# Introduction

## DD2423 Image Analysis and Computer Vision

Mårten Björkman

Computational Vision and Active Perception
School of Computer Science and Communication

November 3, 2015

# 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, Püren, Nils, Ali, Hang, Yuquan, Bala 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.
  - Labs grade: A-F (average of labs rounded downwards)
- Labs are done in Matlab, possibly on your own laptop.
- There are scheduled times in computer halls:
  - 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 pts on exam (of 50 total)

# Grading of labs

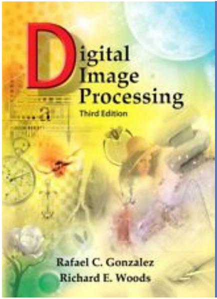
- All labs can be done in **pairs**, but examined **individually**.
- A cumulative definition of grades:
  - E - Lab completed, but many written answers not correct.
  - D - Some written questions have not been answered correctly.
  - C - Minor difficulties in presenting lab results and responding to oral questions posed by TAs.
  - B - No difficulties in presenting lab results and responding to oral questions posed by TAs.
  - 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!

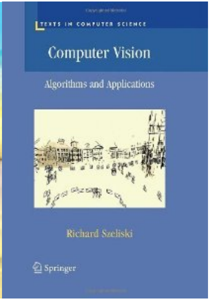
# Important!

- 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

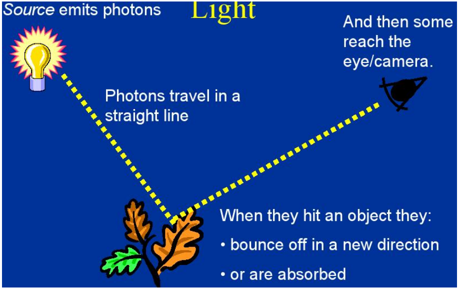
- R. Gonzalez and R. Woods: “Digital Image Processing”, Prentice Hall, 2008.
- R. Szeliski: “Computer Vision: Algorithms and Applications”, Springer, 2010. (available for free: http://szeliski.org/Book)




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

# 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 to solve visual tasks and interact with the world.

# Why is vision relevant?

Safety


Health

Security

Comfort

Fun

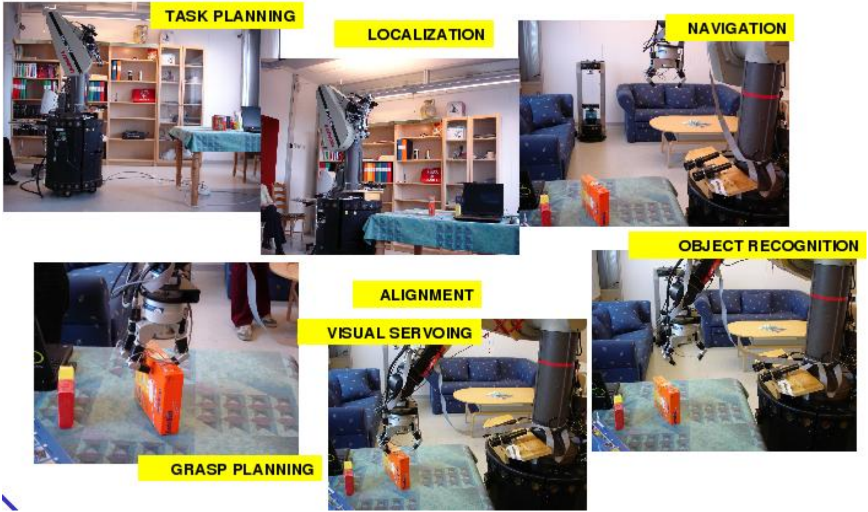

Access

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

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


- Detect and track football players, collect statistics.

# Example - A robot task



# 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.

# 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?

# Image processing $\Longleftrightarrow$ Signal processing

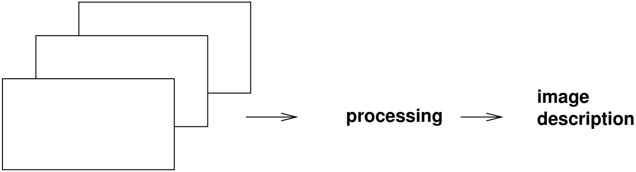
- The image is enhanced for easier interpretation.
- Different levels of processing (often used as pre-processing).

