Digital image processing vs. computer vision Higher-level anchoring

Václav Hlaváč

Czech Technical University in Prague

Czech Institute of Informatics, Robotics and Cybernetics

160 00 Prague 6, Jugoslávských partyzánů 1580/3, Czech Republic

http://people.ciirc.cvut.cz/hlavac, vaclav.hlavac@cvut.cz

also Center for Machine Perception, http://cmp.felk.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.

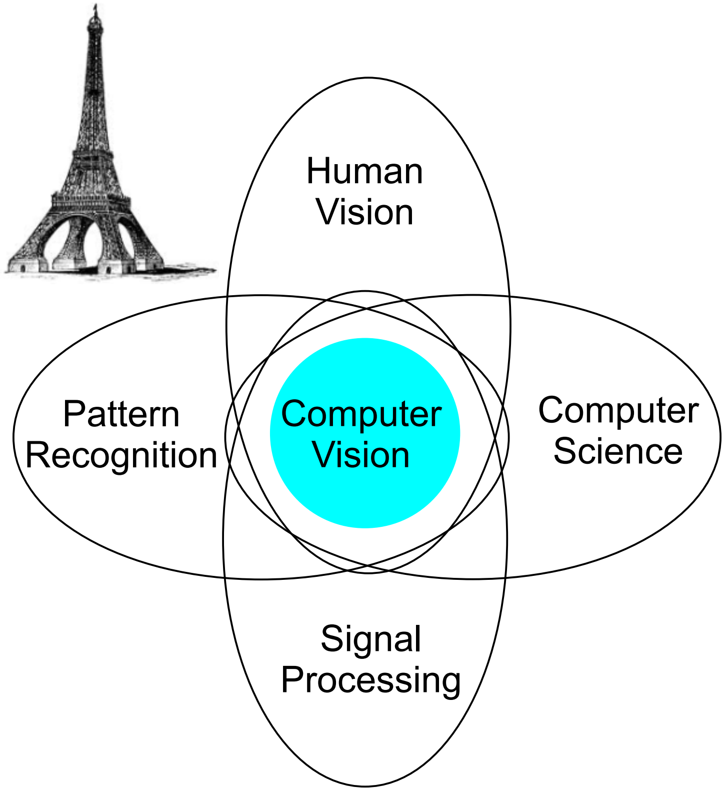
What is computer vision?

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?

- Detecting, segmenting, describing, recognition (more poetically: understanding, perceiving) object of interest in 2D or 3D imagery.
- Detecting events (e.g., for visual surveillance, people counting, detecting a launching ballistic missile from a satellite).
- Organizing information based on imagery (e.g., for indexing databases of images and image sequences).
- Controlling processes (e.g., an industrial robot or an autonomous vehicle).
- Modeling objects or environments (e.g. in industrial inspection, medical image analysis or topographical modeling).
- Human-computer interaction (e.g. as an input to computer game, cf. a depth measuring using Kinect).
- etc.

Perception

- Process of attaining awareness or understanding of sensory percepts.
- A task is far more complex than it was imagined in the 1950s and 1960s:
  - Back then: "Building perceiving machines would take about a decade."
  - However, it is 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.

Image courtesy:
http://www.richardsonthebrain.com/

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 info. (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 info. (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 info. (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).
-

