DD2423 Image Analysis and Computer Vision

Mårten Björkman

Computational Vision and Active Perception
School of Computer Science and Communication

Novermber 8, 2013

General course information

- 7.5 hp course (labs 4.0 hp, exam 3.5 hp)
- Course Web: https://www.kth.se/social/course/DD2423/
- 2-3 lectures a week
- 16 lectures in total (3 exercise sessions)
- TAs: Cheng, Magnus, Virgile, Ali, Püren, Martin and others.
- If you have questions: preferably use the Course Web.

Assessment

- 3 labs (LAB1) and exam (TEN1)
- Grading:
  - Final grade: Average of exam and labs, rounded towards exam.
  - Exam: A-F, Labs: A-F (average of Lab2 and Lab3 rounded downwards)
- Labs are done in Matlab, possibly on your own laptop.
- There are scheduled times in computer halls in the main CSC building:
  - Help: get help while working on labs
  - Redovisning (examination): book a slot though web - no help!
- **Do not use only these to work on the labs!**
- Doing labs before the deadline - up to 5 points on the exam (of 50 total)

Grading of labs

- Lab1 is not graded and can be done and presented in **pairs**.
- Lab2 and Lab3 can be done in **pairs** but are examined **individually**.
- A cumulative definition of grades:

  E - Lab is completed, but many written questions not correctly answered.

  D - Some written questions have not been answered correctly.

  C - Minor difficulties in presenting results and responding to oral questions.

  B - No difficulties in presenting results and responding to oral questions.

  A - Is able to reason about questions beyond the scope of the lab.

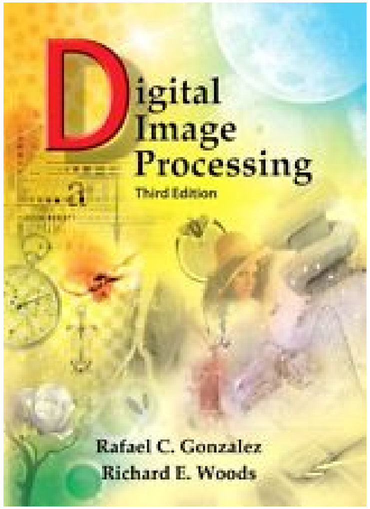
- More detailed definitions on the web page.
- Good idea: Present to each others!

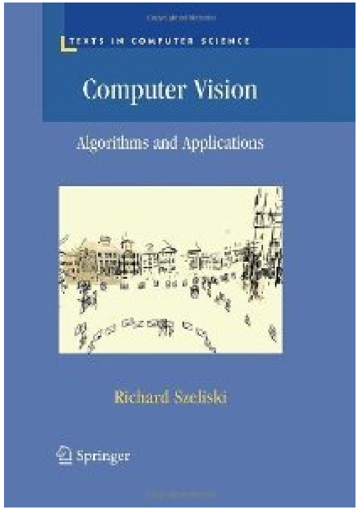
Important!

- Lab1 is easy, Lab2 and Lab3 require time
- RAPP system for administration
- Requirements:
  - Go through the lab instructions.
  - Implement the required functions.
  - **Answer the questions in the instruction file.**
  - **You will not be examined, if you did not write the answers to questions in the instructions!**
  - Make sure that the TA signs the printed instructions (the last page).
- Start to work on labs as soon as possible!

Course books

- R. Gonzalez and R. Woods: “Digital Image Processing”, Prentice Hall, 2008.
- R. Szeliski: “Computer Vision: Algorithms and Applications”, Springer, 2010. (available for free: http://szeliski.org/Book)





- Note: course books are used to help understanding, while assessment is based only on lecture and lab notes.

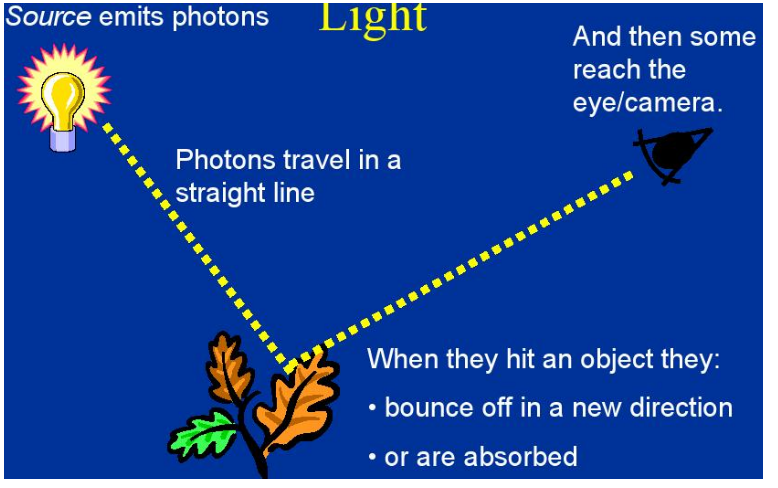
DD2423 Image Analysis and Computer Vision

INTRODUCTION AND OVERVIEW

Mårten Björkman

November 8, 2013

What does it mean to see?



- Vision is an active process for deriving efficient symbolic representations of the world from the light reflected from it.
- Computer vision: Computational models and algorithms that can be used for solving visual tasks and for interacting with the world.

Vision is an active process!

- Active:
  - In nature seeing is always (?) associated with acting.
  - Acting can simplify seeing, e.g. move your head around an object.
  - A computer vision system may control its sensory parameters, e.g. viewing direction, focus and zoom.
- Process:
  - No “final solution”. Perception is a result of continuous hypothesis generation and verification.
  - Vision is not performed in isolation, it is related to task and behaviors.
  - Attention is essential for low complexity.

