بسم الله الرحمن الرحیم

علوم شناختی

جلسه ۶

یک مدل میان‌رشته‌ای از بینایی

An Interdisciplinary Model of Vision

کاظم فولادی قلعه
دانشکده مهندسی، دانشکدگان فارابی
دانشگاه تهران

http://courses.fouladi.ir/cogsci

PART 1:
HISTORICL LANDMARKS

Chapter 2:

The Discipline Matures: Three Milestones

CAMBRIDGE
UNIVERSITY PRESS

Chapter 2.3:
An interdisciplinary model of vision

CAMBRIDGE
UNIVERSITY PRESS

Overview

- Introduce Marr's distinction between levels of explanation
- Explain Marr's top-down model of how the levels relate
- Sketch out the basic elements of Marr's theory of early vision

Background

The different disciplines involved in cognitive science operate at very different **levels** and use different **techniques**

- Experimental psychology
- Computer science/AI
- Theoretical perspectives from **information theory** and **theory of computation**

Different levels

- **Miller**’s psychophysics experiments ⇒ explore the cognitive capacities and limitations of the subject
- **Winograd**’s SHRDLU ⇒ explores how complex processes occur algorithmically
- **Broadbent**’s information-processing model ⇒ explores how information flows within and between various subpersonal sub-systems and mechanisms
- **Kosslyn** focusing on different ways of coding information
- And we haven’t even started thinking about the brain …

Marr’s approach

- Distinguished different explanatory tasks at different levels
- Gave a general theoretical framework for combining them
- Apply the framework in considerable detail to a single example ⇒ **the early visual system**

Marr’s three levels

- 3 different types of analysis of an information-processing system
  - Computational
  - Algorithmic
  - Implementational

Computational analysis

- Form of task analysis of a cognitive system

a) Identifies the **specific** information-processing problem that the system is configured to solve