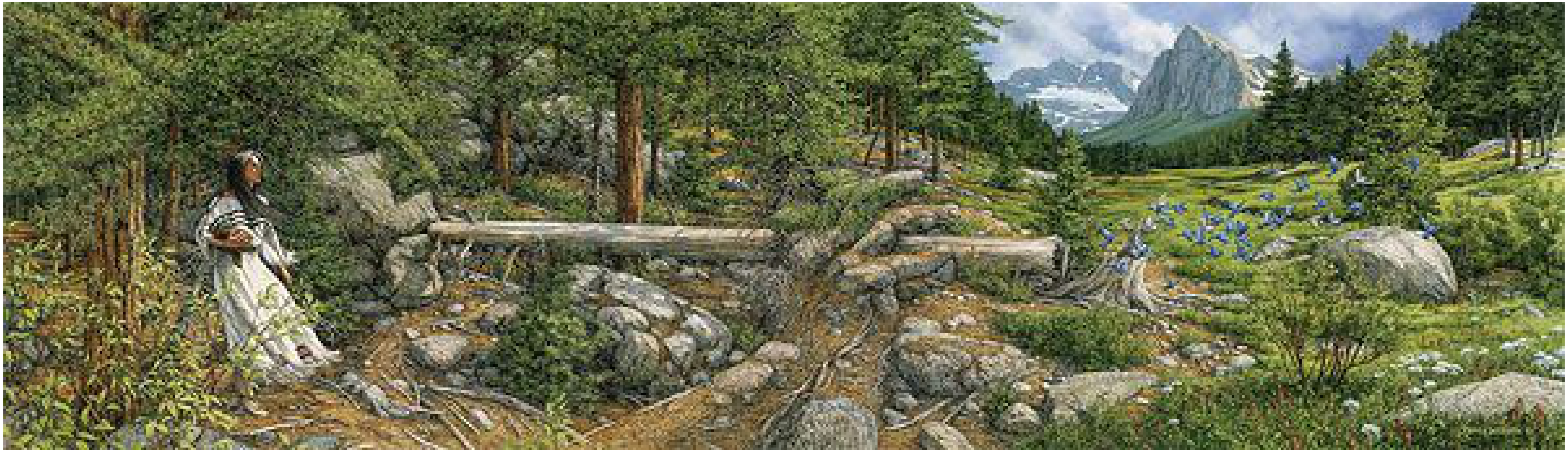
Notes on human (visual) perception

What do you see in the picture?

Vision is very natural for humans and many animals.

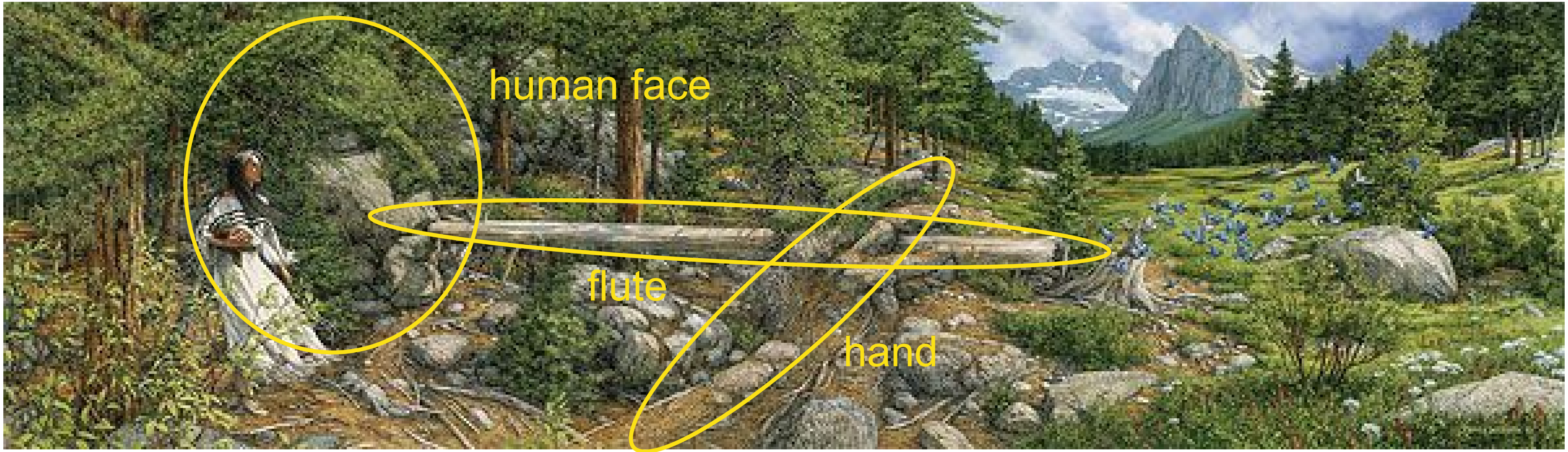
- Vision is not for free::
  - About 50% of the primate's cortex deals with the processing of visual info. (Felleman-van Essen 1991).
  - The brain consumes approximately 20 % of the human's energy.
- Making a computer see and perceive like humans do means to solve a large part of the AI problem (which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

