j-q}^{(1)}$$

$$R_{ij}^{(l)} = \sum_{p=-2}^{p=2} \sum_{q=-2}^{q=2} G_{pq} R_{i-p*2^{l-1}, j-q*2^{l-1}}^{(l-1)}$$

$$[G_{pq}] = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} * \frac{1}{256}$$

# 基于响应显著度的注意选择模块

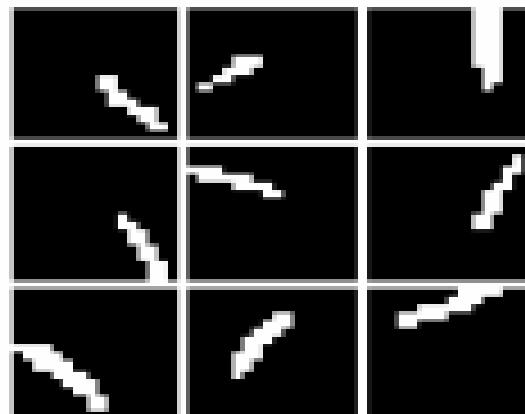
---

响应显著度（Response Salience）是度量一个神经细胞相对一组神经细胞而言对输入刺激产生响应的显著程度。

- 响应显著性是一个标量；
- 被选择作为下一级神经细胞输入的参考量  
    响应显著性值大的神经细胞输出作为下一级神经细胞的输入被进一步处理，相反，显著性值小的神经细胞输出则被忽略。

# 模式距离

- ❖ 模式距离 (**Pattern Distance**) 表示两个简单细胞表征的响应模式之间的距离，记为  $D$ 。
- ❖ 它是简单细胞的三个响应特性：位置 (**L**)，方向 (**O**) 和空间频率 (**F**) 的函数
  - 兴奋区域的中心表示位置 (**L**)；
  - 兴奋区域主轴与x轴的夹角表示方向 (**O**)；
  - 兴趣区域的面积大小表示空间频率 (**F**)。



$$D(A_i, A_j) = W_1 * N \sqrt{(L_{ix} - L_{jx})^2 + (L_{iy} - L_{jy})^2} \\ + W_2 * N |O_i - O_j| + W_3 * N |F_i - F_j|$$

# 区分距离



- ❖ 区分距离(*Discrepancy Distance*) 表示一个简单细胞与作用于同一刺激的简单细胞群之间响应模式的区别，记为 $Diff$ 。
- ❖ 它不仅与模式距离相关，还与单个细胞的响应值相关。

$$Diff(A_i, S) = \left( \sum_{A_j \in S} N(D(i, j)) * \frac{|R_j|}{\sum_{A_k \in S} R_k} \right)$$

# 响应显著度值 (RS)

- 内因—简单细胞自身的响应值 ( $R_i$ )  
响应值越大则响应显著性值也越大
- 外因—作用于同一刺激的简单细胞集合 ( $S$ ) 的影响  
按照格式塔知觉原理，具有相似特性的神经细胞之间的响应相互增强，反之，相互抑制。我们把这种影响用区分距离来度量。

$$RS(A_i) = N(|R_i|) + \lambda * (1 - Diff(A_i, S))$$

# 选择策略

## ■ 阈值选择机制 (TSM)

TSM 是一种阈值过滤算法。当响应显著性值大于阈值时，这个神经元的响应被选择进入下一级神经元的处理，否则被忽略。

$$Output(A_i) = \begin{cases} 0 & \text{if } RS(A_i) \leq T_i \\ R_i & \text{if } RS(A_i) > T_i \end{cases}$$

## ■ 比例选择机制 (PSM)

PSM 是一种瓶颈过滤算法。根据响应显著性值进行排序，选择前面的一定比例（假设为 $p$ ）的神经元进入下一级处理。

# 阈值选择算法

**算法 1:** 阈值选择算法

**Input:** 重构误差上限 (ET), 基函数集 (A), 训练图像集 (I), 训练图像集对应的稀疏编码系数集合 (R) 和响应显著度集合 (RS)。

**Output:** 阈值 T

**Method:** 算法步骤如下。

1、选择一个 T 的初始估计值;

2、用 T 过滤编码系数。如果  $RS_i$  大于 T, 则  $R'_i = R_i$ , 否则  $R'_i = 0$ ;

3、对于训练集计算重构误差值:

$$\text{Error}(R'_i, A) = \sum_{I_i} \sum_{x,y} \left[ I_i(x,y) - \sum_i R'_i A_i(x,y) \right]^2$$

4、如果  $\text{Error} \geq ET$ , 则  $T = \eta T$  其中  $0 < \eta < 1$ , 重复步骤 2 到 4; 否则算法终止, 输出 T。

