Digital image processing vs. computer vision Higher-level anchoring

Václav Hlaváč

Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics
Center for Machine Perception
http://cmp.felk.cvut.cz/~hlavac, hlavac@fel.cvut.cz

Outline of the lecture:

- Digital image processing × image analysis × computer vision.
- Vision vs. perception.
- Why is vision hard?
- Interpretation, its significance for images.

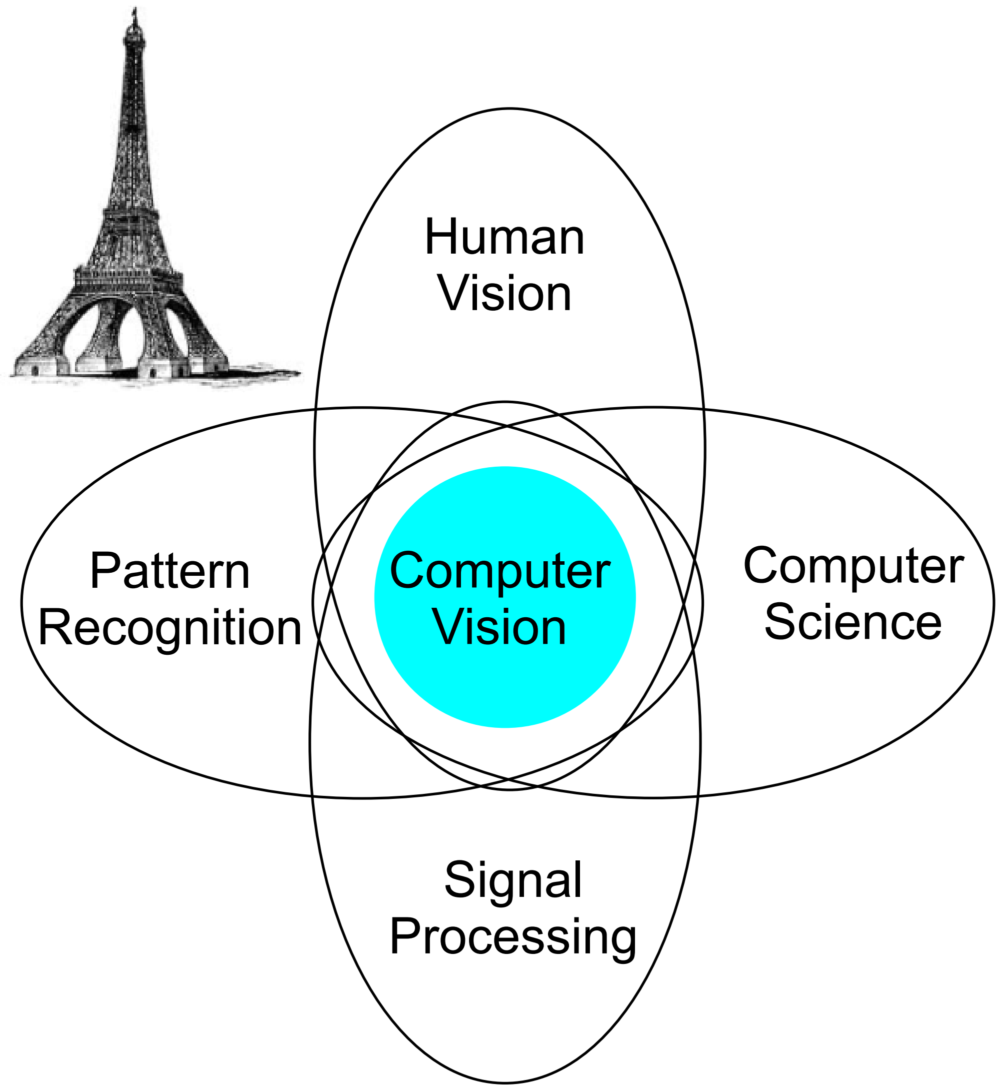
What is computer vision?

Computer vision is the science and technology of machines that see and perceive.

- As a scientific discipline:
  the theory for building artificial systems that obtain information from images.
- As a technological discipline:
  construction of computer vision systems.
  Computer vision $=$ Camera $+$ Computer $+$?
- Images (e.g.):
  - views from multiple cameras,
  - a video sequence,
  - multi-dimensional data from a medical scanner.