Example, painting with camouflaged object



American painter (she) Bev Dolittle *1949, progagonist of a camouflage technique

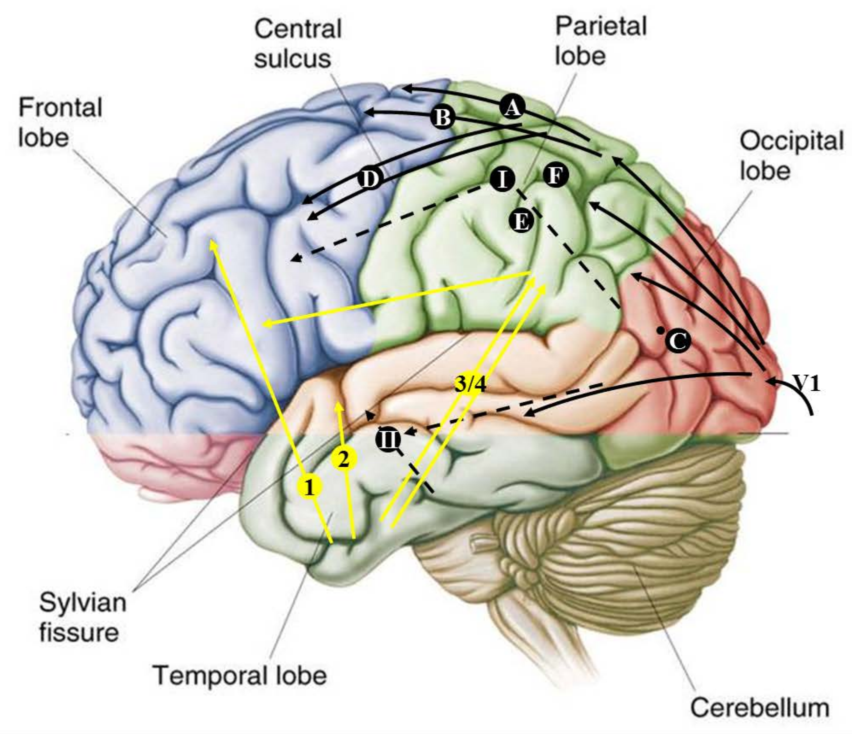
Example, painting with camouflaged object B



American painter (she) Bev Dolittle *1949, progagonist of a camouflage technique

Human vision

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



- A Navigation, direction, obstacles
- B Saccade control
- C Pursuit control
- D Understanding, focus, manipulation control
- E Drawing, writing
- F Calculation
- 1 Categorization
- 2 Recognition, objects in context
- 3 Calculation
- 4 Reading
- I Recognition of actions
- II Emotions

Human vision as opposed to computer vision

Vision allows both humans and animals perceiving and understanding the world surrounding them.

Cognitive science investigates vision in biological systems:

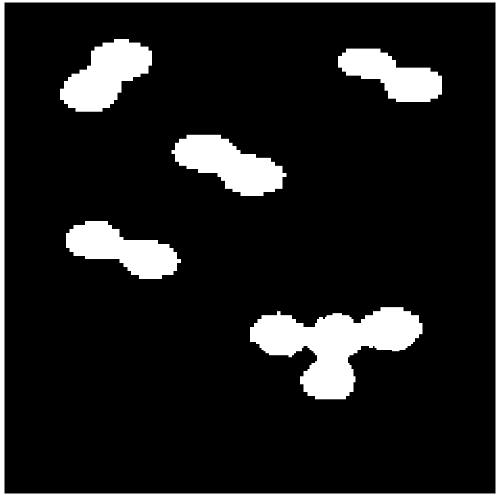
- It seeks empirical models which adequately describe biological vision.
- It sometimes describes vision as a computational system.

