Introduction

DD2423 Image Analysis and Computer Vision

Mårten Björkman

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

October 29, 2019

General course information

- 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: Elena, Jesper, Marcus, Taras, Wenjie, Zehang, Leonard, Sahba, Antonio, Alexandra, Corentin, Jacob and possibly others.
- If you have questions: preferably use Canvas.

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 for labs:
  - Help: get help while working on labs
  - Examination: book a slot in Canvas - no help!
- Do not use only these to work on the labs!
- Doing labs before the deadline - up to 3 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 formal definition on the web page.
- Good idea: Present to each others!

Lab procedure and requirements

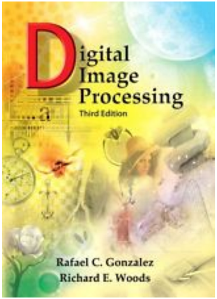
- What to do for each lab:
  - Book a slot for presentation in Canvas.
  - Go through the lab instructions.
  - Implement the required functions and run experiments.
  - Answer the questions in the attached answer sheet.
  - Upload (in a zip file) to Canvas
    1. All your code from the lab
    2. A Matlab script that steps through the lab
    3. Filled in answer sheet
  - Bring a print out of the answer sheet, when you present your lab.
- Start to work on labs as soon as possible!
- Think! What am I supposed to have learned?

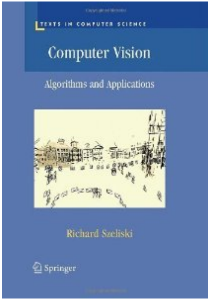
Quizzes for feedback

- 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 to needs 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

Course books

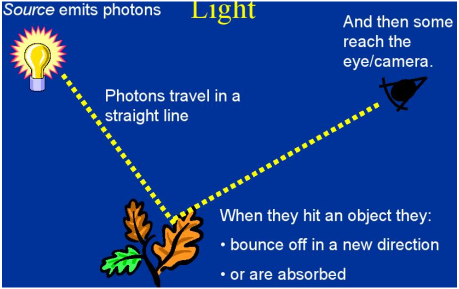
- R. Gonzalez and R. Woods: “Digital Image Processing”, Prentice Hall, 2008.
- R. Szeliski: “Computer Vision: Algorithms and Applications”, Springer, 2010. (available for free: http://szeliski.org/Book)





- 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.

Computer vision examples

Figure: Google self-driving cars

Computer vision examples

Figure: Tracking in 1000 Hz (Tokyo Uni)

Computer vision examples

Figure: Fast book scanning (Tokyo Uni)

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?

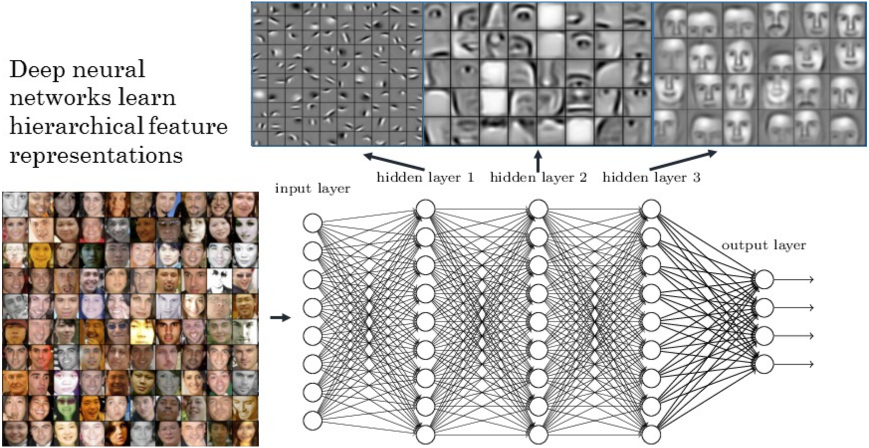
Related courses at RPL

- DD2380 Artificial Intelligence, 6 hp
- DD2421 Machine Learning, 7.5 hp
- DD2434 Machine Learning, Advanced Course, 7.5 hp
- DD2410 Introduction to Robotics, 7.5 hp
- DD2425 Robotics and Autonomous Systems, 9 hp
- DD2424 Deep Learning in Data Science, 7.5 hp
- DD2437 Artificial Neural Networks and Deep Architectures, 7.5 hp

What about deep learning?

