Introduction

DD2423 Image Analysis and Computer Vision

Mårten Björkman

Robotics, Perception and Learning Division
School of Electrical Engineering and Computer Science

November 3, 2021

General course information

This year the course will run in hybrid mode, with labs online, but lectures and exam on campus.

- 7.5 hp course (labs 4.0 hp, exam 3.5 hp)
- Course Web in Canvas under course code DD2423
- 2-3 lectures a week
- 16 lectures in total (3 exercise sessions)
- TAs: Wenjie, Marcus, Zehang, Ioanna, Lihao, Alberta, Alfredo, Ci, Yiming, Jiarui, and possibly others.
- If you have questions: preferably use Canvas.

General course information

The course is a broad introduction to computer vision, including image processing and image analysis.

The goal is for students to ...

- learn about basic concepts, terminology, models, and methods in computer vision,
- become acquainted with common methods in computer vision by developing and evaluating them in practice.

The focus is on ...

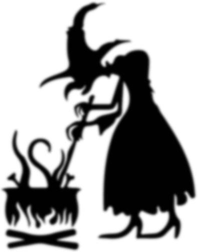
- general problems in computer vision (segmentation, recognition, feature detection, stereo matching, etc),
- the theoretical basis behind those problems, with
- examples of new and traditional methods to solve them.

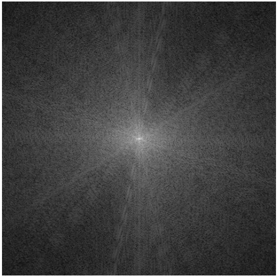
Assessment

- There are 3 labs (LAB1) and 1 exam (TEN1)
- Grading: A-F
  - Final grade: average of exam and labs, rounded towards exam
  - Labs grade: average of labs, rounded towards nearest grade
- Labs are done in Matlab (or Python), possibly on your own laptop.
- There are scheduled times for labs:
  - This year all sessions are fully in Zoom
  - Help: ask for help at `queue.csc.kth.se`
  - Presentation: book a slot in Canvas - no help!
- Doing labs before the deadline - up to 3 pts on the exam

Lab1: Filtering Operations

- Learn about image filtering and the effect of image noise.
- Frequency based representations using the Fourier Transform.



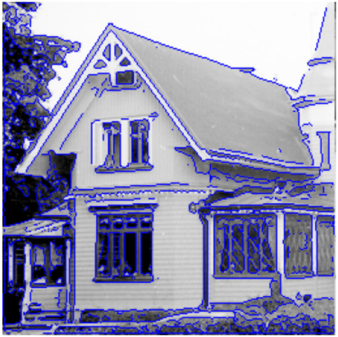


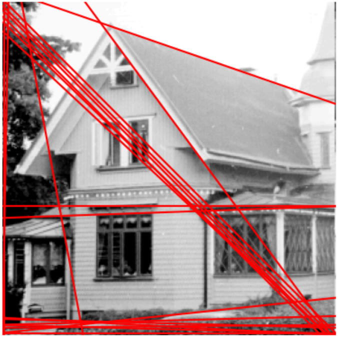
blurring might help
later segmatation

sometimes it is easier to know what
needs to be done when glanced at
the frequency domain.

Lab2: Edge Detection & Hough Transform

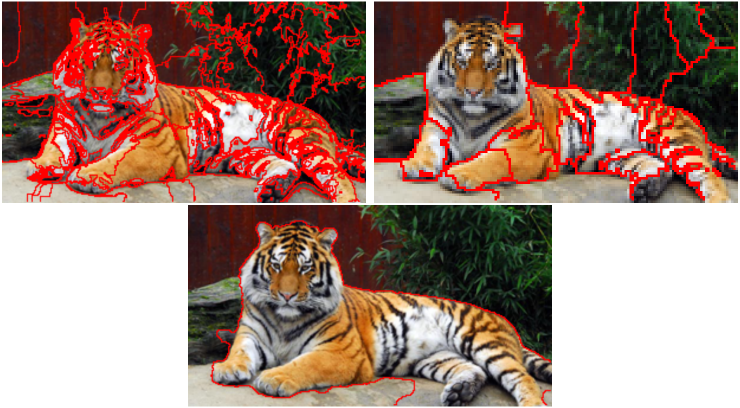
- Learn about how to detect edges and lines in images.
- Get your hands dirty by building something from scratch.