Why to study image processing, analysis and computer vision?

- Computer vision has grown on four pillars (at least): (1) Computer science; (2) Signal processing; (3) Pattern recognition; (4) Human vision.
- Attempts since 1960s.
- A rich methodology.
- Interesting interdisciplinary ties.
- Exciting insights into human vision.
- An important information source and modality in the information age.



What is computer vision used for?

- Controlling processes (e.g., an industrial robot or an autonomous vehicle).
- Detecting events (e.g., for visual surveillance, people counting, detecting a launching ballistic missile from a satellite).
- Organizing information (e.g., for indexing databases of images and image sequences).
- Modeling objects or environments (e.g. industrial inspection, medical image analysis or topographical modeling).
- Interaction (e.g. as the input to a device for computer-human interaction).
- . . .

Perception

- Process of attaining awareness or understanding of sensory information.
- A task is far more complex than it was imagined in the 1950s and 1960s:
  "Building perceiving machines would take about a decade."
  However, it still very far from reality.
- Aristotle's five senses are: sight, hearing, touch, smell, taste.
- Perception conjectures a dynamic relationship between:
  "description" (in the brain)
  - $\leftrightarrow$ senses,
  - $\leftrightarrow$ surrounding,
  - $\leftrightarrow$ memory.

Notes on human (visual) perception

What do you see in the picture?

Notes on human (visual) perception

What do you see in the picture?

Vision is very natural for humans and many animals.

Notes on human (visual) perception

What do you see in the picture?

Vision is very natural for humans and many animals.

- Vision is not for free::
-
-

Notes on human (visual) perception

What do you see in the picture?

Vision is very natural for humans and many animals.

- Vision is not for free::
  - About 50% of the primate's cortex deals with the processing of visual (Felleman-van Essen 1991).

Notes on human (visual) perception

What do you see in the picture?

Vision is very natural for humans and many animals.

- Vision is not for free::
  - About 50% of the primate's cortex deals with the processing of visual (Felleman-van Essen 1991).
  - The brain consumes approximately 20 % of the human's energy.
-
-

Notes on human (visual) perception

What do you see in the picture?

Vision is very natural for humans and many animals.

- Vision is not for free::
  - About 50% of the primate's cortex deals with the processing of visual (Felleman-van Essen 1991).
  - The brain consumes approximately 20 % of the human's energy.
- Making a computer see and perceive like humans do means to solve a large part of the AI problem (which is difficult, close to impossible).
-