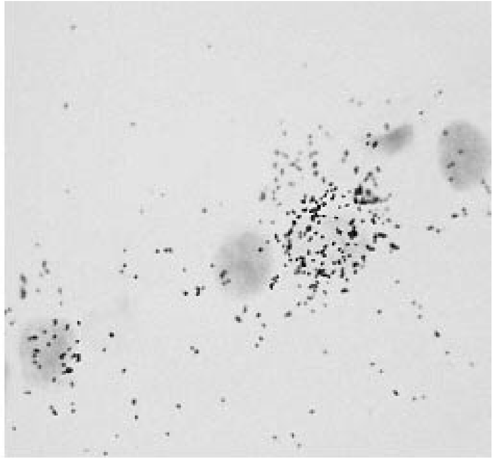
Computer vision aims at engineering solutions, but its research is also interested in biological vision:

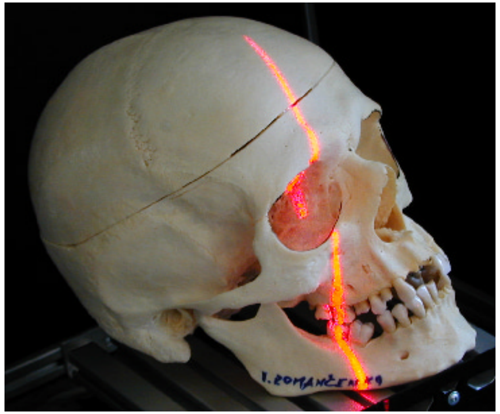
- Biological vision systems cope with tasks not yet solved in computer vision. They provide ideas for engineering solutions.
- Technical requirements for vision systems often match requirements for biological vision.

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.

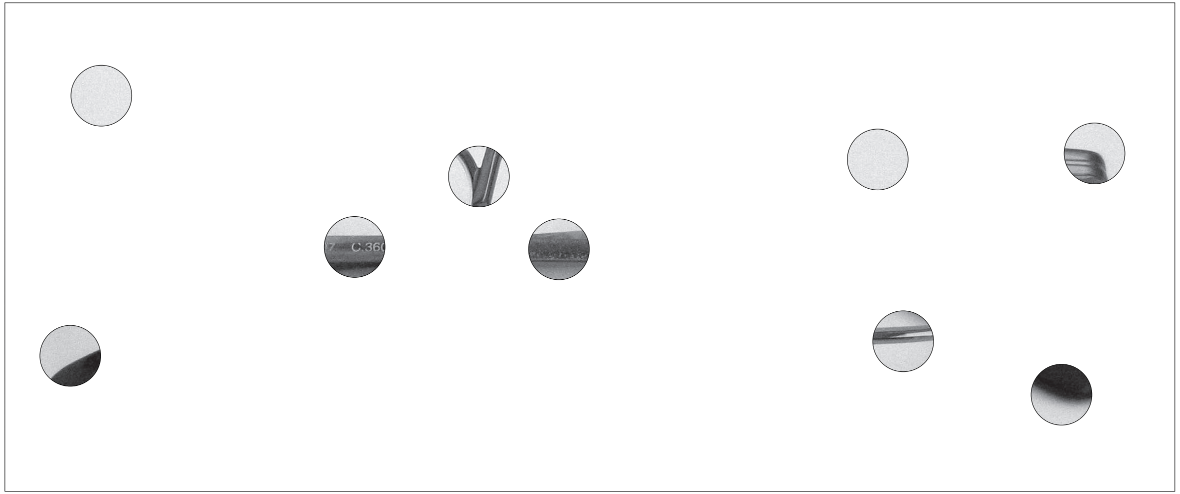
Inherent presence of noise as each real world measurement is corrupted by noise.

A lot of data Sheet A4, 300 dpi, 8 bit per pixel = 8.5 Mbytes.
Non-interlaced 4k video 3840×2160, RGB (24 bit) $\approx$ 9.96 Gbits/second.