Lab3: Object Segmentation

- Learn about how to divide an image into image regions.
- Study different methods and understand their challenges.



Matlab or Python? – That is the question!

- Matlab Pros:
  - It has worked well many times before.
  - Everything is in one environment.
  - Little fuss with image formats.
- Matlab Cons:
  - Not the natural environment for machine learning.
  - More restricted to the scientific community.
- Python Pros:
  - Dominating language for machine learning.
  - Allows for a lot of flexibility.
- Python Cons:
  - The first time we try it!
  - Need for additional packages: `NumPy`, `SciPy`, `Matplotlib`.
  - Different packages represent images in different ways. !

Lab presentations

- What to do for each lab:
  - Book a slot for presentation in Canvas in the Calendar.
  - Go through the lab instructions!
  - Implement the required functions and run experiments.
  - Answer the questions in the attached answer sheet
  - Upload the following to Canvas in a zip file:
    1. All your code from the lab
    2. A Matlab/Python script that steps through the lab
    3. Filled in answer sheet
  - Present your lab online using Zoom
- Start to work on labs as soon as possible!
- Don’t over-do the answer sheets! Consider it as notes for yourself.
- Think what you learned from the lab, not what you did!

Lab grading

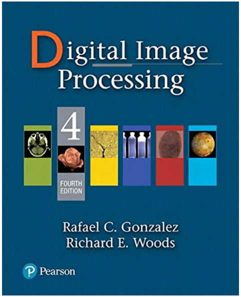
- All labs can be done in **pairs**, but examined **individually**.
- A cumulative definition of grades:
  - E - Lab completed, but many written answers not correct.
  - D - Some written questions have not been answered correctly.
  - C - Minor difficulties in presenting lab results and responding to 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 formal definition on the web page.
- Good idea: Present to each others for practice!

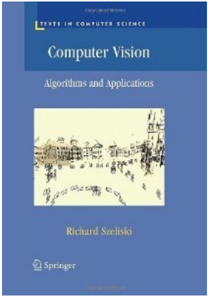
Quizzes for feedback

- Every week quizzes will be posted on Canvas
  - Should not take more than 10–15 minutes to complete
  - Quizzes are recommended, but not compulsory
- Quizzes provide feedback:
  - For you to test your degree of understanding
  - For me to know what requires rehearsal
- Recommendation:
  - After each week, do the corresponding quiz
  - Before attending the exam, redo the quizzes
- Last year I saw a strong correlation between those doing the quizzes and those passing the exam.

Recommended books

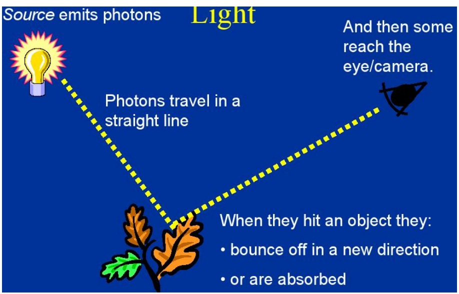
- R. Gonzalez and R. Woods: “Digital Image Processing, 4e edition”, Pearson, 2018.
- R. Szeliski: “Computer Vision: Algorithms and Applications”, Springer, 2021? (available for free: http://szeliski.org/Book)





- Note: the books can give you a second opinion on topics, but assessment is based only on lecture and lab notes.

What does it mean to see?


Source emits photons
Light
And then some reach the eye/camera.
Photons travel in a straight line
When they hit an object they:
• bounce off in a new direction
• or are absorbed

mind that cameras do not see what we see (our brain processes incoming images).

- 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.

Computer vision examples

Figure: Google self-driving cars

Computer vision examples

Figure: Tracking in 1000 Hz (Tokyo Uni)

Why is vision interesting?