# Purpose of image processing

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




- 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 vs classification

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

Image feature detection $\rightarrow$ object classification

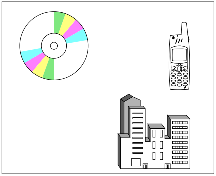
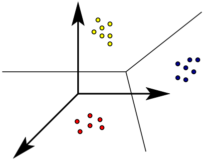
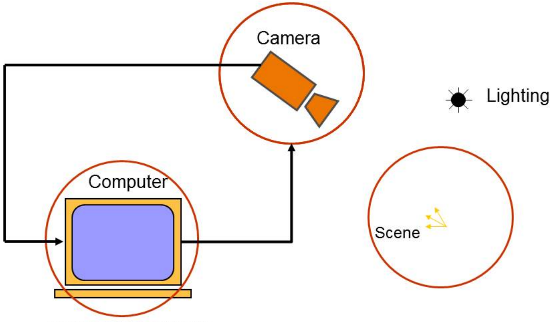


**image domain**



**feature space**

# Computer vision




- Purpose: Achieve an understanding of the world, possibly under active control of the image acquisition process.
- Examples: object tracking, activity recognition
- Often people say computer vision, instead of image analysis.
- Subjects in this course: stereo, motion, object recognition, etc.

$< \text{underdetermined 2D} \rightarrow \text{3D problem} >$

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: A discontinuity in image brightness may correspond to a discontinuity in either
  - depth
  - surface orientation
  - surface structure
  - illumination

# The importance of discontinuities

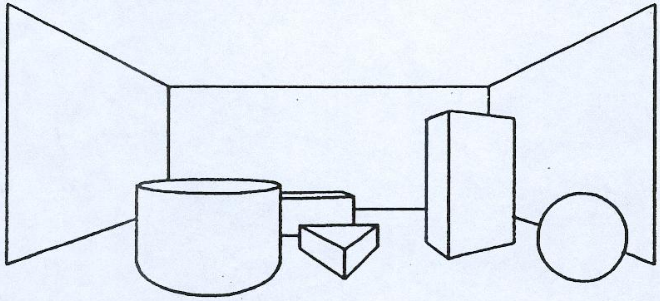


FIG. 1.14. This figure shows a two-dimensional edge-map representation of the imaginary scene from Fig. 1.1. Note that the perspective appears distorted because the image covers a wide horizontal angle (150 degrees); but of course, this diagram does not itself cover the same visual angle. The image is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

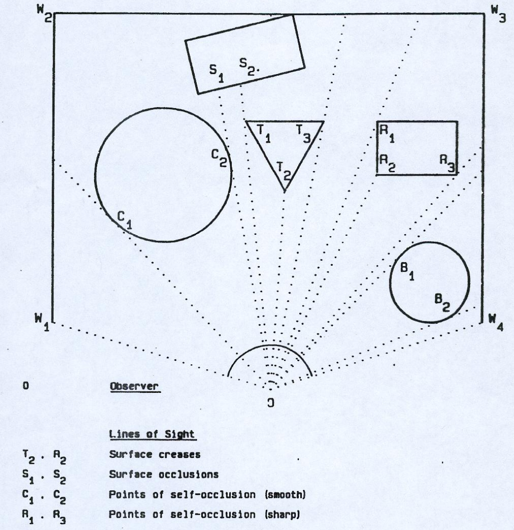



FIG. 1.1. A plan of an imaginary scene. The image formed by the observer at $O$ will contain segments corresponding in sequence to the various visible surfaces. These segments are bounded by the lines-of-sight that are drawn from $O$ to each point of surface occlusion or surface creasing.

