Introduction to 3D computer vision from the philosophical and computational viewpoints

Václav Hlaváč

Czech Technical University in Prague

Czech Institute of Informatics, Robotics and Cybernetics

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

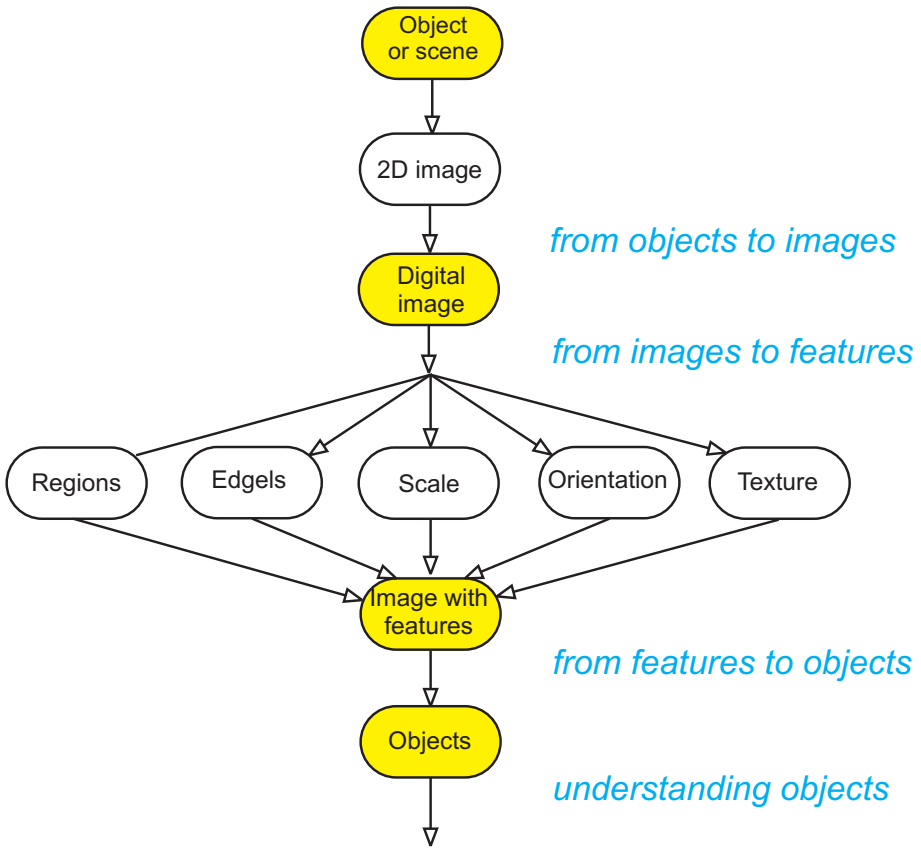
http://people.ciirc.cvut.cz/hlavac, vaclav.hlavac@cvut.cz

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

Traditional approach to 2D image processing/understanding

The following sequence of operations is usually performed in image understanding tasks:

- Image capturing (and digitization).
- Image preprocessing.
- Detecting objects, i.e. candidates for the approximate object location. Often false positives are allowed, because the next step (segmentation) filters them out.
- Object segmentation, i.e. finding regions (contiguous pixels) belonging to the object. (previous and this step can be performed together in one step in some methods, e.g. in the popular grab-cut).
- Objects description.
- Objects recognition, often using statistical pattern recognition (=machine learning) methods.



The methodology behind

- The general system theory provides a general framework, which allows us to treat the understanding of complex phenomena using the machinery of mathematics.
- The inherent complexity of the vision task is solved here by distinguishing the object (everything of interest or system or phenomenon) from the background.
- The objects and their properties need to be characterized. A formal mathematical model with a relatively small number of parameters is typically used for this abstraction.
- Model parameters are typically estimated from the (image) data.
- This methodology allows describing the same object using qualitatively different models (e.g. algebraic or differential equations) when varying resolution is used during observation. Studying changes of models with respect to several resolutions may give deeper insight into the phenomena of interest.

3D vision, motivation ≈ seek for principles

We seek principles related to 3D vision, which might enable us:

- to understand vision of living organisms,
- to equip machines with visual capabilities.