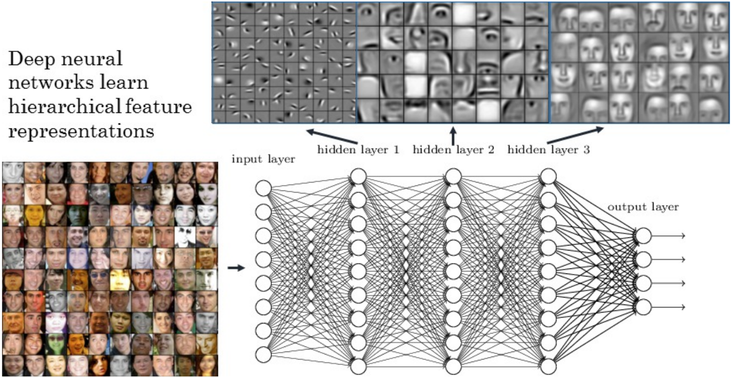
- Intellectually interesting
  - We are too good at it to appreciate how difficult it really is.
  - Going from 2D to 3D is an under-determined inverse problem.
- Psychology:
  - About 50% of cerebral cortex is for vision.
  - Vision is (to a large extent) how we experience the world.
- Engineering:
  - It is fun to build ”intelligent” machines.
  - Computer vision opens up for multi-disciplinary work.
  - We have images everywhere and need ways to organize them.

Multi-disciplinarity

- Neuroscience / Cognition: how do human beings do it?
- Philosophy: how to you e.g. define the concept of an object?
- Physics: how does an image become an image?
- Geometry: how do things look under different orientations?
- Signal processing: how do you work on images in practice?
- Statistics: deal with noise, develop appropriate models.
- Machine learning: how to draw conclusions from lots of data?

What about deep learning?

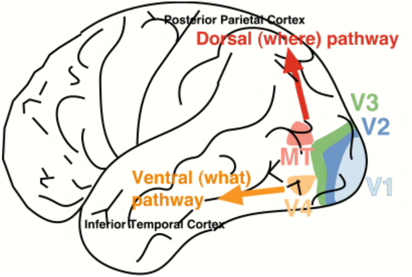
Why study computer vision, when we now have deep learning?



We need to understand the problems, before we can solve them. Deep learning gave us new tools, but the problems are still the same.

What about deep learning?

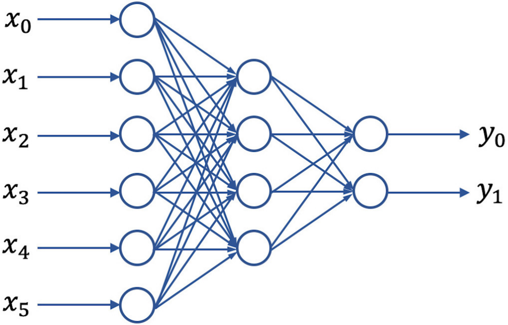
Visual cortex with *what* and *where* pathways.



Deep learning can

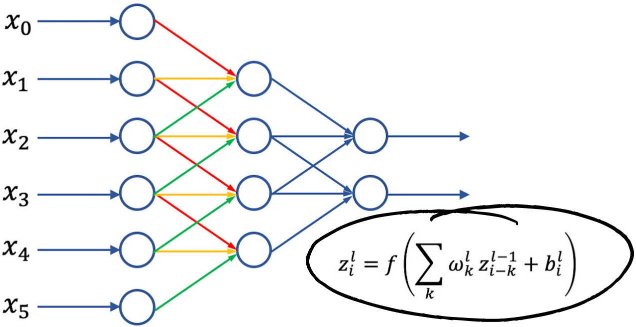
- answer *what*-questions – but *where*-questions are also important.
- benefit from lots of data – but what if you don't have much data?

Fully-connected neural networks (FCN)



- Assume given: Many sets of training samples with matching inputs $\{x_0, ..., x_5\}$ and expected outputs $\{y_0, y_1\}$.