Computer vision

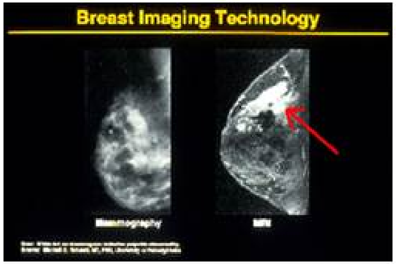
- Digital imaging has become cheap.
- Open-source software available: allows to build on and develop algorithms more easily (e.g. OpenCV)
- Machine learning has become more important.
- Applications:
  - Robotics: vision based control, object detection and recognition
  - Industry: material inspection, packaging
  - Image processing of satellite images; surveillance; teleoperation
  - Analysis and visualization of medical data
  - Internet: image search, classification and sorting.

Why is vision interesting?

- Intellectually interesting
  - How do we figure out what objects are and where they are?
  - Harder to go from 2D to 3D (vision), than from 3D to 2D (graphics).
- Psychology:
  - $\sim$ 50% of cerebral cortex is for vision.
  - Vision is (to a large extent) how we experience the world.
- Engineering:
  - Intelligent machines that interact with the environment.
  - Computer vision opens up for multi-disciplinary work.
  - Digital images are everywhere.

Why is vision relevant?

Safety



Health

Security

Comfort

Fun

Access

There are many applications where vision is the only good solution.

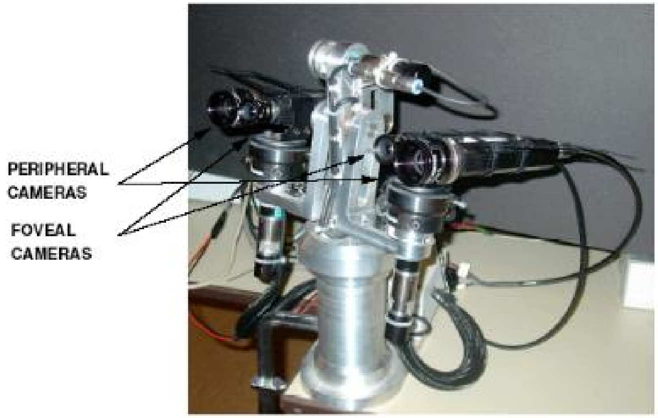
Examples

- Add object into video sequence; detect ground plane and camera position.

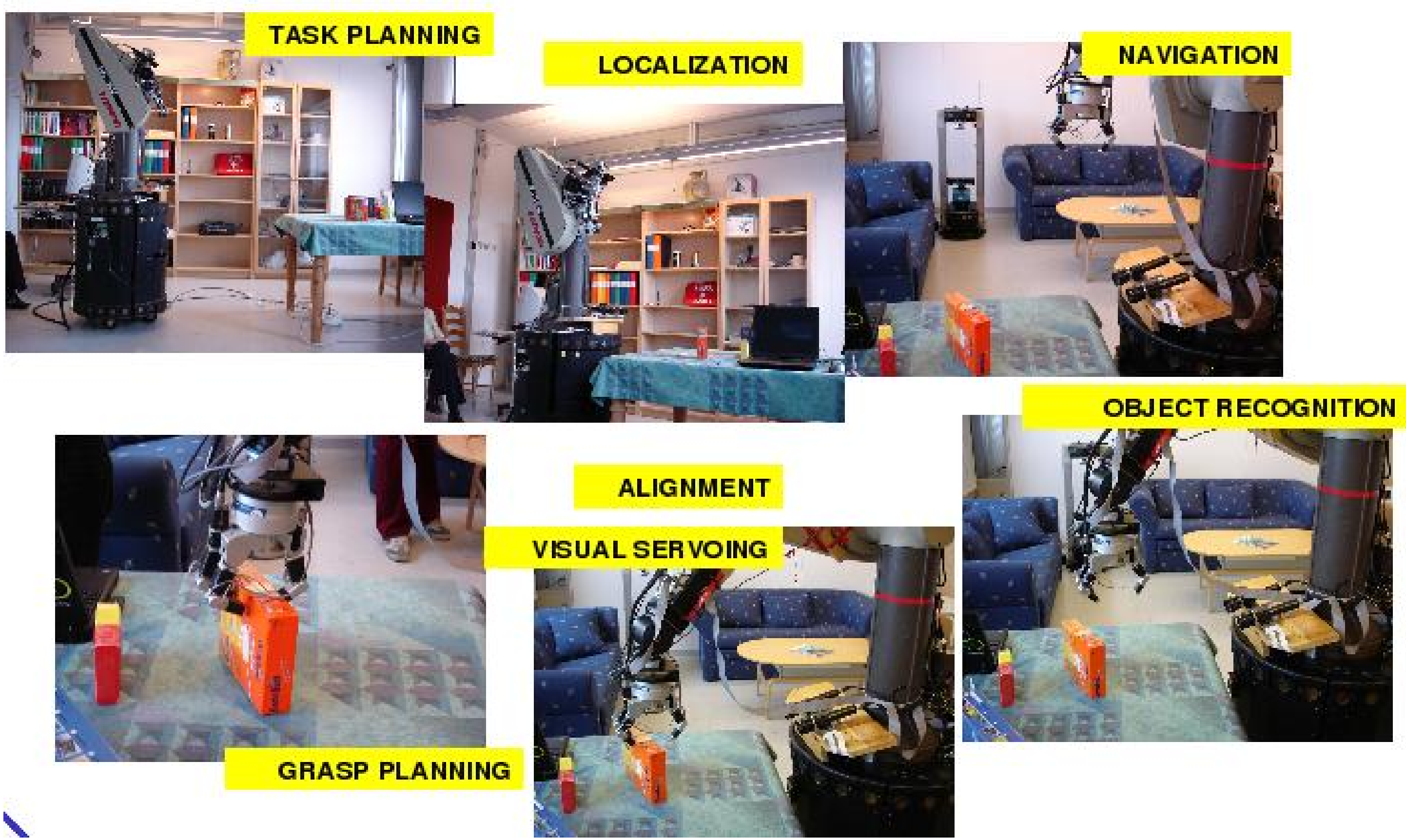
- Detect and track football players, collect statistics.