Types of related questions:

- *Empirical questions* (what is?) determine how existing biological visual systems are designed.
- *Normative questions* (what should be?) deal with classes of animals or robots that would be desirable.
- *Theoretical questions* (what could be?) speculate about mechanisms that could exist in intelligent visual systems.

Computer vision, three intertwined tasks

Vision as an information processing task.

1. Feature observability in images: We need to determine whether the task-relevant knowledge will be present in the primary image data.
2. Representation: This problem is related to the model choice of the observed world, usually performed at various levels of the interpretation complexity.
3. Interpretation: This problem tackles the semantics of the data. In other words, how are the observations mapped to the logical model approximating a subset of a real world.

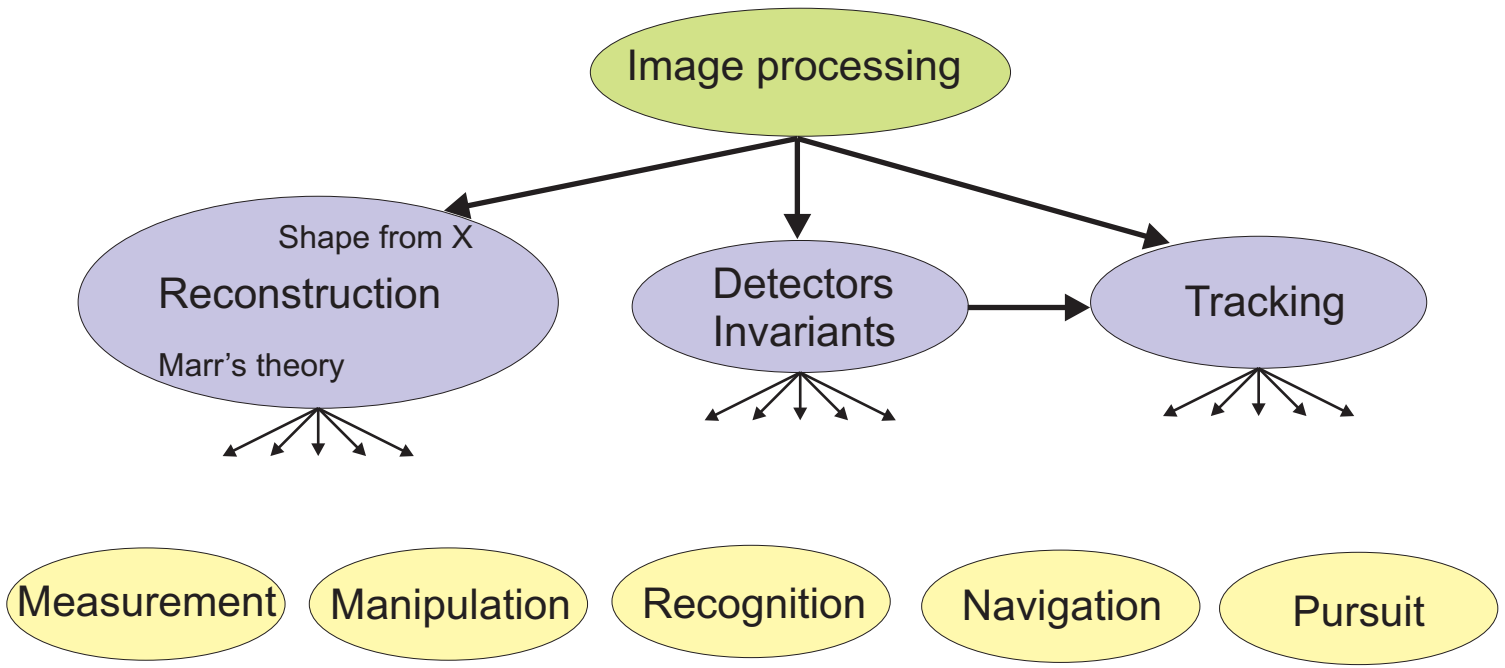
   The task is to make certain information explicit from a mathematical model storing it in an implicit form.