Alternative: Convolutional neural networks (CNN)



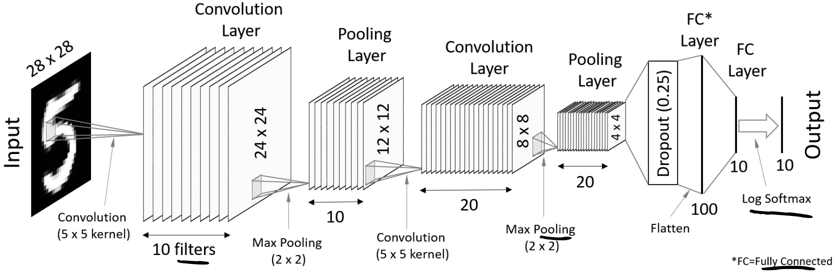
$$z_i^l = f\left(\sum_k \omega_k^l \, z_{i-k}^{l-1} + b_i^l\right)$$

Two important modifications:

- Weight sharing: use the same weight for all links of similar colour.
- Only connect to neurons in a local neighbourhood, not all neurons.

In this case: $3 + 1 = 4$ unknown parameters, instead of $6 \times 4 + 4 = 28$.

Convolutional neural networks (CNN)

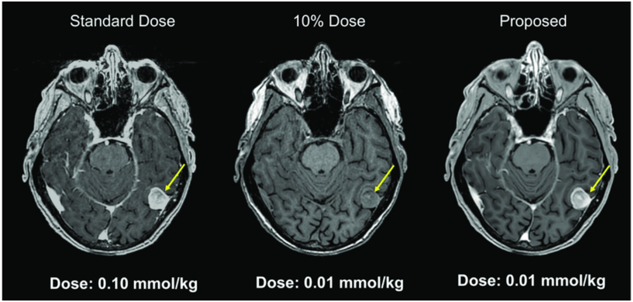


- Instead of a large weight matrix, apply multiple small local filters

  Fewer parameters to learn ⇒ easier to train for images

  ex. FCN: $28^4 = 614'656$, CNN: $5x5x10x20 = 5'000$ parameters
- Pooling: gradually reduce size by maximizing (or averaging) in small local windows
- Finish with a couple of fully-connected layers

Image processing



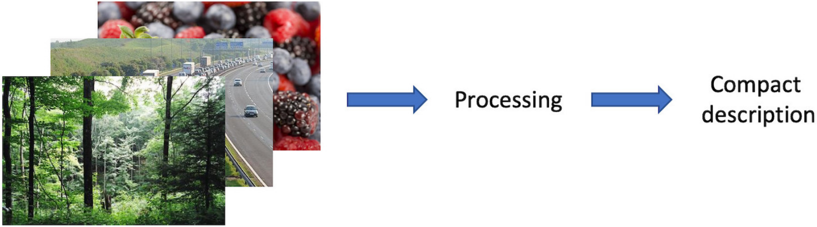
Purpose:

- Enhance important features & suppress disturbances (noise).
- Examples: Poor image data in medicine, astronomy, surveillance.

Subjects treated in this course:

- Image sampling, digital geometry (Lecture 2)
- Linear filter theory, the Fourier Transform (Lecture 3 and 4)
- Enhancement: gray-scale transform, image filtering (Lecture 5)

Image analysis

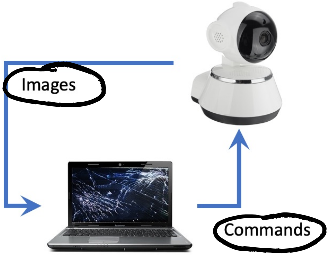


- Purpose: Generate a useful compact description of the image.
- Example: Matching representations of objects to image data.

Subjects studied in this course:

- Image and shape representation (Lecture 6)
- Feature detection and matching (Lecture 7 and 9)
- Object and image segmentation (Lecture 8 and 10)

Computer vision



- Purpose: Achieve an understanding of the world, possibly under active control of the image acquisition process.