Examples

- Find, recognize and pick up objects in using a stereo head.



Example - A robot task



Approaches

- Mathematical modeling - image oriented
  - image formation (geometry, illumination)
  - information content
- Experimental evaluation - task oriented
  - development of algorithms
  - experimental setup
- Relations to biological vision (human, mammals)
  - neuroscience
  - psychophysics

Multi-disciplinarity

- Neuroscience / Cognition: how do animals do it?
- Philosophy: why do we consider something an object? (Hard!)
- Physics: how does an image become an image?
- Geometry: how does things look under different orientations?
- Signal processing: how do you work on images?
- Probability / Statistics: deal with noise, develop appropriate models.
- Numerical methods / Scientific computing: do this efficiently.
- Machine learning / AI: how to draw conclusions from lots of data?

What can be done?

- Improve image quality
- Recognize things (objects, people)
- Locate objects in space (partial or full pose estimation)
- Track features and things (objects, people)
- Reconstruction from 2D images to 3D scenes.
- History:

  - 1960: Image enhancement (2D signal processing), image coding, simple image analysis.

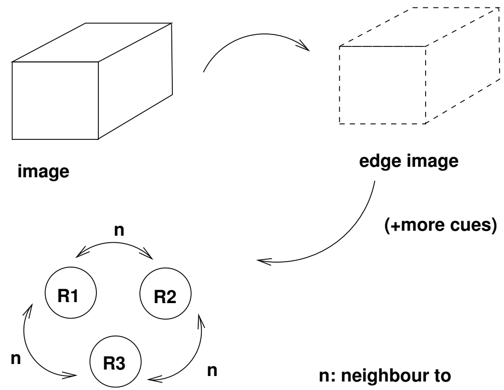
  - 1975: Deeper understanding of the meaning and degree of difficulty of the problem. Geometry and reconstruction.

  - 1985: Vision as an active goal-directed process.

  - 2000: Handle large amount of data. Machine learning.

Representation of image data

- Computer vision is about making certain aspects of the image data explicit.



- Different representations can be good for different purposes.

  Alt1: graph,   Alt2: 1 square, 2 parallelograms,

  Alt3: 9 lines,   Alt4: 1 block,   Alt5: the original pixels.

Observations

- The descriptions are of different types.
  - the edge image is iconic (image-like)
  - other representations are symbolic (*a cube*)
  - some expressed in 2D, some in 3D
- All information is not contained in an image alone.
  - going from 2D to 3D is an underdetermined inverse problem
  - it requires restrictions to be solved
  - we add what we already know about the world

Image processing $\Longleftrightarrow$ Signal processing

- The image is enhanced for easier interpretation.
- Different levels of processing (often used as pre-processing).

Purpose of image processing

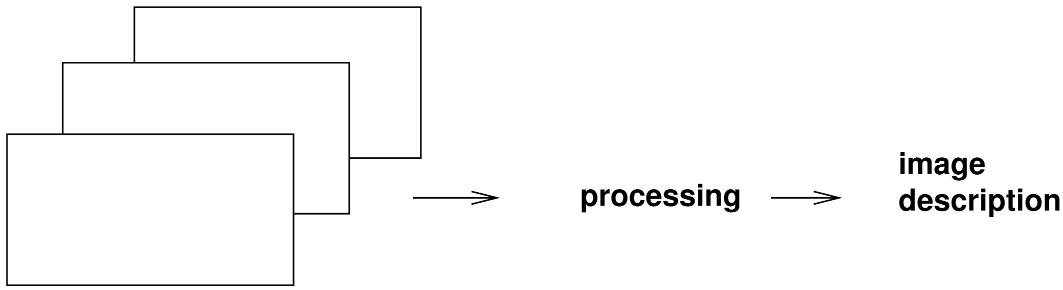
- Enhance important image structures
- Suppress disturbances (irrelevant info, noise)

Example: Poor image data in medicine, astronomy, surveillance.

Subjects treated in this course:

- Image sampling, digital geometry
- Enhancement: gray scale transformation (histogram equalization), spatial filtering (reconstruction), morphology
- Linear filter theory, the sampling theorem

Image analysis



- Purpose: Generate a useful description of the image
- Examples: Character recognition, fingerprint analysis

Subjects studied in this course:

- Feature detection
- Image descriptors
- Image segmentation
- Image recognition and classification

Recognition / Classification?

- Recognition: Is this my cup?
- Classification: Is this a cup?
- Detection: Is there a cup in the image?