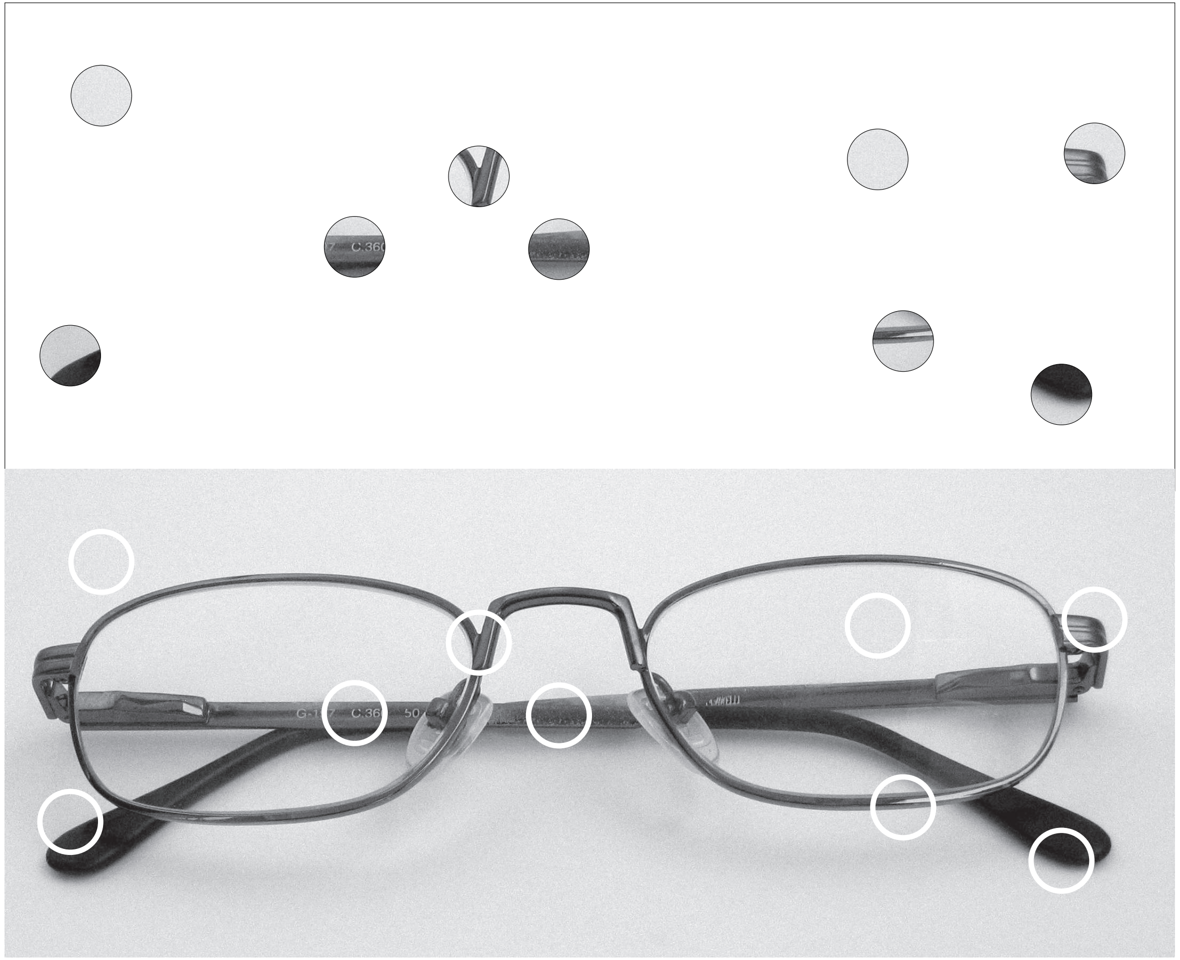
Interpretation needed (to be discussed soon).

Local window vs. the need for a global view.

Insufficiency of local view, illustration



Insufficiency of local view, illustration



Interpretation and its role, semantics

A human can embed percepts into a context known to her/him. This ability is formalized by the concept interpretation expressed mathematically as a mapping.

Interpretation: Observation $\rightarrow$ logical model *or* syntax $\rightarrow$ semantics

Examples:

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

Theoretical background: mathematical logic, theory of formal languages.

A deep philosophical problem: Gödel's incompleteness theorems ($\approx$1930); David Hilbert's program to find a complete and consistent set of axioms for all mathematics is impossible. More precisely, no consistent system of axioms whose theorems can be listed by an effective procedure (i.e., an algorithm) is capable of proving all truths about the arithmetic of the natural numbers.

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; independent on a specific application area.
- Signal processing methods are used, e.g., the 2D Fourier transform.