Subjects studied in this course:

- Recognition and classification (Lecture 11)
- Stereo geometry (Lecture 12)
- Motion and optical flow (Lecture 13)

However, whole field often called computer vision (incl. image analysis)

Computer vision examples

Figure: Scene parsing (Hong Kong)

Computer vision examples

Figure: OpenPose: Multi-person tracking (CMU)

From a 2D image to a 3D world

$<$ Underdetermined 2D → 3D problem $>$

Main assumptions:

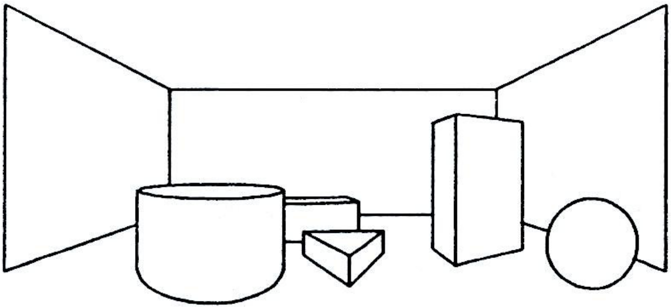
- The world we observe is constructed from coherent matter.
- We can therefore perceive it as constructed from smooth surfaces separated by discontinuities.

The importance of discontinuities: A discontinuity in the image (and edge) may correspond to a discontinuity in either

- Viewing distance
- Surface orientation
- Illumination changes
- Surface texture

The importance of discontinuities

What are the explanations for the discontinuies you see?

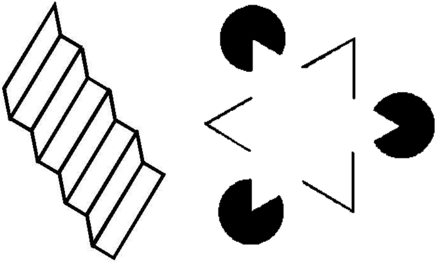


Vision is an active process!

- 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.

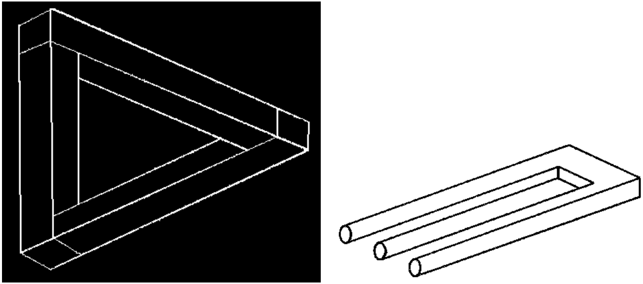
Human vision is not perfect!

Reversing staircase illusion and subjective contours:



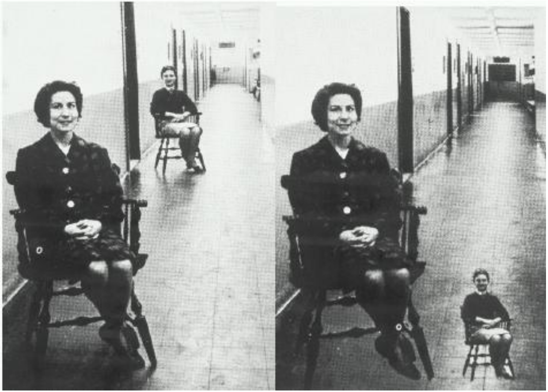
- Our perceptual organization process continues after providing a (first) interpretation. Continue viewing the reversing staircase illusion and you will see it flip into a second staircase.

Impossible objects



Another example that vision is an ongoing process. We don't immediately see that is doesn't make sense.

Depth illusion - size constancy



Projective size = Object size + Viewing distance, but we tend to invert the problem and directly see the Object size.

Color illusion - color constancy

A = B ?