Image feature detection → object classification

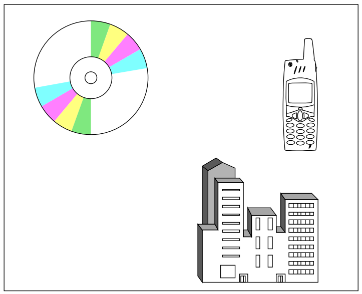
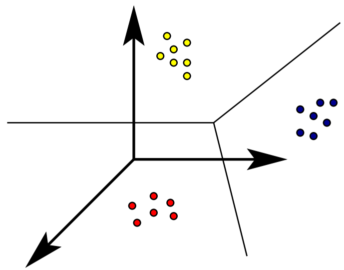
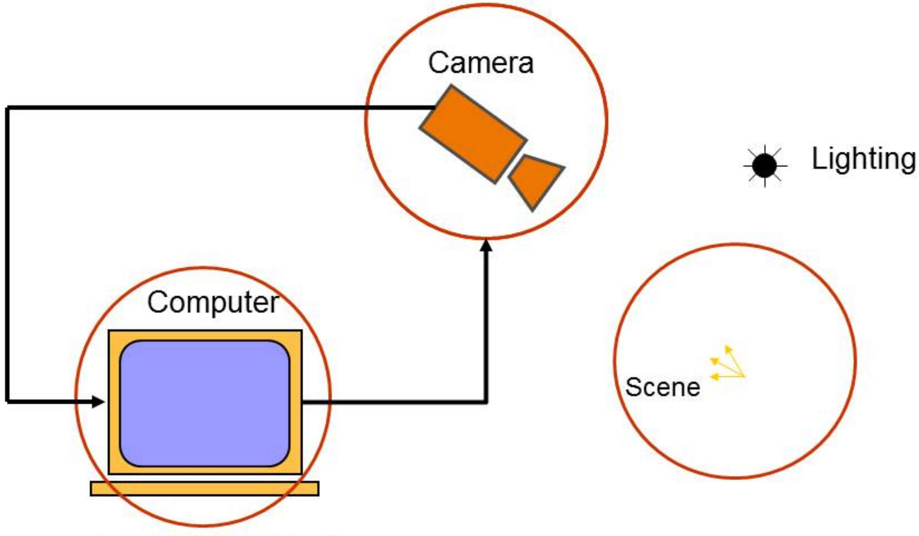


image domain



feature space

Computer vision



- Purpose: Achieve an understanding of the world, possibly under active control of the image acquisition process.
- Often people say computer vision, even if they mean image analysis.

Subjects treated in this course: stereo, motion, object recognition, etc.

From 3D world to 2D image and back

$<$ underdetermined 2D $\rightarrow$ 3D problem $>$

Main assumptions:

- The world we observe is constructed from coherent matter.
- We can therefore perceive it as constructed from smooth surfaces separated by discontinuities.

In human vision, this way of perceiving the world can be said to precede understanding.

- The importance of discontinuities: Under general assumptions, a discontinuity in image brightness may correspond to a discontinuity in
  - depth
  - surface orientation
  - surface structure
  - illumination

The importance of discontinuities

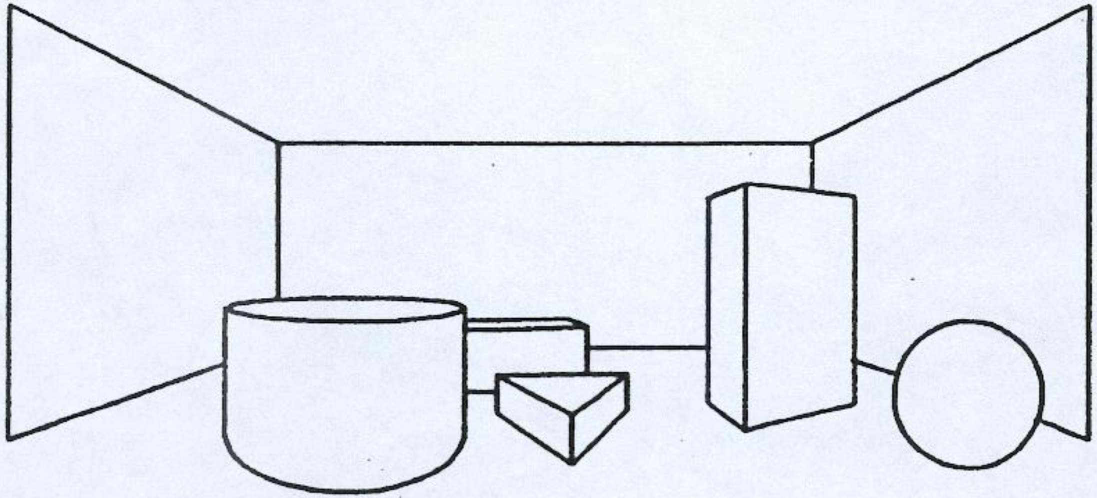


FIG. 1.14. This figure shows a two-dimensional edge-map representation of the imaginary scene from Fig. 1.1. Note that the perspective appears distorted because the image covers a wide horizontal angle (150 degrees); but of course, this diagram does not itself cover the same visual angle. The image is made up of elongated contours that are oriented at all angles, and contour intersections.