Notes on human (visual) perception

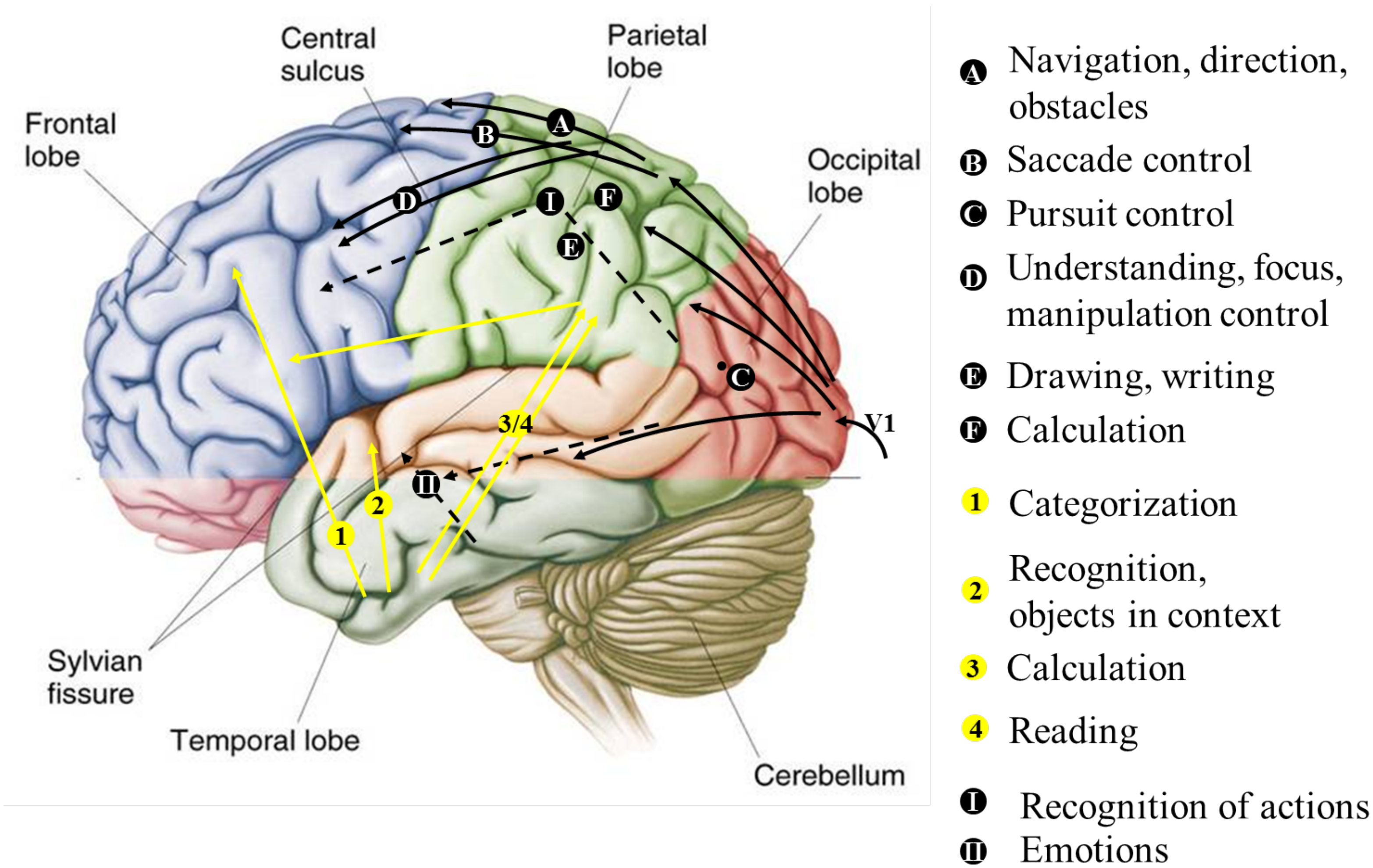
What do you see in the picture?

Vision is very natural for humans and many animals.

- Vision is not for free::
  - About 50% of the primate's cortex deals with the processing of visual (Felleman-van Essen 1991).
  - The brain consumes approximately 20 % of the human's energy.
- Making a computer see and perceive like humans do means to solve a large part of the AI problem (which is difficult, close to impossible).
- A lot of high level knowledge, semantic information and context is explored

Human vision

- Visual cortex occupies about 50% of the Macaque brain.
- More human brain is devoted to vision than to anything else.



Human vision as opposed toh computer vision

Vision allows both humans and animals to perceive and understand the world surrounding them.

Cognitive science investigates vision in biological systems:

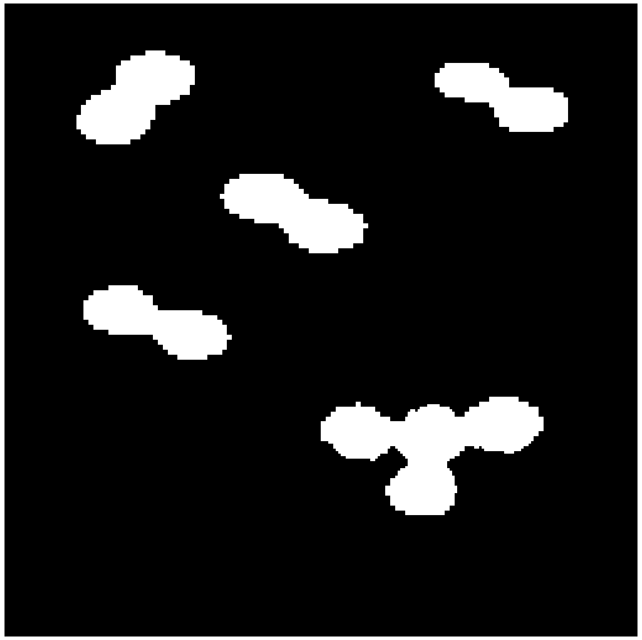
- It seeks empirical models which adequately describe biological vision.
- It sometimes describes vision as a computational system.