Two main approaches to artificial vision

The taxonomy is created according to the flow of information and the amount of a priori knowledge available/explored:

1. Reconstruction, bottom-up: The aim is, e.g., to reconstruct the 3D shape of the object from an image or set of images, which might be either intensity or range images.
   - One extreme here is given by Marr's theory, which is strictly bottom-up with very little a priori knowledge about the objects needed.
   - Some more practical approaches aim at creating a 3D model from real objects.
2. Recognition, top-down, model-based vision: The a priori knowledge about the objects is expressed by means of the models of the objects, where 3D models are of particular interest.

Structure of 3D vision



- D. Marr thought in 1970s that all processing should involve 3D reconstruction.
- However, it can be seen that there are other tasks too.

Situation in 1970s

- 1950s and 1960s brought many important discoveries in many different fields. AI, neuroscience, psychophysics.
- But nothing similar happened after 1970s.
- No neurophysiologists had recorded new and clear high-level correlates of perception.
- The leaders of 60s have turned away from what they had been doing.
  - David Hubel's (1923-2013) and Torsten Wiesel's (1924-) breakthrough discoveries from 1960s about the visual system and visual processing earned them the Nobel Prize for Physiology or Medicine in 1981. The have focused on anatomy since 1970s.
  - Horace Barlow (1921-2020) who discovered in 1950s that frog's retina not only detects and also predicts the position of passing fly. It opened research establishing that neurons estimate movement, color, position, and orientation of objects in visual world. In 1970s, he turned to psychophysics.

Courtesy: Allen Houng, https://www.powershow.com/view1/f5bfd-ZDc1Z/Marr_powerpoint_ppt_presentation

Neurophysiologists in a similar situation

- C.G. Gross told us in 1969 that hand-detectors exist, but not why and how they do.

  C.G. Gross, D.B. Bender, C.E. Rocha-Miranda: Visual Receptive Fields of Neurons in Inferotemporal Cortex of the Monkey. Science Vol. 166, Issue 3910, pp. 1303-1306, December 1969.

- Single-unit recording tells us nothing about how to program them on a computer.
- Neurophysiology and psychophysics can describe the behavior of cells or of subjects but not to explain such behavior.

Questions left out

- What are the visual area of the cerebral cortex actually doing?
- What are the problems in doing it that need explaining?
- At what level of description should such explanation sought?

David Marr's vision theory

Vision is based on the following processes:

- extracting visual information form the iconic representation,
- organizing the visual information,
- transforming the visual information into the explicit form for subsequent processing.

Marr's theory (David Marr, *1945 – †1980, died of leukemia)

- The primal sketch corresponds to edgels in the image.
- Partial findings are grouped using Gestalt principles.
- Book.

  David Marr: Vision: A computational investigation into the human representation and processing of visual information. MIT Press 1982.

Information processing task according to D. Marr

0. Task formulation.
   - *Inspiration from biology and neurophysiology.*
1. Computational analysis.
   - *Mathematical formulation of the achievable goal;*
   - *finding the solution and its proof;*
   - *identifying general constraint.*
2. Algorithms.
   - *Algorithmic realization of the theory from the item 1.*
3. Hardware tools.

Table 5.1 The three levels at which any machine carrying out an information-processing task must be understood

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?

Information processing task (2); Flying analogy

Example: Flying objects heavier than the air.

0. Task formulation – create the aircraft heavier than the air.
1. Computational theory – aerodynamics.
2. Algorithm – either waving with wings or the propeller.
3. Hardware – wooden propeller, piston combustion engine, linen wings.

Note: The advantage of the approach is a quite natural separability of individual steps.

David Marr, task formulation

The task is extracting the object shape from the information borne by edgels in a static monochromatic image.

Two levels of visual information processing

1. Preprocessing – context-free, preattentive, no semantics available, i.e. no interpretation possible..
2. High-level processing – explores context, is attentive, generates and tests hypotheses.

Two main conclusions from D. Marr's analysis

1. The visual system provides a 3D representation serving as input to recognition and classification (shape, spatial relations).
2. 3D representation is on an object-centered rather than viewer-centered frame of reference.

The procedure is bottom-up purely, which reduces the computational complexity significantly.

Towards observer-independent representation

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

Zero crossings

- Information at retina, iconic image, light intensity values at each point.
- Changes in intensity value provide clues as to surface boundaries. E.g., zero crossing of the second derivative of the intensity function.
- D. Marr proposed a Laplacian/Gaussian filter to detect zero crossings.