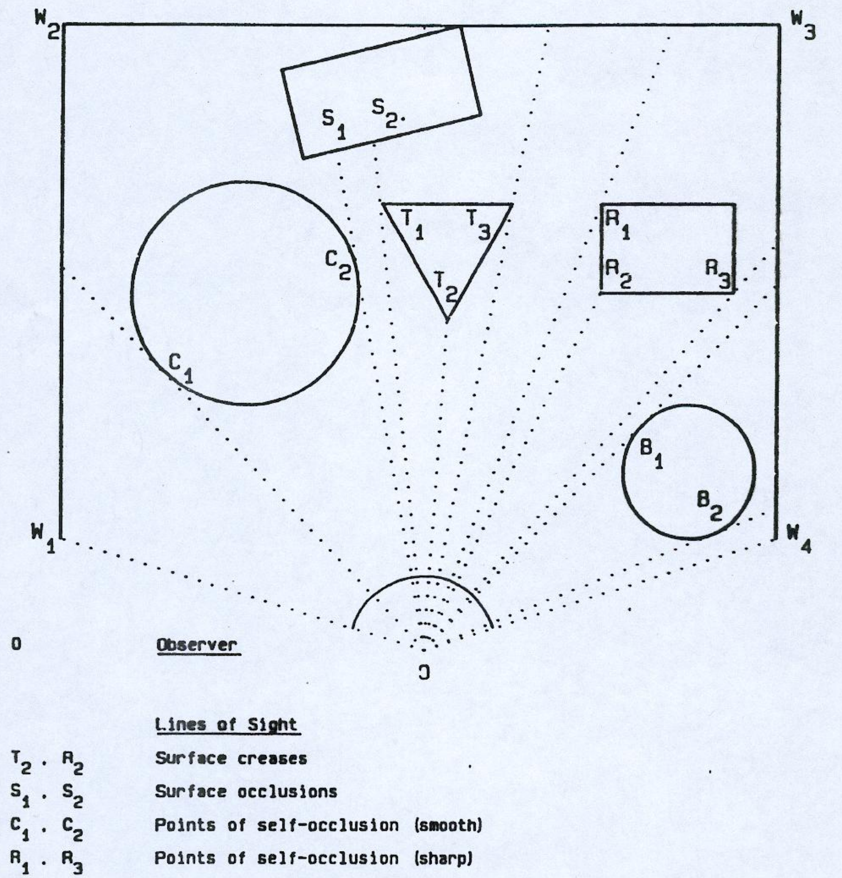


FIG. 1.1. A plan of an imaginary scene. The image formed by the observer at *O* will contain segments corresponding in sequence to the various visible surfaces. These segments are bounded by the lines-of-sight that are drawn from *O* to each point of surface occlusion or surface creasing.

Geometric discontinuities

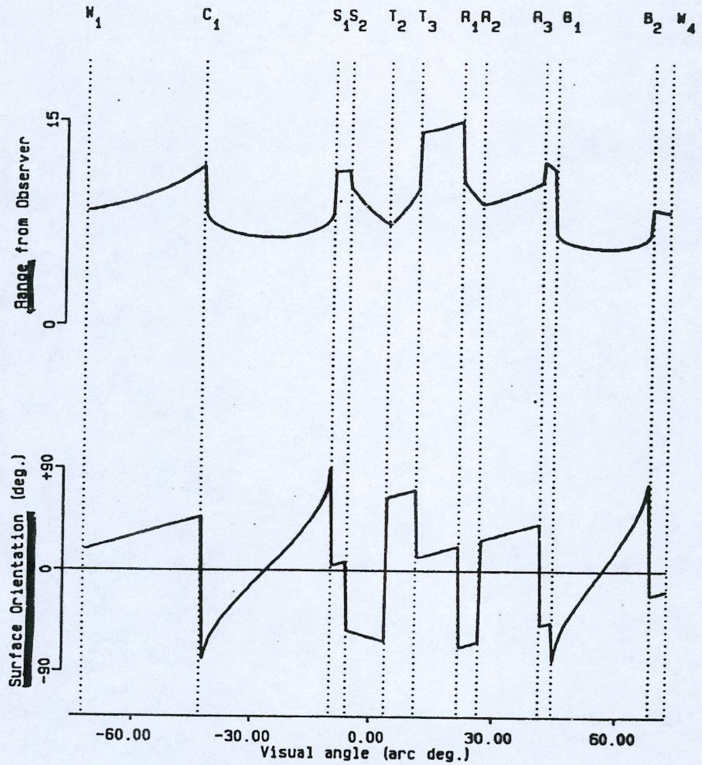


FIG. 1.2. A range map (top) and surface orientation map (bottom) for the scene of Fig. 1.1 from the point *O*. The different surface segments shown in that figure now correspond to areas where range and surface orientation change smoothly, and the segment boundaries are marked by sharp changes of direction and discontinuities in these maps.

Discontinuities: Ideal illumination

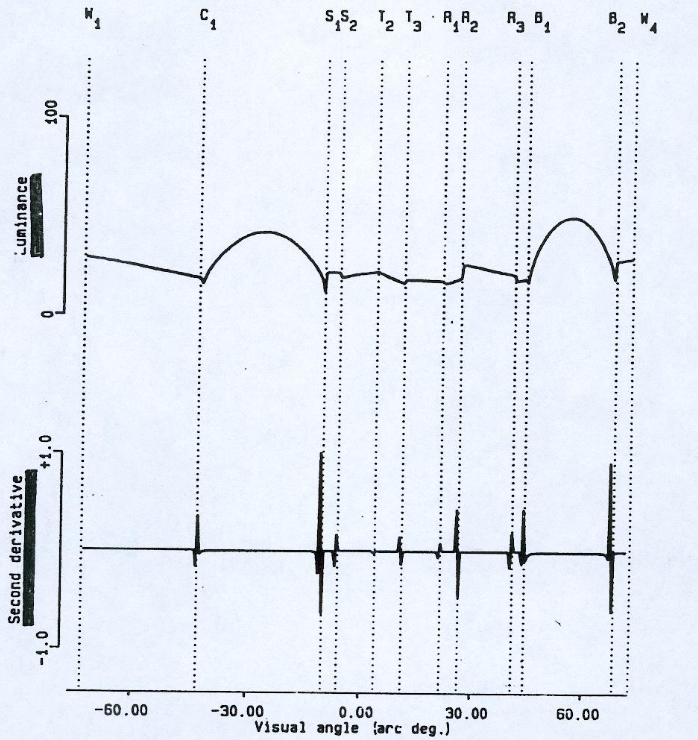


FIG. 1.3. The top of this figure shows the luminance profile of the image formed of the scene at the point *O*. The different surface segments shown in the figure now correspond to areas where luminance changes smoothly. The bottom of the figure shows the second derivative of the luminance profile. Surface creases and discontinuities are now marked by sharp peaks in this new signal.