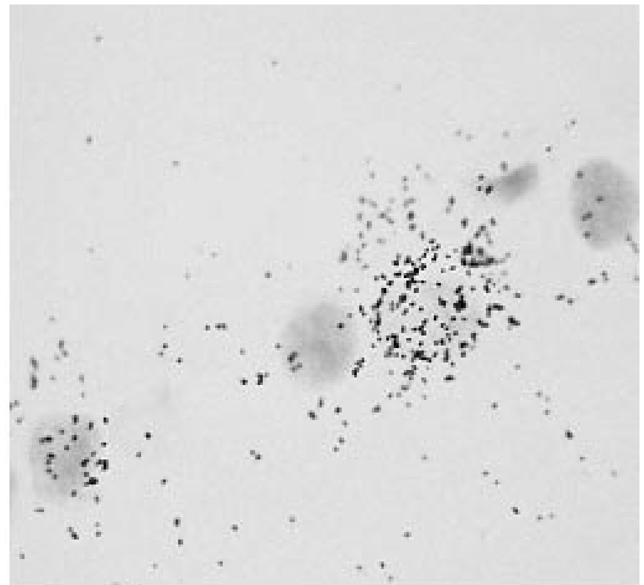
Computer vision aims at engineering solutions, but its research is also interested in biological vision:

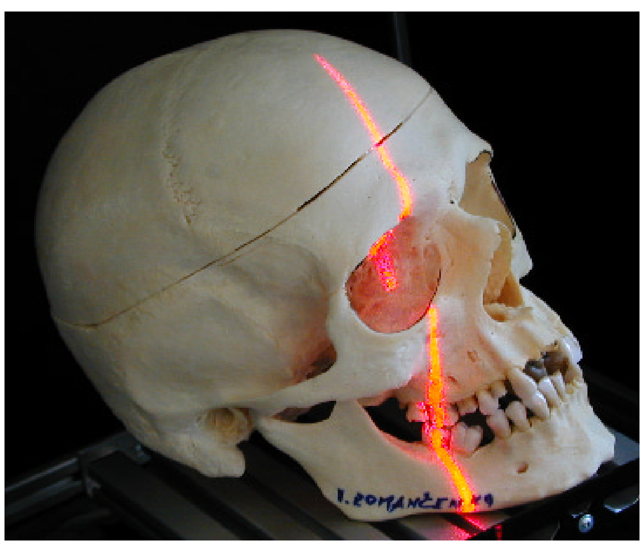
- Biological vision systems cope with tasks not yet solved in computer vision. They provide ideas for engineering solutions.
- Technical requirements for vision systems often match requirements for biological vision.

Caution: Mimicking biological vision does not necessarily provide the best solution for a technical problem.

Examples of input images







Why is computer vision hard?
Let us find six reasons (at least).

Why is computer vision hard? Let us find six reasons (at least).

Loss of information in 3D $\rightarrow$ 2D due to perspective transformation (mathematical abstraction = pinhole model).

Measured brightness is given by a complicated image formation physics. Radiance ($\approx$ brightness) depends on light sources intensity and positions, observer position, surface local geometry, and albedo. Inverse task is ill-posed.