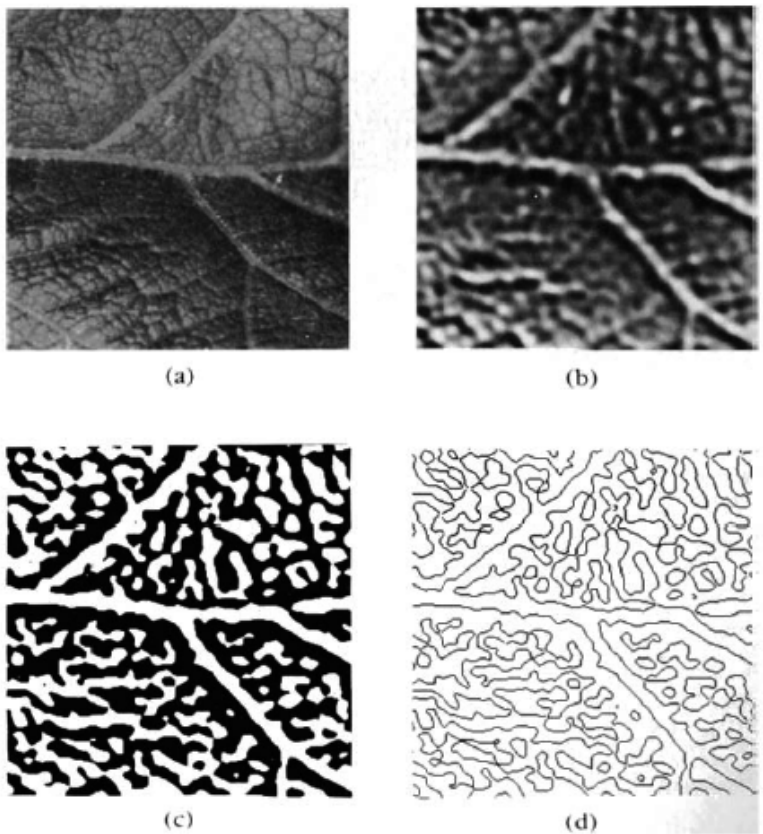


Figure 2–14.

Primal sketch

- Identifies intensity changes in the 2D image;
- Provides basic information about the geometric organization of those intensity changes;
- Primitives include: zero-crossings, virtual lines groups.

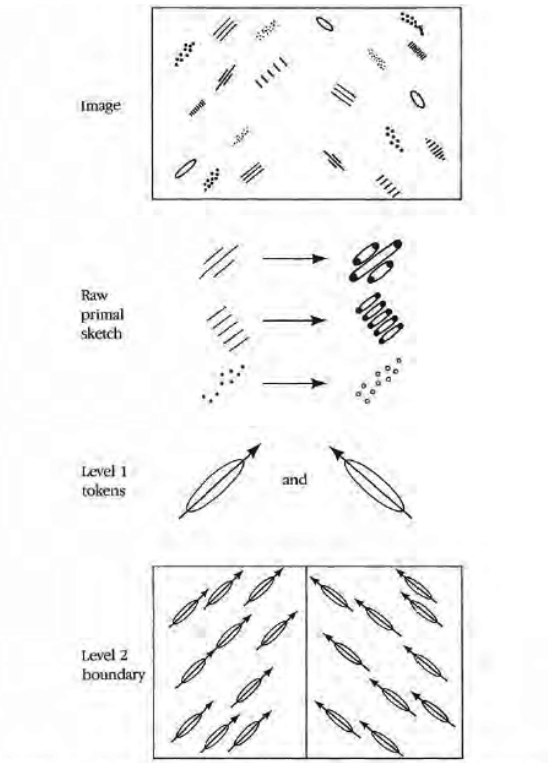
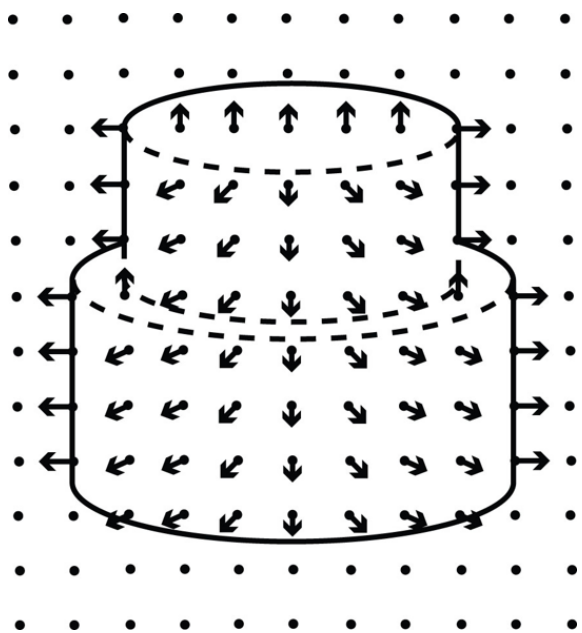