Discontinuities: Texture + noise

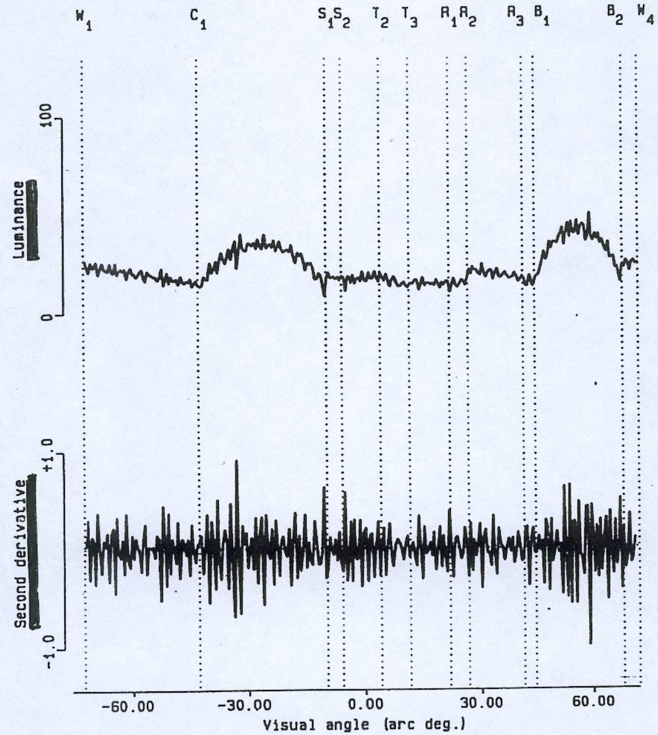
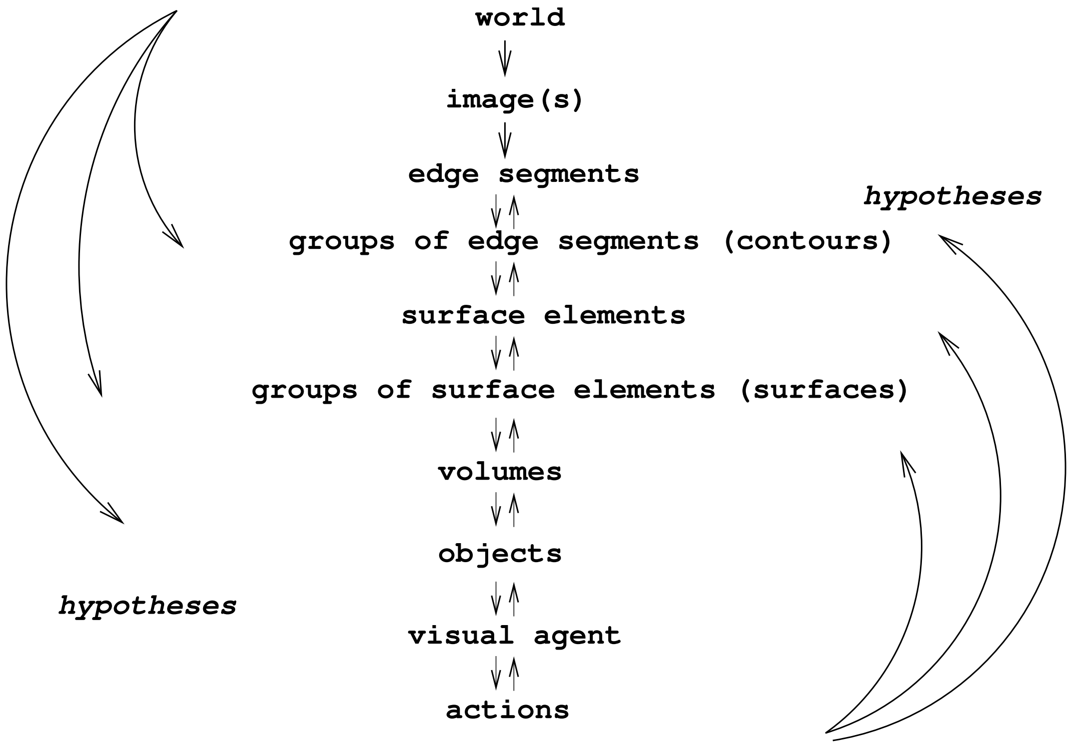


FIG. 1.7. This figure shows the luminance profile (top) and second derivative (bottom) of the scene image, as before, where all the surfaces are marked with a pattern of alternating light and dark texture elements. The points of special interest with respect to the original scene are now hidden in the second derivative signal by its response to the texture.