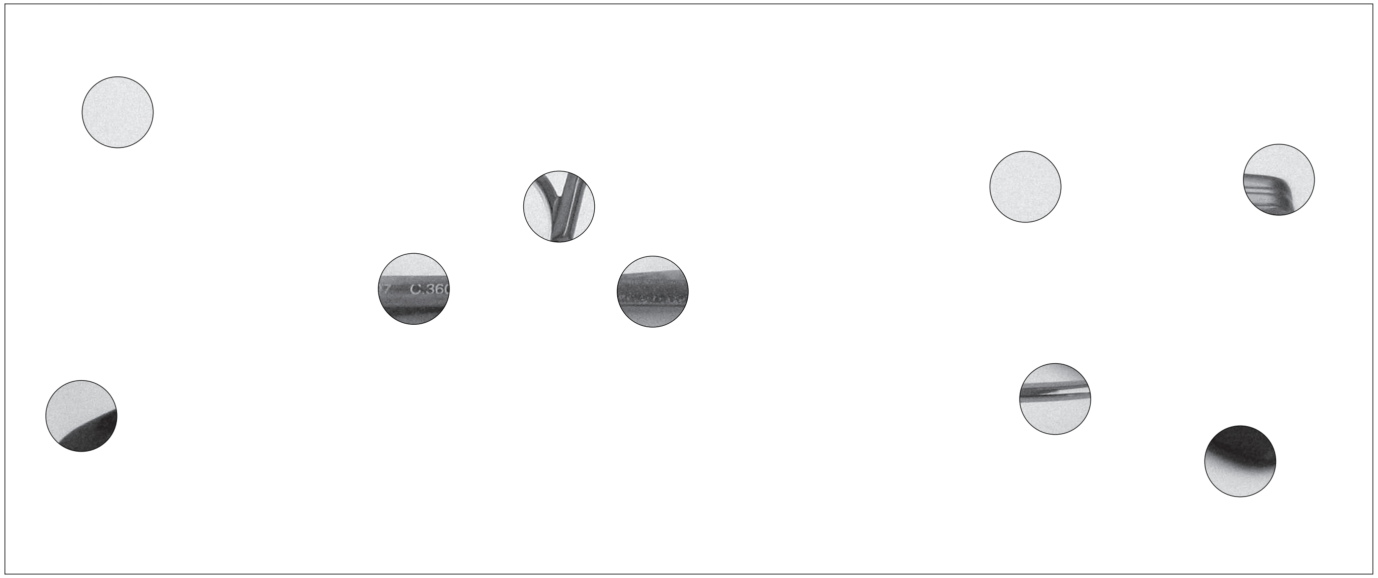
Inherent presence of noise as each real world measurement is corrupted by noise.

A lot of data Sheet A4, 300 dpi, 8 bit per pixel = 8.5 Mbytes. Non-interlaced video 512 $\times$ 768, RGB (24 bit) = 225 Mbits/second.

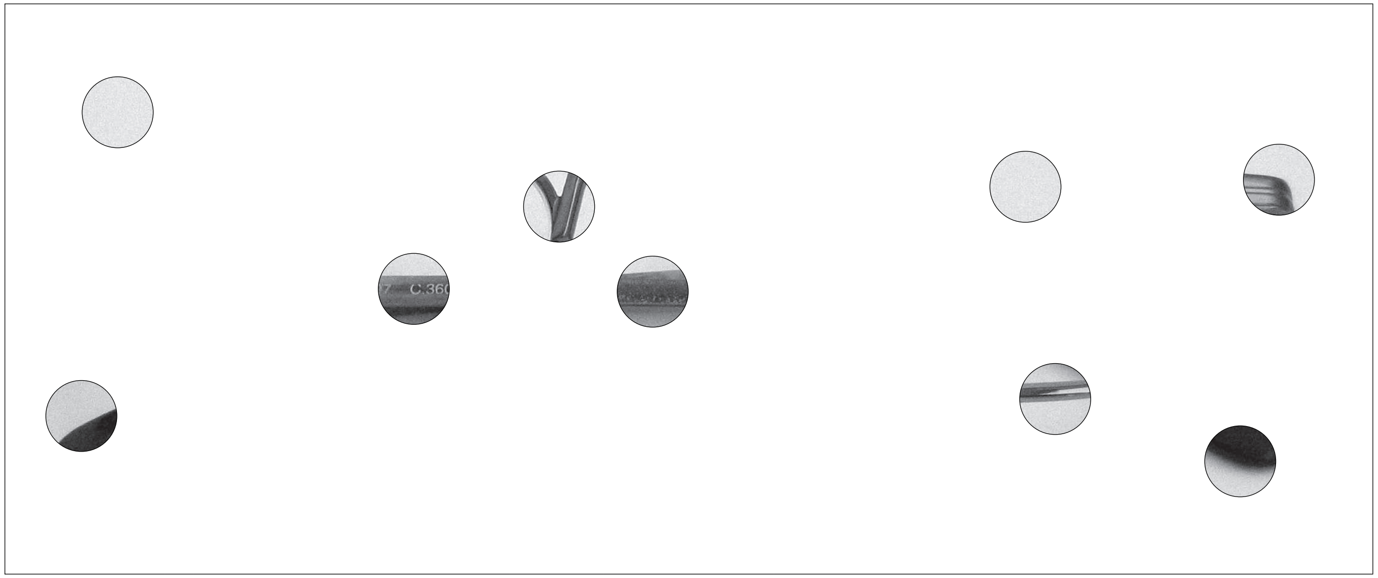
Interpretation needed (to be discussed soon).