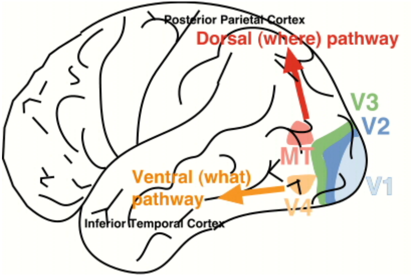
Why study computer vision, when we now have deep learning?

Deep neural networks learn hierarchical feature representations



What about deep learning?

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

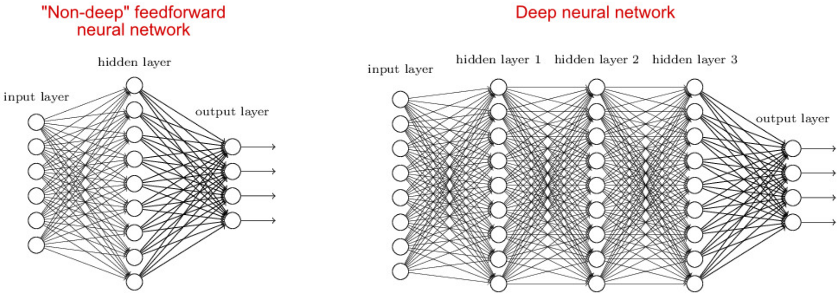


Deep learning can

- benefit from lots of data – but what if you don't have much data?
- answer *what*-questions – but not good at *where*-questions.

Computer vision is so much more than image classification.

Fully-connected neural networks (FCN)



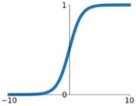
- Neurons on one layer depends on neurons from layer before

$$z_{n+1} = f\left(W_n z_n + b_n\right)$$

with hidden neurons z_n, input neurons $y = z_0$, output neurons $y = z_N$, weight matrix W, bias vector b_n, activation function f.

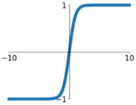
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



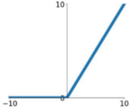
tanh

$\tanh(x)$



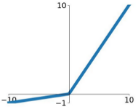
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

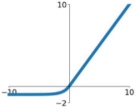


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

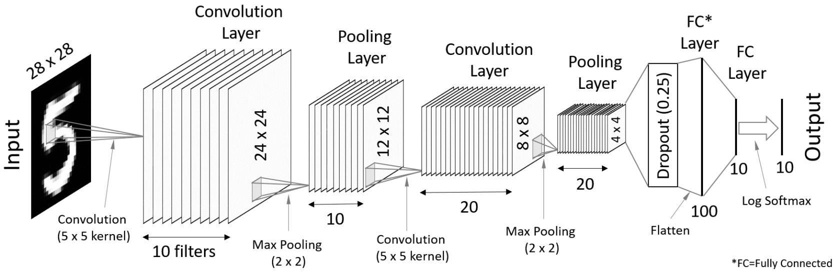
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



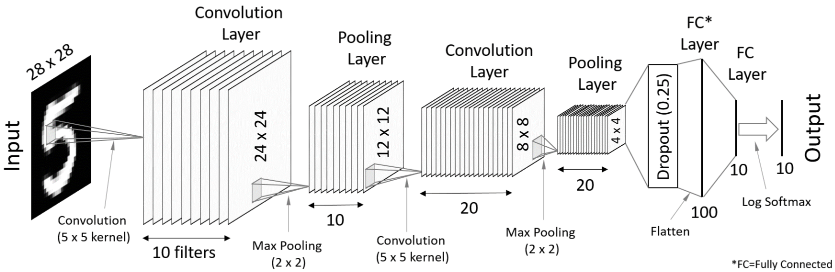
ReLU is the simplest function and is the most widely used.

Convolutional neural networks (CNN)



- Instead of a large weight matrix, apply multiple small local filters
  Fewer parameters to learn $\Rightarrow$ 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 fully-connected layers (like previous slides)

Convolutional neural networks (CNN)



- Convolution layers are based on convolutions

$$z_{n+1}^{c'} = f\,(\sum_{c} w_n^{c,c'} * z_n^c + b_n^{c'})$$

with filter kernels $w_n^{c,c'}$ and neurons z_n^c organized in channels c.

- More on convolutions will be covered in lecture 3.