Figure 2–7. A diagrammatic representation of the descriptions of an image at different scales which together constitute the primal sketch. At the lowest level, the raw primal sketch faithfully follows the intensity changes and also represents terminations, denoted here by filled circles. At the next 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.

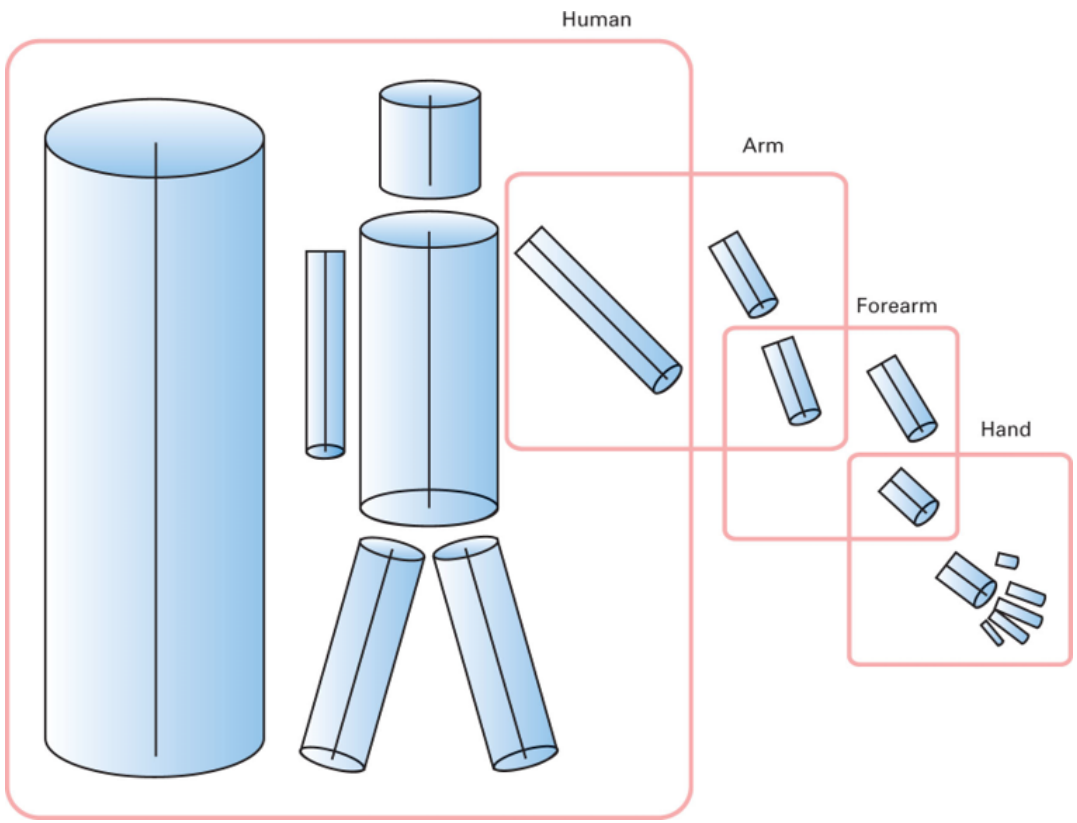
2.5D sketch

- 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 which facilitates recognition.



Marr's bottom-up approach summarized