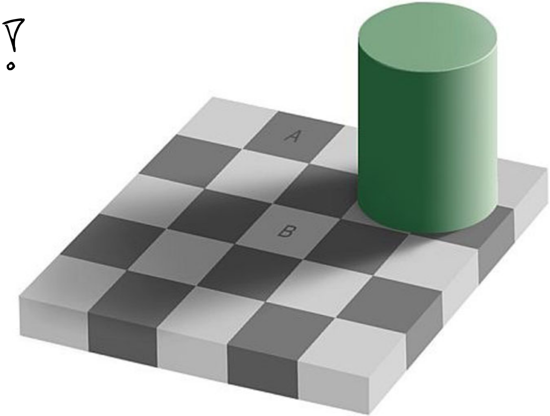
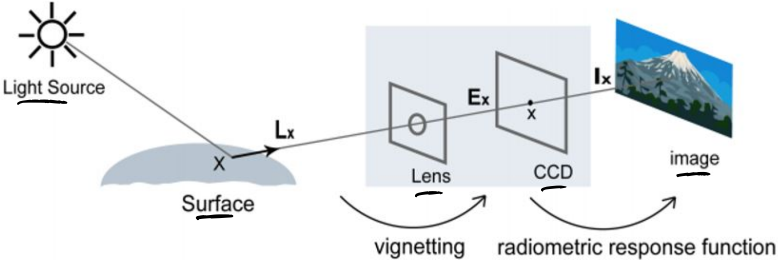


Image color = Object color + Illumination, but (again) we tend to invert the problem and directly see the Object color.

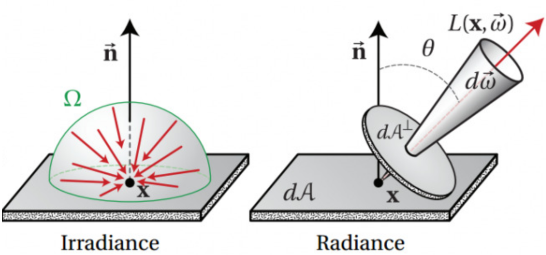
Image formation

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



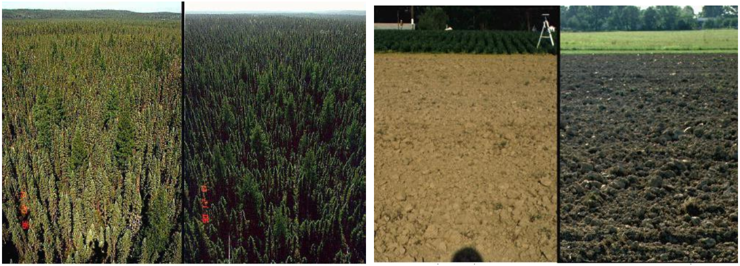
Basic concepts

- 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

Light source examples

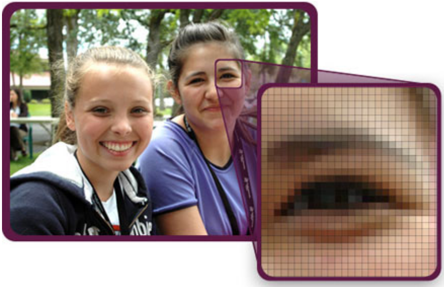


Forest (left) and field (right) with the sun behind (left) or in-front of (right) the camera. Colors change a lot, but structures are similar.

Digital imaging

Image irradiance E × area × exposure time → 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 → 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,

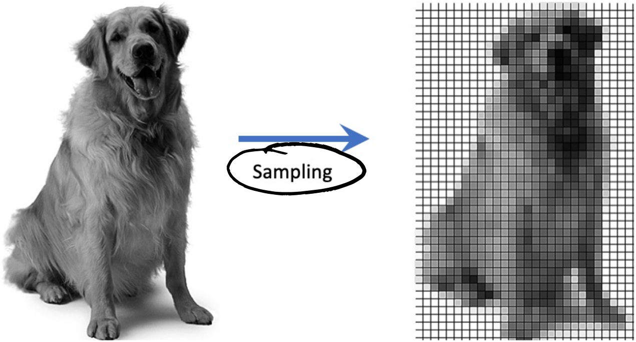
  - aspect ratio of the sensor (xsize/ysize),