# Geometric discontinuities

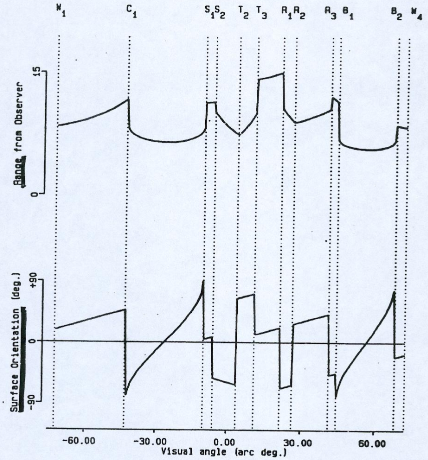



FIG. 1.2. A range map (top) and surface orientation map (bottom) for the scene of Fig. 1.1 from the point *O*. The different surface segments shown in that figure now correspond to areas where range and surface orientation change smoothly, and the segment boundaries are marked by sharp changes of direction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

# Discontinuities: Ideal illumination

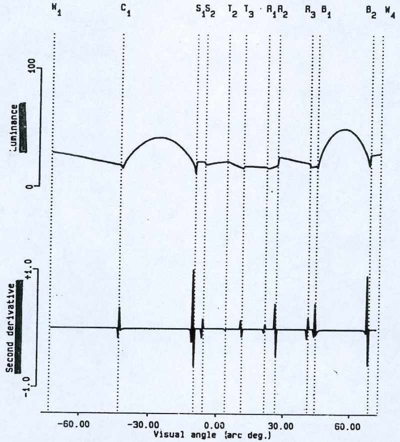



FIG. 1.3. The top of this figure shows the luminance profile of the image formed of the scene at the point *O*. The different surface segments shown in the figure now correspond to areas where luminance changes smoothly. The bottom of the figure shows the second derivative of the luminance profile. Surface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

# Discontinuities: Texture + noise

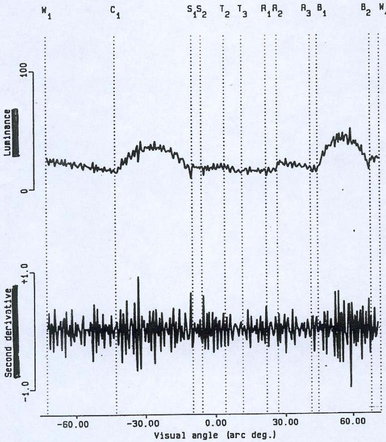
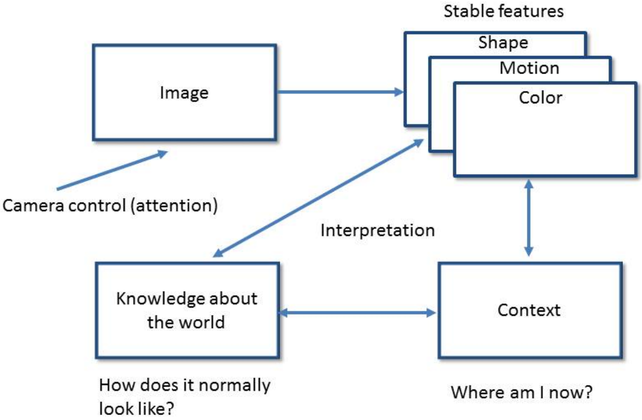



FIG. 1.7. This figure shows the luminance profile (top) and second derivative (bottom) of the scene image, as before, where all the surfaces are marked with a pattern of alternating light and dark texture elements. The points of special interest with respect to the original scene are now hidden in the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

# Components of a computer vision system



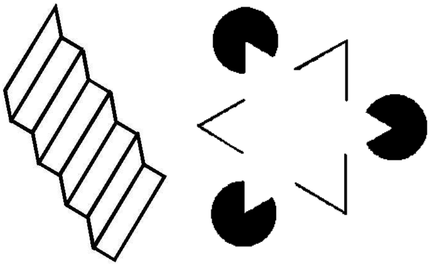

Everything is connected to everything. No straight sequence.

# 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.

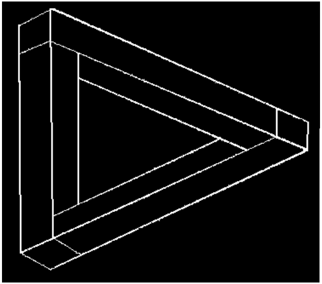
# Human vision is not perfect!

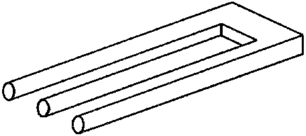
Reversing staircase illusion and subjective contours:



- 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.