Local window vs. need for global view

Insufficiency of local view, illustration



Insufficiency of local view, illustration



Interpretation and its role, semantics

Interpretation: Observation $\rightarrow$ Model
Syntax $\rightarrow$ Semantics

Examples:

- Looking out of the window $\rightarrow$ {rains, does not rain}.
- An apple on the conveyer belt $\rightarrow$ {class 1, class 2, class 3}.
- Traffic scene $\rightarrow$ seeking number plate of a car.

Theoretical background: mathematical logic, theory of formal languages.

Deep philosophical problem: Gödel's incompleteness theorems, informally: a logic system with propositions cannot be proved or disproved.

From a low to a high level processing from the apriori knowledge point of view

Low level of knowledge (or none) = digital image processing

- Images are not interpreted. Methods are independent on a specific application area.
- Signal processing methods are used, e.g., the 2D Fourier transform.

Middle level of knowledge = image analysis

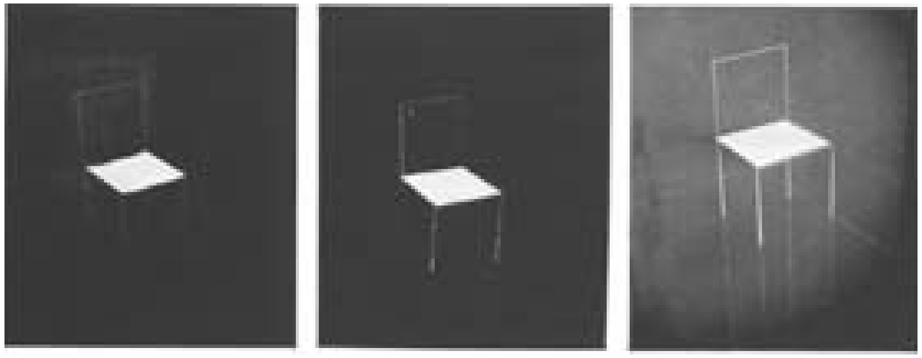
- Often 2D images only, e.g. cell images in an optical microscope.
- Interpretation explores an important additional knowledge allowing to solve tasks unsolvable otherwise.

High level of knowledge = computer vision, e.g., understanding content of a 3D scene from images and videos

- The most general task formulations, 3D world, changing scenes.
- Complicated, interpretation is explored, feed back explored, artificial intelligence methods.
- Goals are overambitious. Involved tasks are underconstraint and too ambitious. Tasks have to be radically simplified.

Role of the apriori knowledge, counterexample

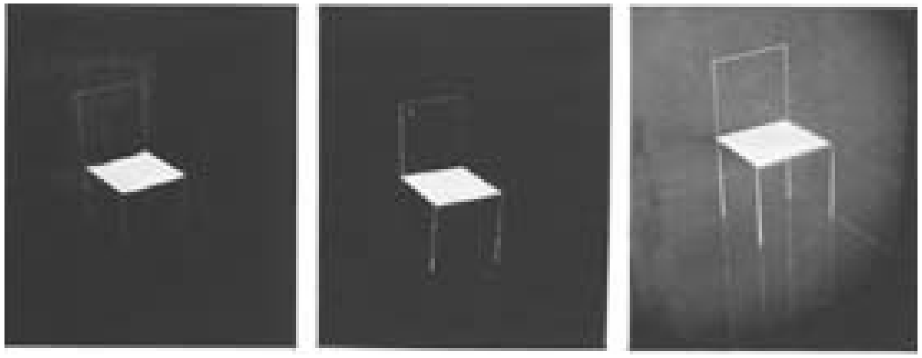
- Apriori knowledge about "our world" enables humans to understand multi-meaning images.
- Of course, apriori assumptions can mislead the human too . . .
- Counterexample: Ames chair.



We can see chairs.

Role of the apriori knowledge, counterexample

- Apriori knowledge about “our world” enables humans to understand multi-meaning images.
- Of course, apriori assumptions can mislead the human too ...
- Counterexample: Ames chair.



We can see chairs.

Actually, there are no chairs.