Middle level of knowledge ≈ image analysis

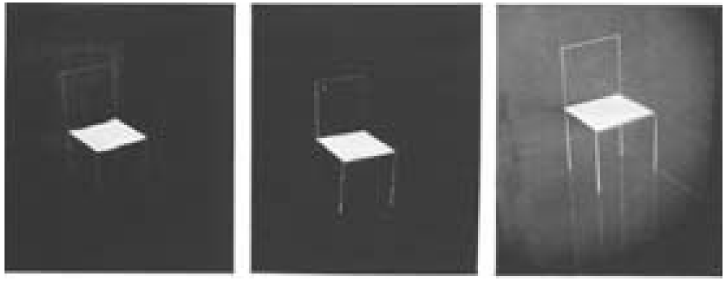
- Often 2D images only, e.g. cell images in an optical microscope.
- Interpretation explores an important additional knowledge allowing to solve tasks unsolvable otherwise.

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 and a feed back are explored; based on artificial intelligence methods.
- Goals are overambitious. Involved tasks are underconstraint have to be simplified radically.

Role of the prior knowledge, counterexample

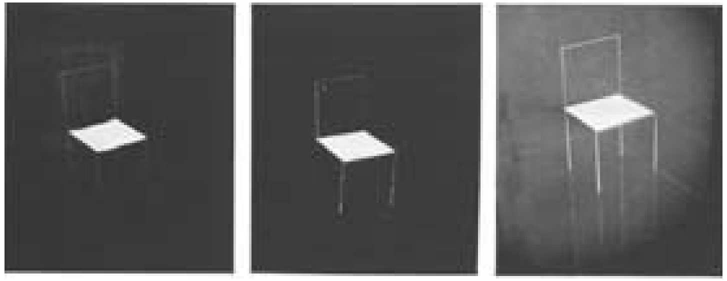
- Prior knowledge about "our world" enables humans to understand multi-meaning images.
- Of course, prior assumptions can mislead the human too . . .
- Counterexample: Ames chair.



We can see chairs.

Role of the apriori knowledge, counterexample

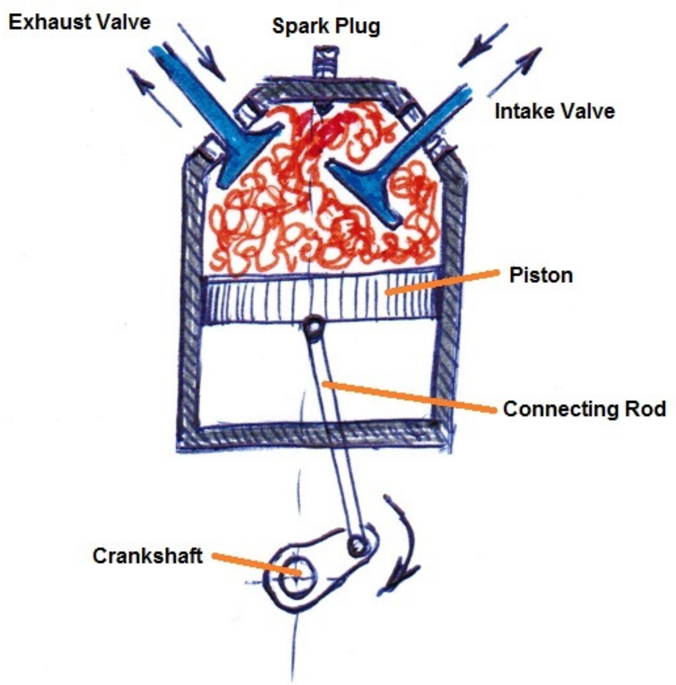
- Apriori knowledge about "our world" enables humans to understand multi-meaning images.
- Of course, apriori assumptions can mislead the human too . . .
- Counterexample: Ames chair.



We can see chairs.