b) Identify **general** constraints upon any solution to that problem

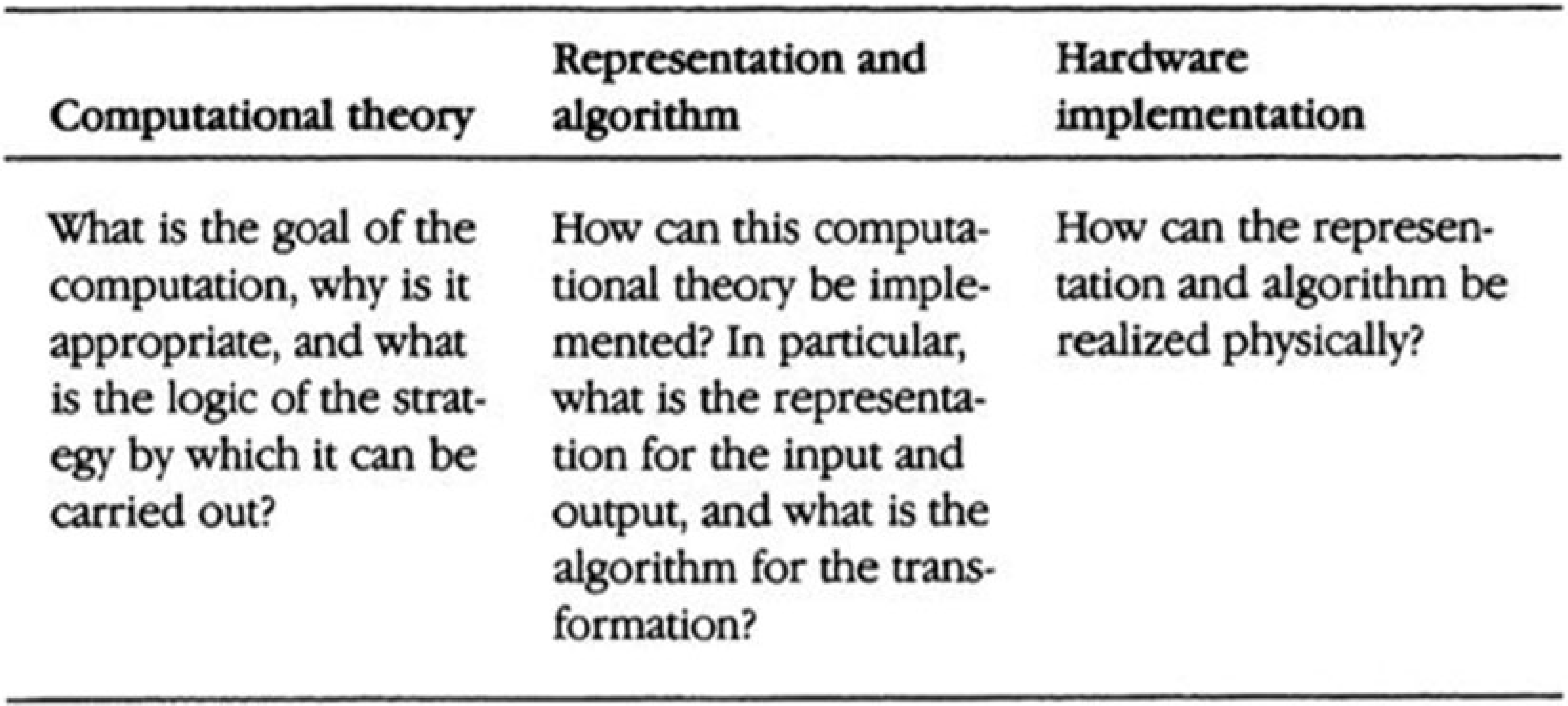
Algorithmic analysis

- Explains how the cognitive system **actually** performs the information-processing task
  - identifies input information and output information
  - identifies algorithm for transforming input into required output
  - specifies how information is **encoded**

Implementational analysis

- Finds a **physical realization** for the algorithm
  - Identify neural structures realizing the basic representational states to which the algorithm applies [e.g. populations of neurons]
  - Identify neural mechanisms that transform those representational states according to the algorithm

Marr’s three levels



Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

A table illustrating the three different levels that Marr identified for explaining information-processing systems. Each level has its own characteristic questions and problems. (From Marr 1982)

Turing machine example

• Computational	Characterization of multiplication function
• Algorithmic	Turing machine table
• Implementational	Construction of a physical Turing machine

Cognitive Science © José Luis Bermúdez / Cambridge University Press 2020

Marr's computational analysis of visual system

- Two basic conclusions from his task analysis
  - The visual system's job is to provide a 3D representation of the visual environment that can serve as input to recognition and classification processes – primarily information about shape of objects and their spatial distribution
  - This 3D representation is on an object-centered rather than viewer-centered frame of reference

Experimental evidence

- Possibility of double dissociations between perceptual abilities and recognitional abilities
- Right parietal lesions – recognition abilities preserved, but problems in perceiving shapes from unusual perspectives
- Left parietal lesions - shape perception intact, but recognition and identification impaired
- Suggested to Marr that visual system provides input to recognition systems

Theoretical considerations

- Recognitional abilities are constant across changes in how things look to the perceiver due to
  - orientation of object
  - its distance from perceiver
  - partial occlusion by other objects
- So - visual system provides information to recognition systems that abstracts away from these perspectival features – observer-independent representation

Algorithmic analysis

- Input = light arriving at retina
- Output = 3D representation of environment
- Questions:
  - what sort of information is extracted from the light at the retina?
  - how does the system get from this information to a 3D representation of the environment?