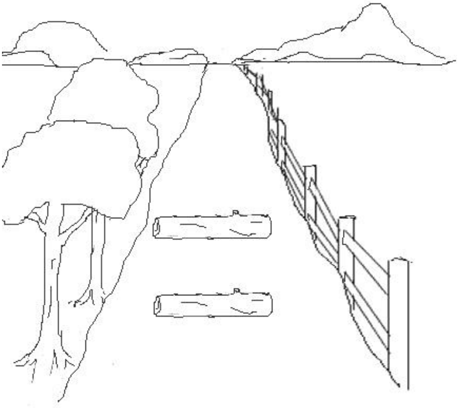
# Impossible objects





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

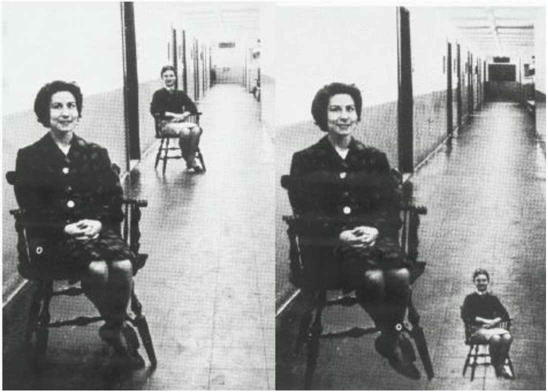
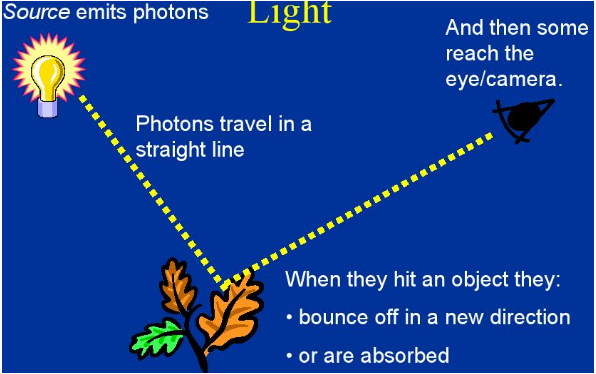
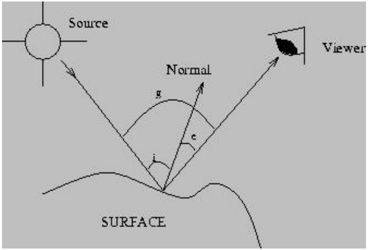


Image formation is a physical process that captures scene illumination through a lens system and relates the measured energy to a signal.

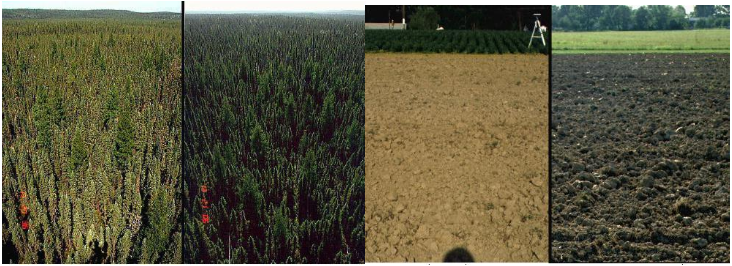


# Basic concepts

- Irradiance E: Amount of light falling on a surface, in power per unit area (watts per square meter).
- Radiance L: Amount of light radiated from a surface, in power per unit area per unit solid angle. Informally “Brightness”.




- Image irradiance E is proportional to scene radiance

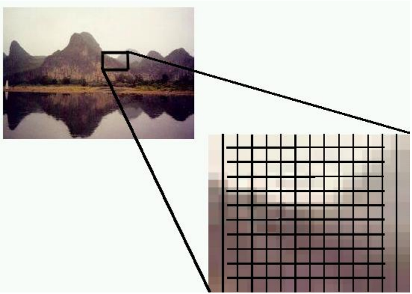


Left: Forest image (left): sun behind observer, (right): sun opposite observer
Right: Field with rough surface (left): sun behind observer, (right): sun opposite observer.

# Digital imaging

Image irradiance E $\times$ area $\times$ exposure time $\rightarrow$ Intensity

- Sensors read the light intensity that may be filtered through color filters, and digital memory devices store the digital image information either as RGB color space or as raw data.
- An image is discretized: sampled on a discrete 2D grid $\rightarrow$ array of color values.