# 比例因子选择算法

## 算法 2：比例因子选择算法

**Input:** 重构误差上限 (ET)，基函数集 (A)，训练图像集 (I)，训练图像集对应的稀疏编码系数集合 (R) 和响应显著度集合 (RS)。

**Output:** 比例因子 P

**Method:** 算法步骤如下。

1、选择一个 P 的初始估计值；

2、用 P 过滤编码系数。如果  $RS_i$  属于前  $P \times$  简单细胞总个数，则  $R'_i = R_i$ ，

否则  $R'_i = 0$ ；

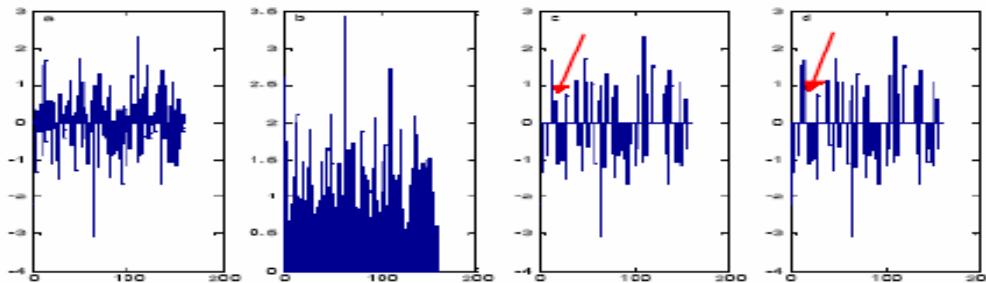
3、对于训练集计算重构误差值：

$$\text{Error}(R'_i, A) = \sum_{I_i} \sum_{x,y} \left[ I_i(x,y) - \sum_i R'_i A_i(x,y) \right]^2$$

4、如果  $\text{Error} \geq ET$ ，则  $P = (1+\eta)P$  其中  $0 < \eta < 1$ ，重复步骤 2 到 4；否则

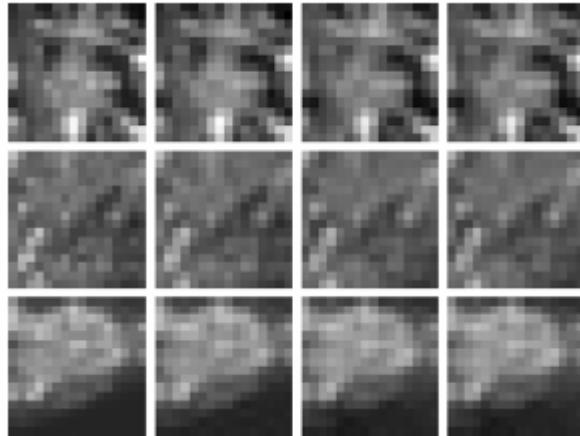
算法终止，输出 P。

# 仿真结果 (1)



输入图像刺激的各种响应系数图。a) 稀疏编码模型产生的系数; b) 各个简单细胞响应显著度值; c) AGSC-P模型产生的系数 ( $p=45\%$ ) ; d) AGSC-T模型产生的系数 ( $T=1.372$ )。

# 仿真结果 (2)



输入图像和重构图像。第一列是原始的输入图像；第二列是通过完整稀疏编码系数重构的图像；第三列是通过AGSC-P模型系数重构的图像；第四列是通过AGSC-T模型系数重构的图像。

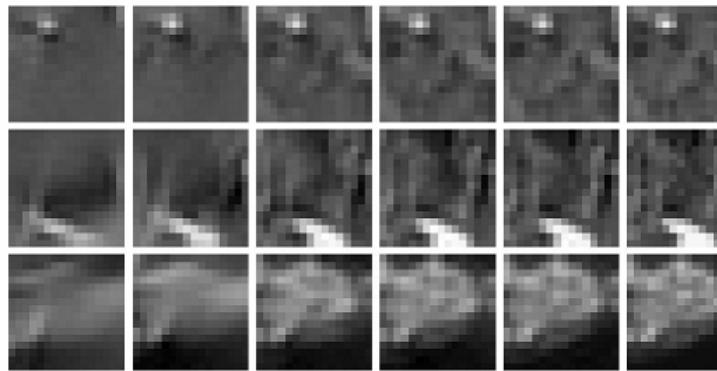
# 仿真结果 (3)

$$\text{Error}(s, A) = \sum_{x,y} \left[ I(x, y) - \sum_i s_i a_i(x, y) \right]^2$$

	Mean error	Maximum error	Minimum error
AGSC-T	0.1113	0.3929	0.0078
AGSC-P	0.1105	0.3913	0.0060
SC	0.1002	0.3209	0.0032

稀疏编码模型（简称SC）和基于注意机制的稀疏编码模型（AGSC）的重构误差比较图。

# 仿真结果



AGSC模型在比例因子P取不同值的情况下重构图像，从左到右P的取值分别为 0.1, 0.20, 0.30, 0.45, 0.80, 最右边的图像为原始图像。