A simplified (traditional) model of a system



Edge labeling

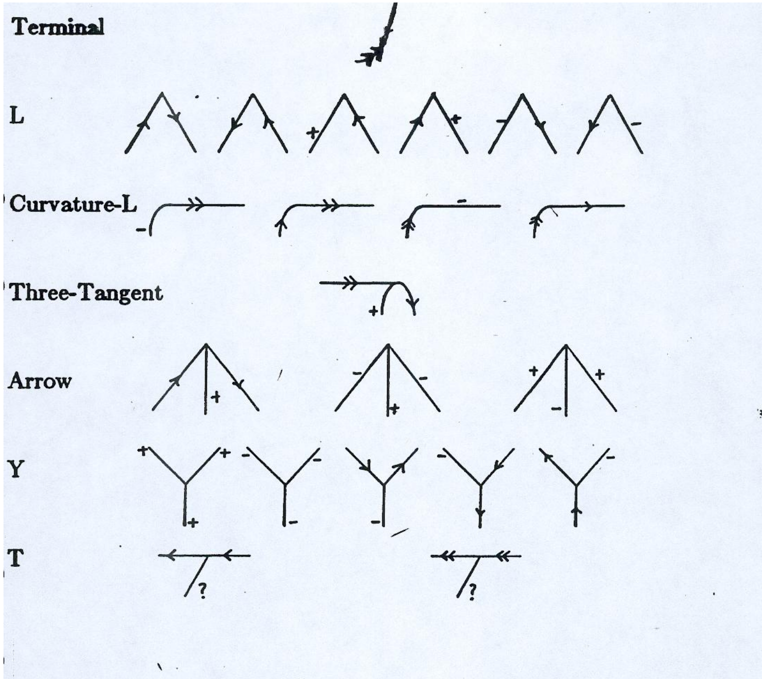


Figure 17: The Junction Catalog for Piecewise Smooth Surfaces

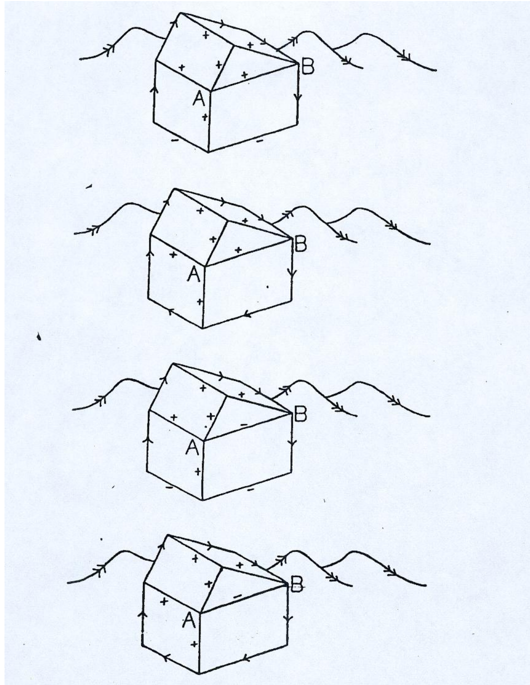
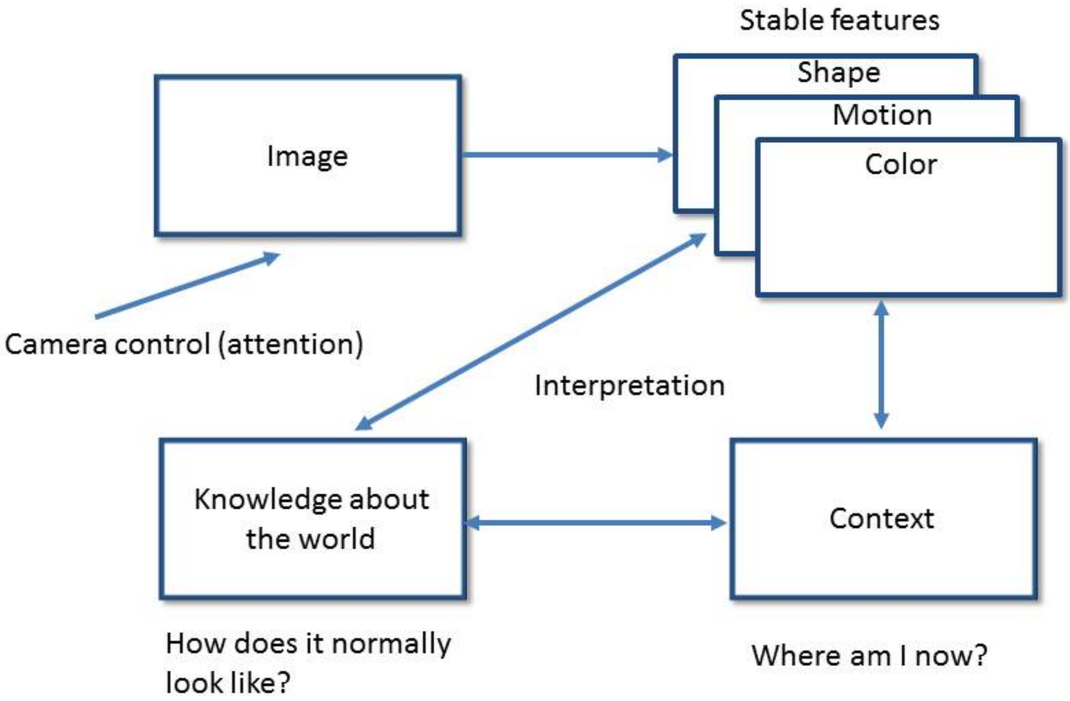


Fig. 31. Possible labellings of a curved object.

Turned out to be very hard. Many possible solutions. Research got stuck!

Components of a computer vision system



Everything is connected to everything. No straight sequence.

The three levels (David Marr 1982)

1. Computational theory:
   What is the goal of the computation? Why is it appropriate?
2. Representation and algorithm:
   How can this computational theory be implemented?
3. Hardware implementation:
   How can this representation and its algorithms be realized physically?

In human vision several computations seem to take place:

- primitive grouping processes
- primitive 3D shape interpolation (based on hypothesis about 3D)
- constancy phenomena size, shape, color

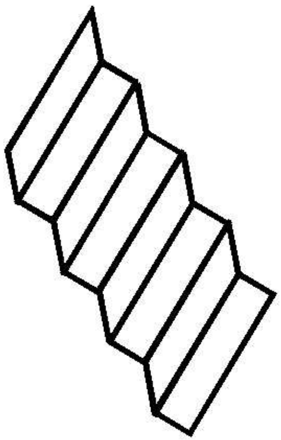
What is missing in Marr's theory?

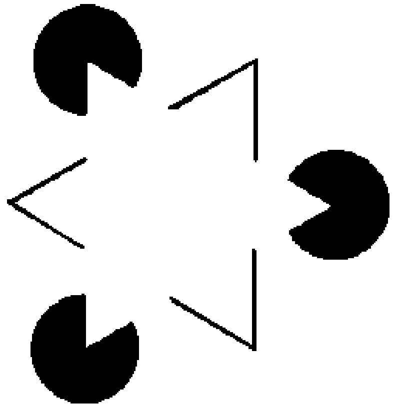
Although the models contain "top-down" and "bottom-up" processing there is no representation of

- goals and behaviors
- eye movements and visual feedback
- attention
- theories/models for integration of different types of information
- learning

Human vision is not perfect!