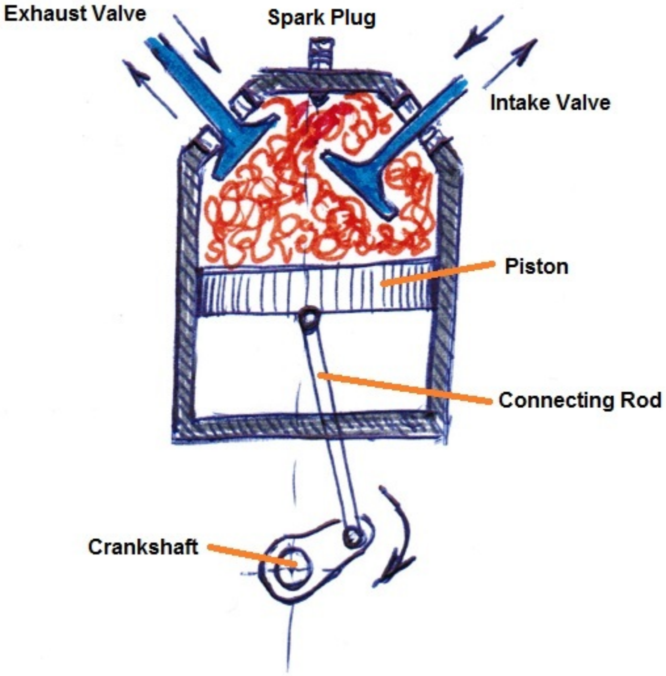
Actually, there are no chairs.

Is computer vision a typical engineering problem?



The idea is simple . . .

Is computer vision a typical engineering problem?



The idea is simple . . .

However, the idea needs only a refinement . . .

The ultra brief history of computer vision

1966 Seymour Papert assigns summer vision project to students.

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

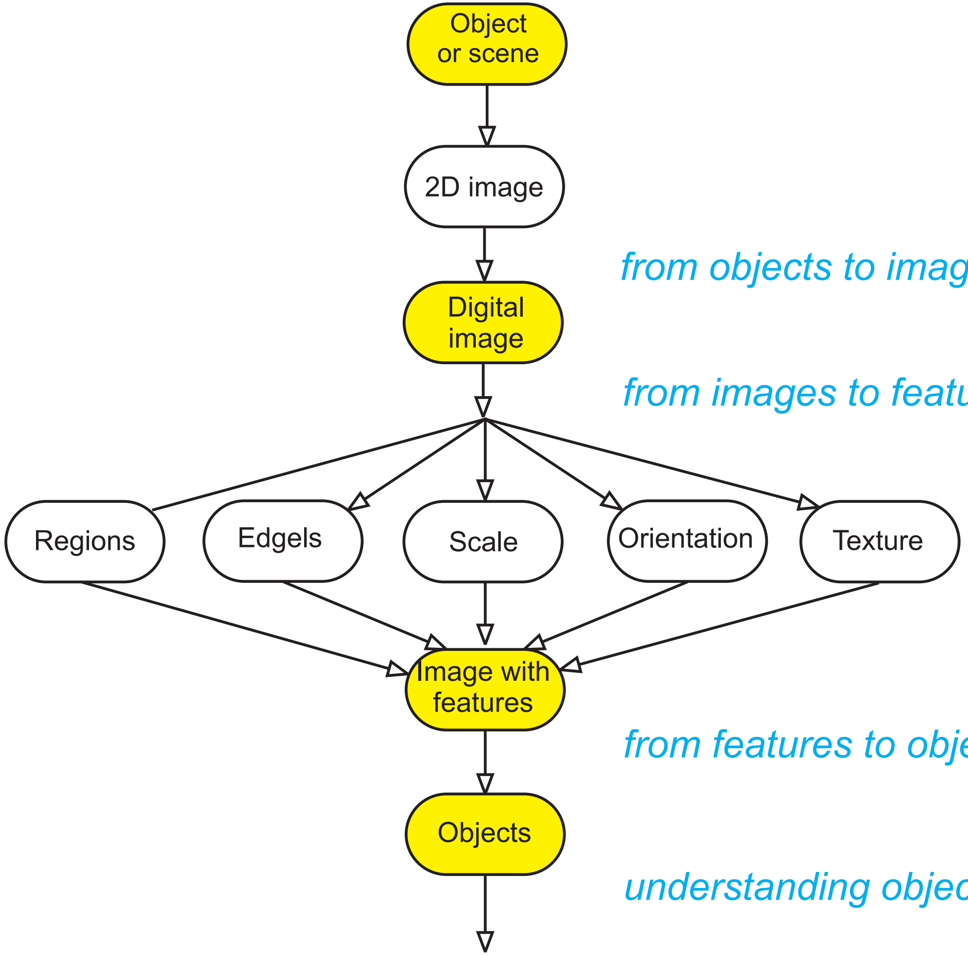
~**1970s** Some progress on interpreting selected images.

~**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