The challenge

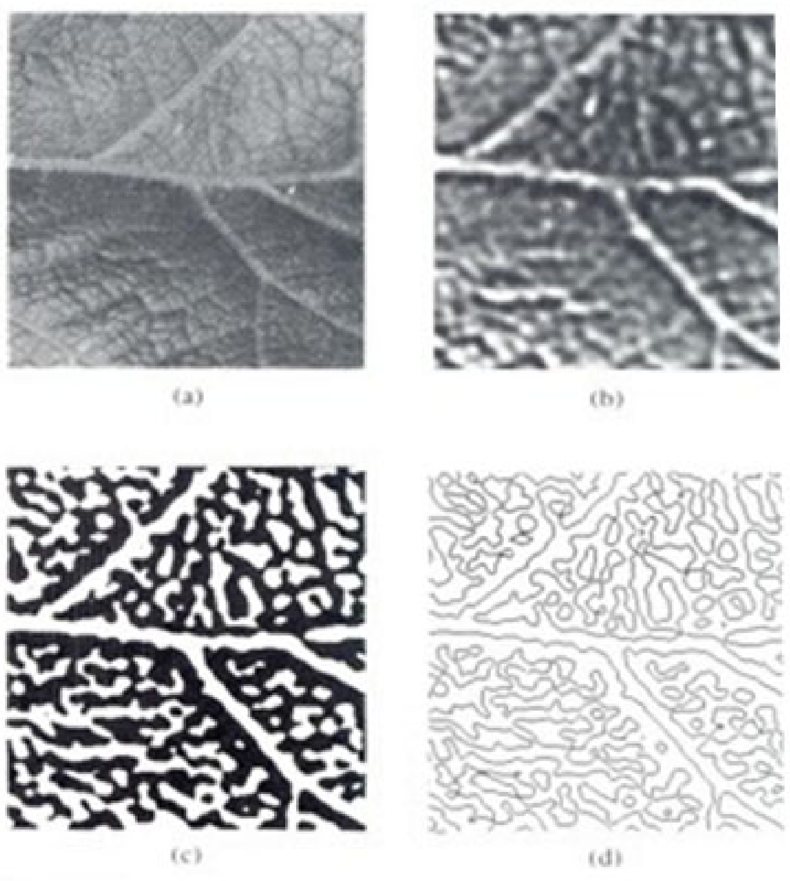
- “From an information-processing point of view, our primary purpose is to define a representation of the image of reflectance changes on a surface that is suitable for detecting changes in the image’s geometrical organization that are due to changes in the reflectance of the surface itself or to changes in the surface’s orientation or distance from the viewer” (Marr, *Vision* p. 44)

- Need to find representational primitives that allow inference backwards from structure of image to structure of environment

Representational primitives

- Basic information at retina =
  intensity value of light at each point in the retinal image
  - changes in intensity value provide clues as to surface boundaries
- Primitives allow structure to be imposed on patterns of intensity changes
  - E.g. zero crossings (sudden intensity changes)

Zero crossings



- If we plot changes in intensity on a graph, then radical discontinuities will be signalled by the curve crossing zero