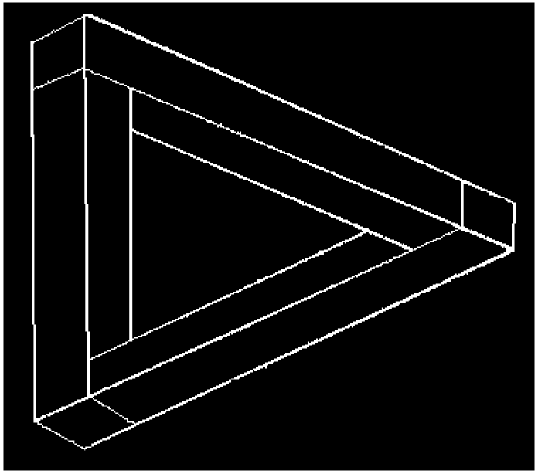
Reversing staircase illusion and subjective contours:

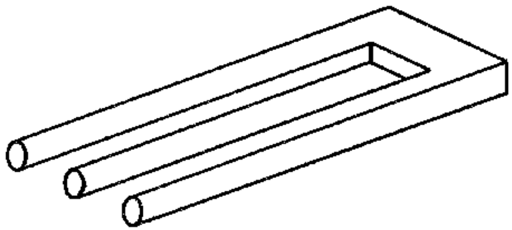




- Our perceptual organization process continues after providing a (first) interpretation. Continue viewing the reversing staircase illusion and you will see it flip into a second staircase.

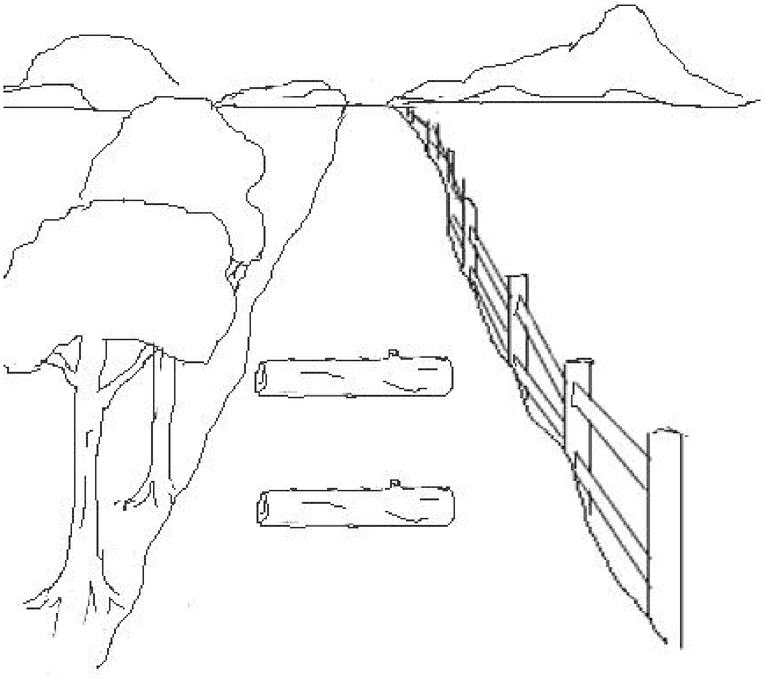
Impossible objects





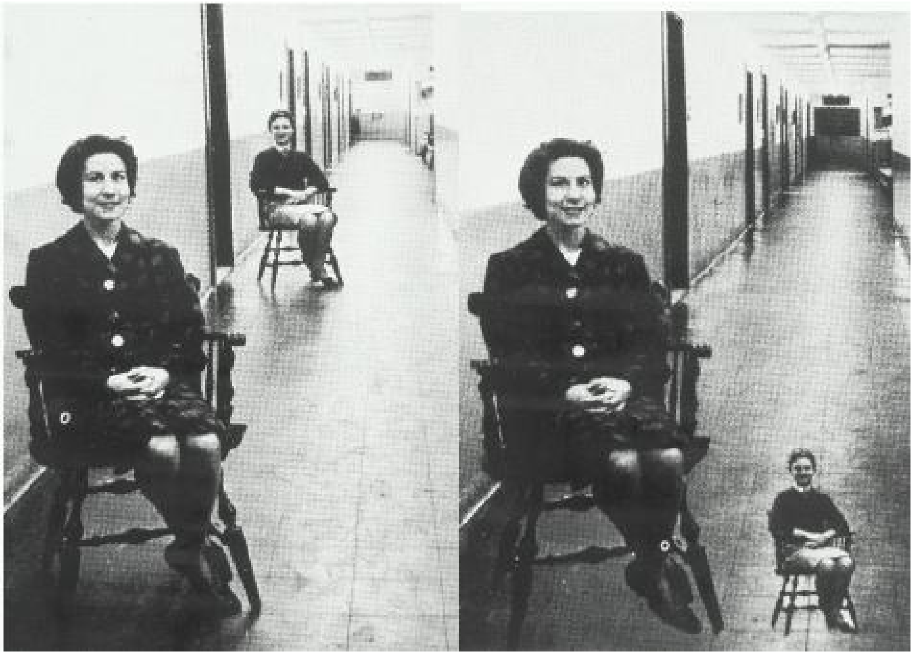
Another example that vision is an ongoing process.

Depth illusion - size constancy



We tend to “normalize” things, such as size, shape and colors.

Depth illusion - size constancy



Summary of good questions

- Why is vision an active process?
- What is computer vision good for?
- In what ways is computer vision multi-disciplinary?
- How to cope with the fact that it is an underdetermined inverse problem?
- What is image processing, image analysis and computer vision?
- Why are image edges so important in vision?
- What could a possible vision system consist of?

Recommended readings

- Gonzalez and Woods: Chapter 1
- Szeliski: Chapter 1
- Introduction to labs (on web page)