  - number of sensor elements in total,

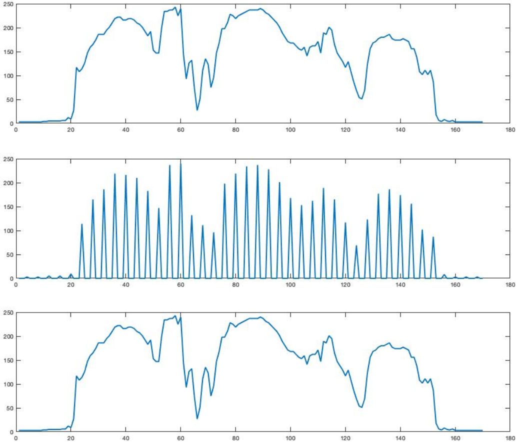
  - image center of sensor chip relative to the lens system.

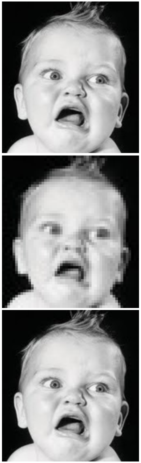
Sampling and quantization

- 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 Δt determine how rapid spatial and temporal variations can be captured.



Sampling and quantization

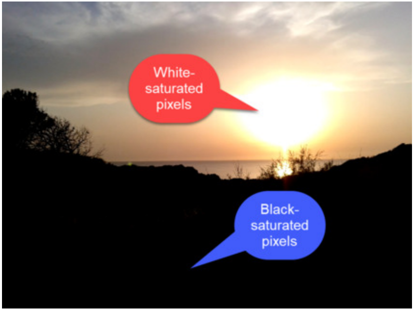




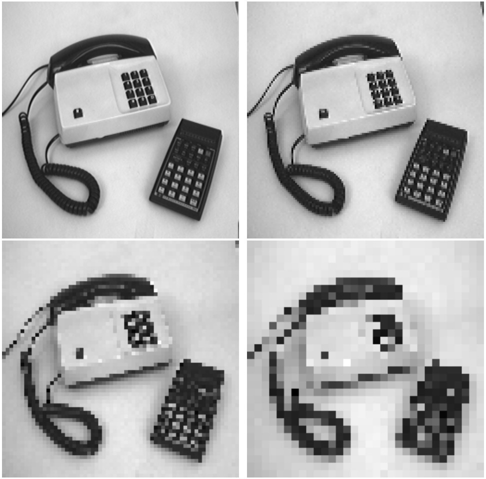
If sampling rate is high enough, the original image can (at least in theory) be perfectly reconstructed.

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.

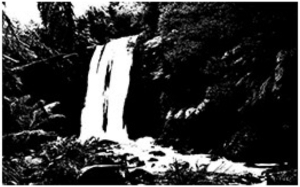


Different image resolutions

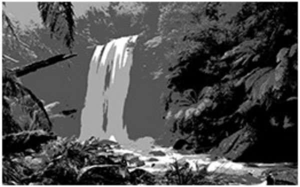


Sampling due to limited spatial resolution. Often the resolution is so high that we don't view it as a problem, but it still is for small objects.

Different number of grey levels

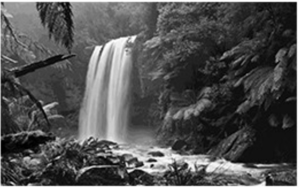


Monochrome (1-bit)

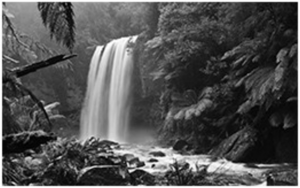


2-bit Grayscale

no need for higher resolution



4-bit Grayscale



8-bit Grayscale

no needed

Quantization due to limited intensity resolution. In most cases only 6-bit (64 gray-levels) is enough to make an image visually pleasing.

Summary of good questions

- What is computer vision good for?
- In what ways is computer vision multi-disciplinary?
- In what sense is vision is an underdetermined inverse problem?
- What is image processing, image analysis and computer vision?
- Why are discontinuities 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 1.4
- Szeliski: Chapters 1.1 - 1.3, 2.3
- Introduction to labs (on web page)