- Marr proposed a Laplacian/Gaussian filter to detect zero crossings

Primal sketch

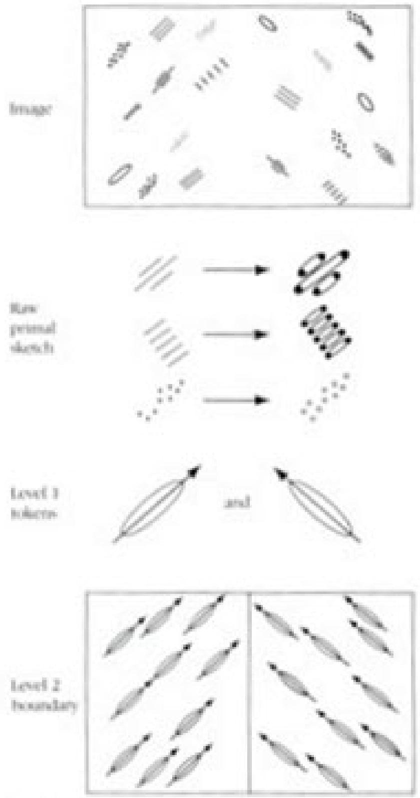
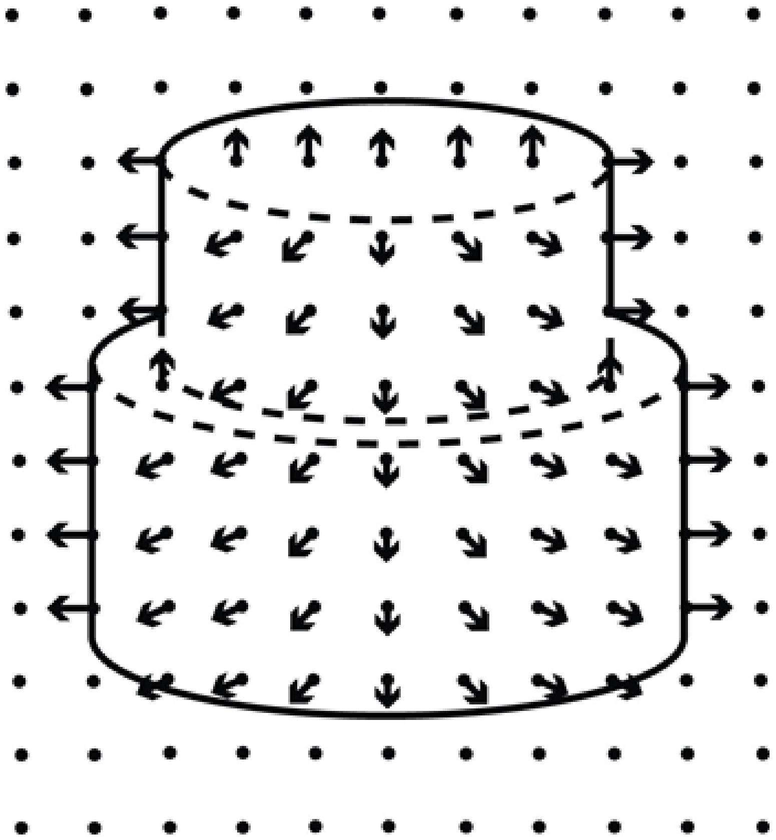


Figure 2–7. A diagrammatic representation of the descriptions of an image at different scales which together constitute the primal sketch. At the lowest level, the raw primal sketch faithfully follows the intensity changes and also represents terminations, denoted here by filled circles. At the next level, oriented tokens are formed for the groups