# Imaging acqusition - From world point to pixel

- World points are projected onto a camera sensor chip.
- Camera sensors sample the irradiance to compute energy values.
- Positions in camera coordinates (in mm) are converted to image coordinates (in pixels) based on the intrinsic parameters of the camera:

  - size of each sensor element,

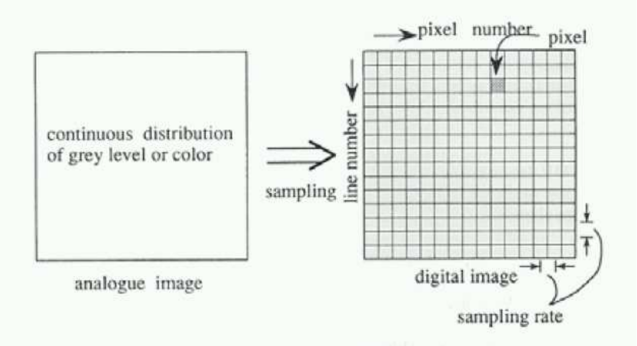
  - aspect ratio of the sensor (xsize/ysize),

  - number of sensor elements in total,

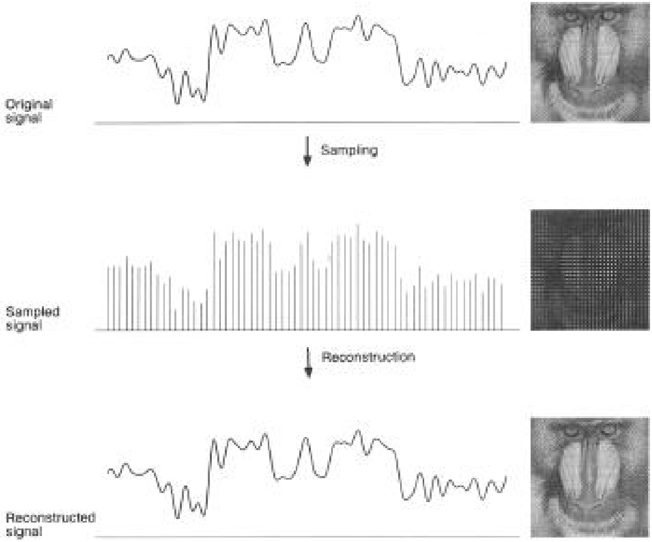
  - image center of sensor chip relative to the lens system.

# Sampling and quantization

- Sample the continuous signal at a finite set of points and quantize the registered values into a finite number of levels.
- Sampling distances $\Delta x, \Delta y$ and $\Delta t$ determine how rapid spatial and temporal variations can be captured.



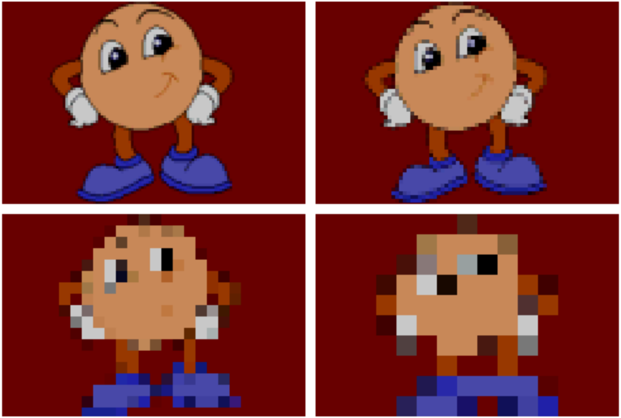
# Sampling and quantization



# Sampling and quantization

- Quantization: Assigning integer values to pixels (sampling an amplitude of a function).
- Quantization error: Difference between the real value and assigned one.
- Saturation: When the physical value moves outside the allocated range, then it is represented by the end of range value.

# Different image resolutions



Sampling due to limited spatial and temporal resolution.


N=64


N=32


N=16


N=8


N=4


N=2

Quantization due to limited intensity resolution.

# Summary of good questions

- 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?
- Why is vision an active process?
- What parameters affects the quality in the acquisition process?
- What is sampling and quantization?

# Recommended readings

- Gonzalez and Woods: Chapters 1.1 - 1.4
- Szeliski: Chapters 1.1 - 1.2
- Introduction to labs (on web page)