The ultra brief history of computer vision

1966 M. Minsky assigns computer vision as an undergrad summer project.

1960 Interpretation of synthetic worlds, e.g. block world for robots.

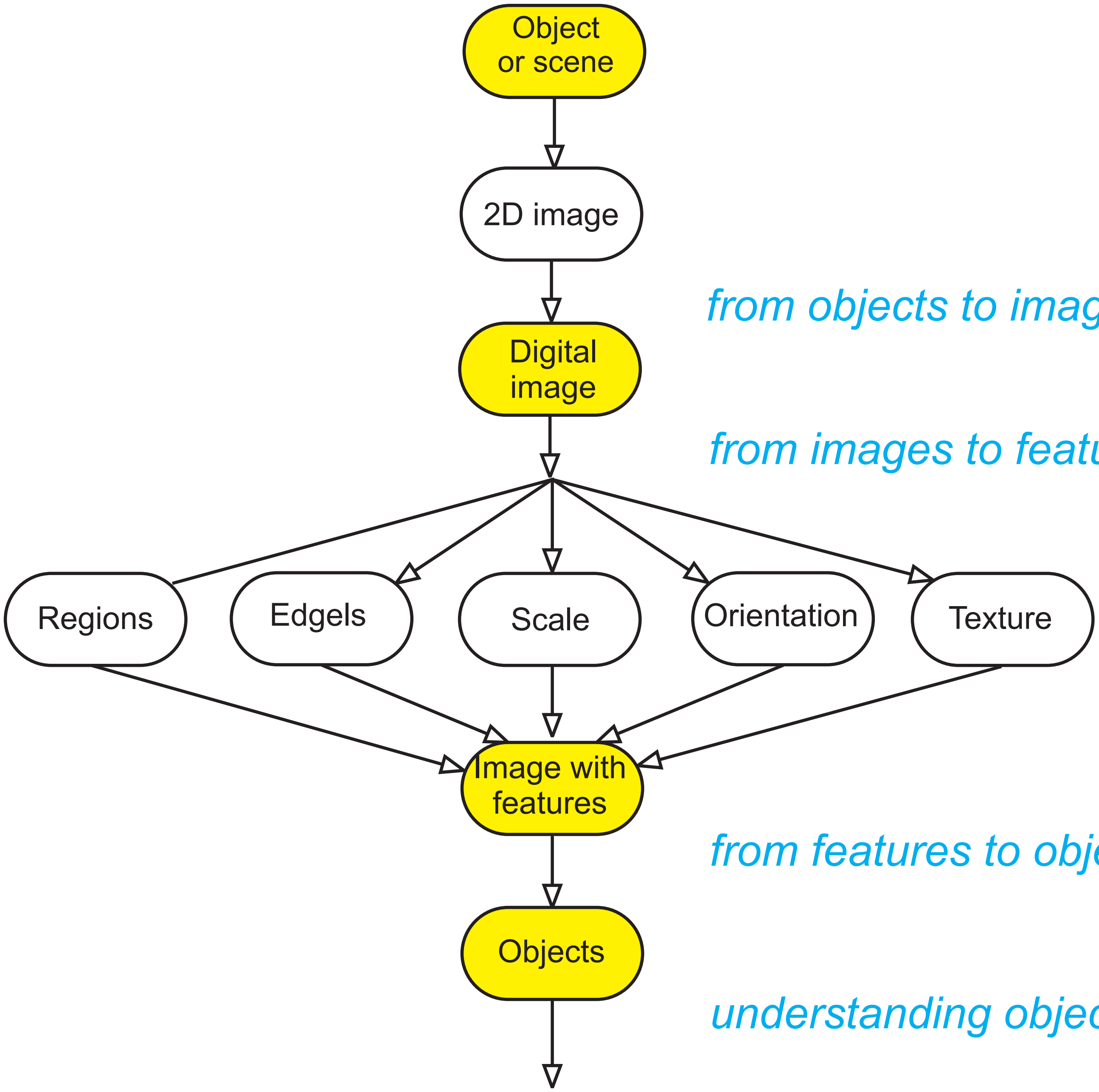
1970s Some progress on interpreting selected images.

1980s Artificial neural nets come and go; shift toward geometry and increased mathematical rigor; inspiration from biological vision (D. Marr et al.)

1990s Face recognition; statistical analysis in vogue; geometry of vision.

2000s Broader recognition; large annotated datasets available; video processing starts.

Image-based recogniton hierarchy of representations



from objects to images

from images to features

from features to objects

understanding objects