in the image. At the next level, the difference in orientations of the groups in the two halves of the image causes a boundary to be constructed between them. The complexity of the primal sketch depends upon the degree to which the image is organized at the different scales.

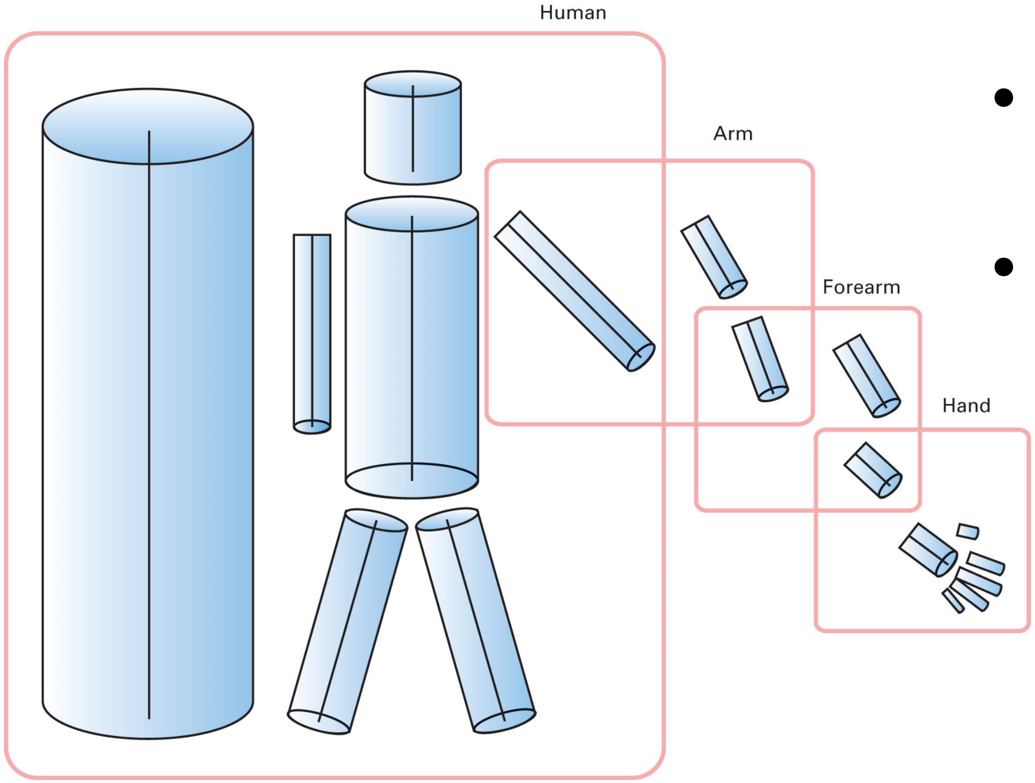
- identifies intensity changes in the 2D image
- basic information about the geometric organization of those intensity changes
- Primitives include:
  - zero-crossings
  - virtual lines
  - groups

2.5D sketch



- Displays orientation of visible surfaces in viewer-centered coordinates
- Represents distance of each point in visual field from viewer
- Also orientation of each point and contours of discontinuities
- Very basic information about depth

3D sketch

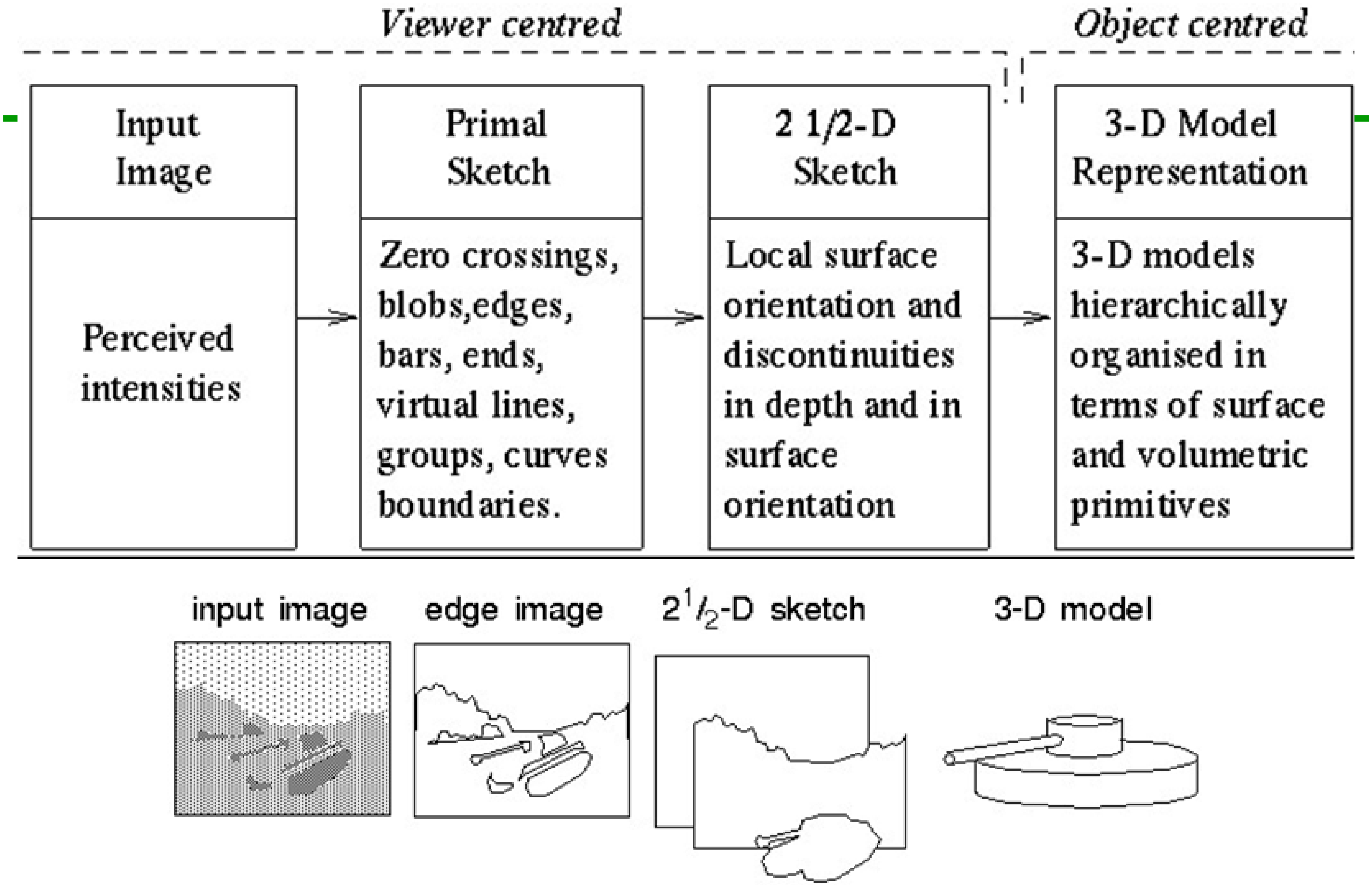


- characterizes shapes and their spatial organization
- object-centered
- basic volumetric and surface primitives are schematic (facilitates recognition)

Representation in the 3D sketch

- Depends upon many shapes being recognizable as ensembles of generalized cones

- Generalized cones are easy to represent
  - vector describing path of the figures axis of symmetry
  - vector specifying perpendicular distance from every point on axis to shape's surface


Viewer centred
Object centred
Input Image
Perceived intensities
Primal Sketch
Zero crossings, blobs,edges, bars, ends, virtual lines, groups, curves boundaries.
2 1/2-D Sketch
Local surface orientation and discontinuities in depth and in surface orientation
3-D Model Representation
3-D models hierarchically organised in terms of surface and volumetric primitives
input image
edge image
2 1/2-D sketch
3-D model

Basic top-down assumption

- Explanation is top-down because of <u>underdetermination</u>
  - Many different algorithms can in principle compute the same task
  - There are many different ways of implementing a given algorithm
- Multiple realizability ⇒
  more informative to work at higher levels
- Relatively little implementational detail in Marr

Implementing zero crossings

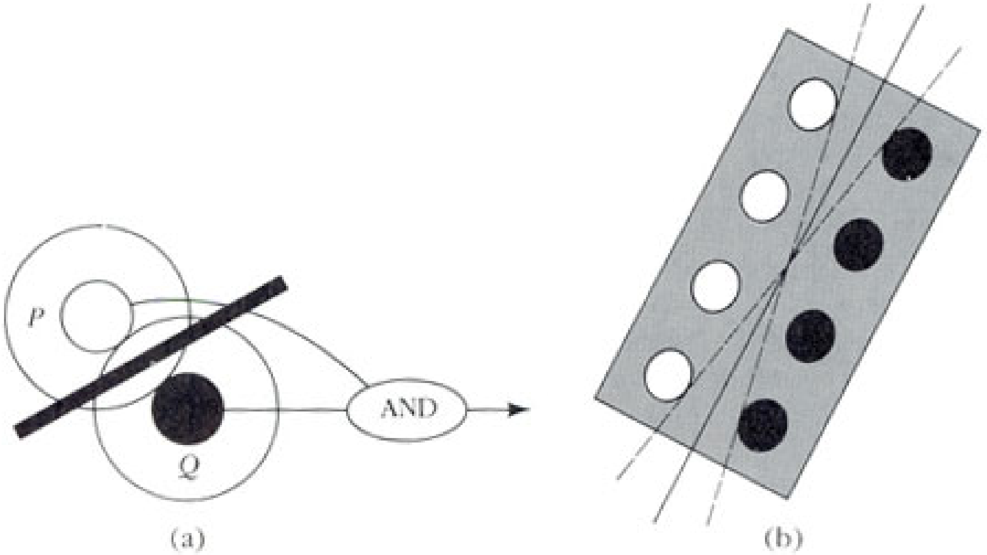
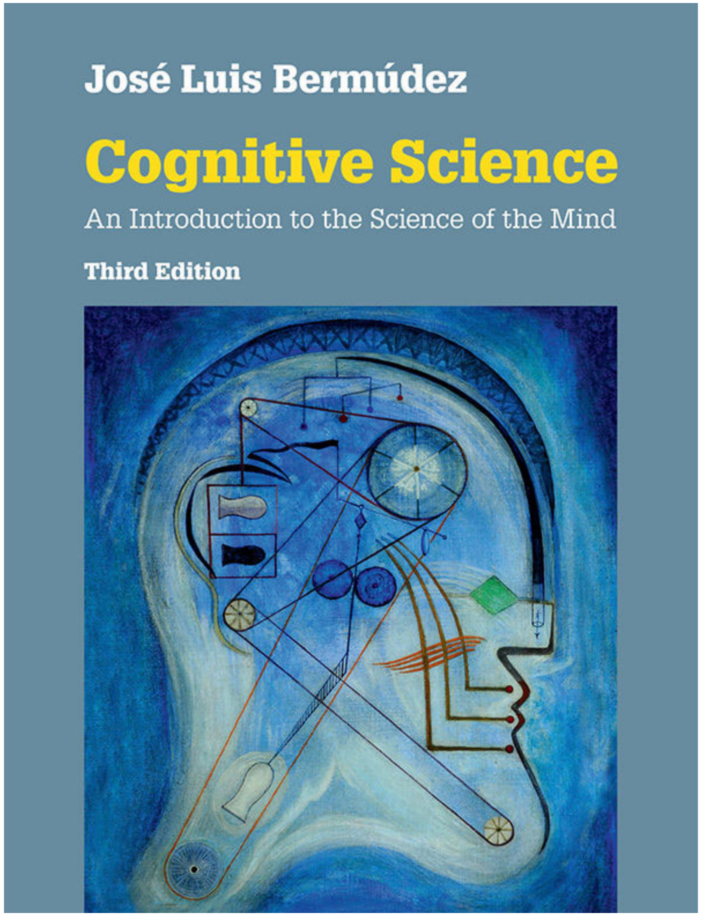


Figure 2–18. A mechanism for detecting oriented zero-crossing segments. In (a), if *P* represents an on-center geniculate X-cell receptive field, and *Q* an off-center, then a zero-crossing must pass between them if both are active. Hence, if they are connected to a logical AND gate as shown, the gate will detect the presence of the zero-crossing. If several are arranged in tandem as in (b) and are also connected by logical AND's, the resulting mechanism will detect an oriented zero-crossing segment within the orientation bounds given roughly by the dotted lines. Ideally, we would use gates that responded by signaling their sum only when all their *P* and *Q* inputs were active. (Reprinted, by permission, by D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond. B 204*, pp. 301–328.)

Marr key points

1) Tri-level hypothesis very influential
2) Classic example of top-down analysis
3) Most detail comes at algorithmic level
4) Neurobiology only comes in at the implementational level

منبع اصلی



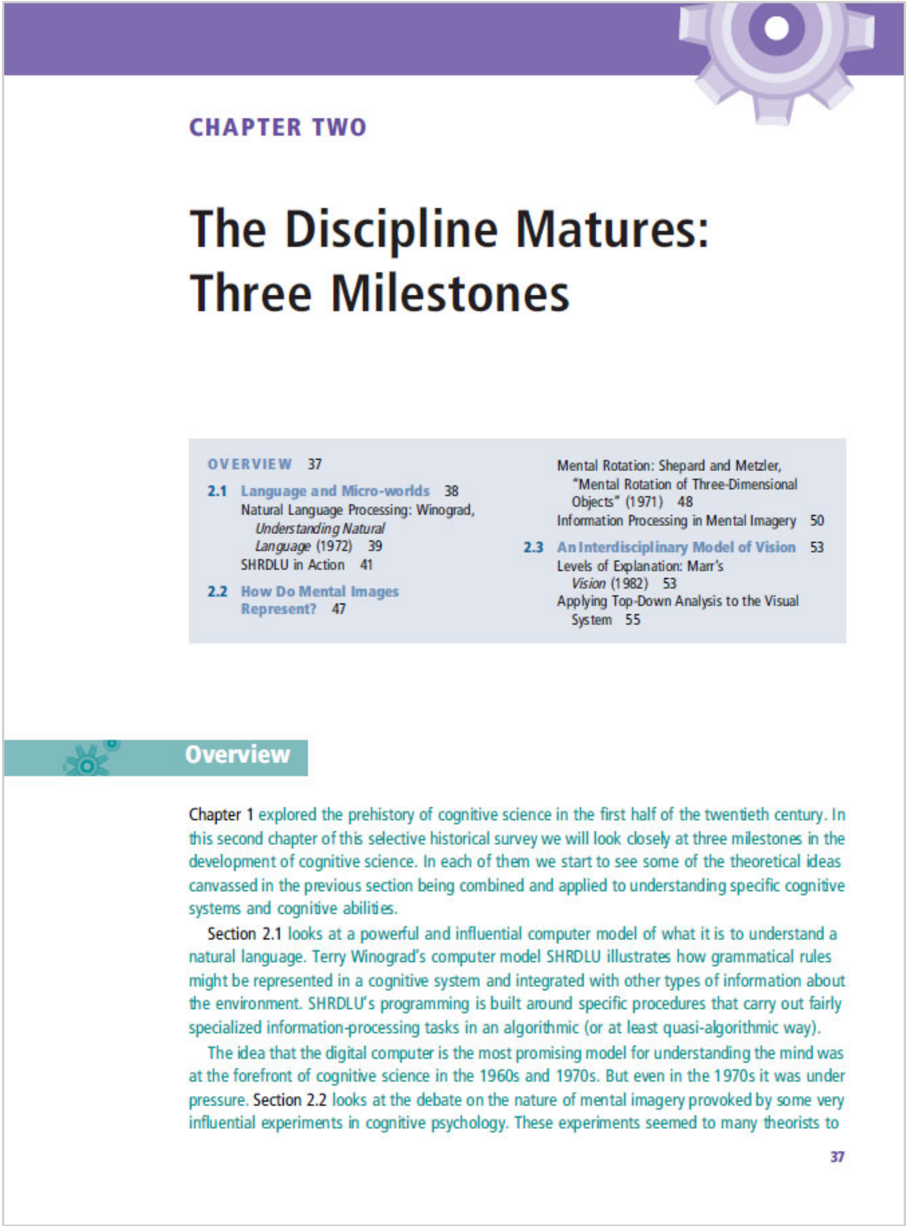
José Luis Bermúdez

Cognitive Science

An Introduction to the Science of the Mind

Third Edition

José Luis Bermúdez,
Cognitive Science:
An Introduction to the Science of the Mind,
3rd ed., Cambridge University Press, 2020.
Chapter 2 (Section 2.3)



CHAPTER TWO

The Discipline Matures: Three Milestones

Overview

Chapter 1 explored the prehistory of cognitive science in the first half of the twentieth century. In this second chapter of this selective historical survey we will look closely at three milestones in the development of cognitive science. In each of them we start to see some of the theoretical ideas canvassed in the previous section being combined and applied to understanding specific cognitive systems and cognitive abilities.

Section 2.1 looks at a powerful and influential computer model of what it is to understand a natural language. Terry Winograd's computer model SHRDLU illustrates how grammatical rules might be represented in a cognitive system and integrated with other types of information about the environment. SHRDLU's programming is built around specific procedures that carry out fairly specialized information-processing tasks in an algorithmic (or at least quasi-algorithmic way).

The idea that the digital computer is the most promising model for understanding the mind was at the forefront of cognitive science in the 1960s and 1970s. But even in the 1970s it was under pressure. Section 2.2 looks at the debate on the nature of mental imagery provoked by some very influential experiments in cognitive psychology. These experiments seemed to many theorists to

37