Image processing $\Longleftrightarrow$ Signal processing

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

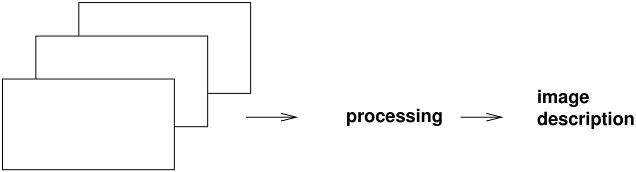
Purpose of image processing

- 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

Image analysis



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

Subjects studied in this course:

- Feature detection and matching
- Shape descriptors
- Image segmentation
- Recognition and classification

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

Image feature detection → object classification

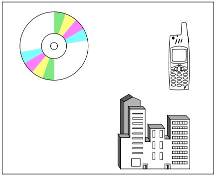
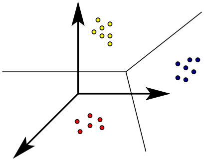
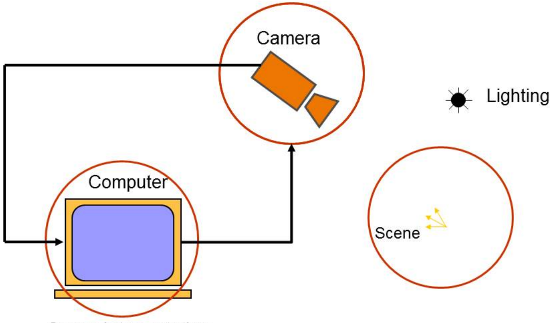


image domain



feature space

Computer vision



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

Computer vision examples

Figure: Scene parsing (Hong Kong)

Computer vision examples

Figure: OpenPose: Multi-person tracking (CMU)

From 3D world to 2D image and back

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

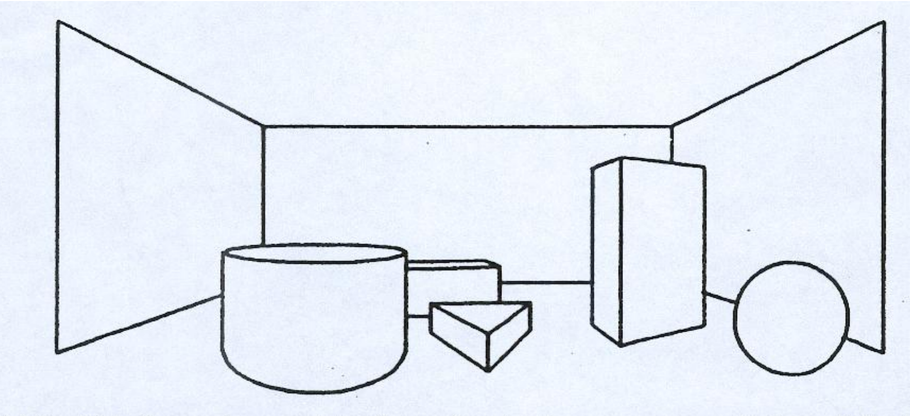
Main assumptions:

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

What are the explanations for the discontinuies you see?

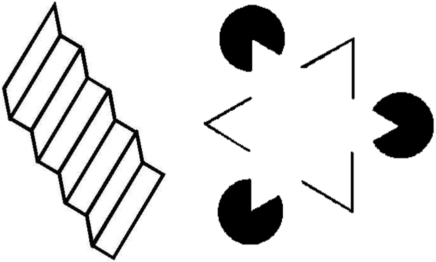


Vision is an active process!

- 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.

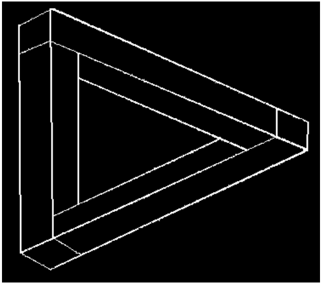
Human vision is not perfect!

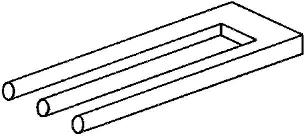
Reversing staircase illusion and subjective contours:



- 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.

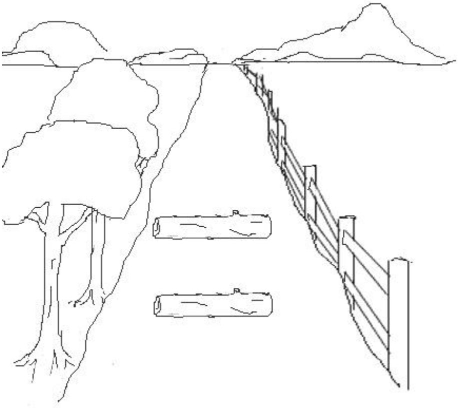
Impossible objects





Another example that vision is an ongoing process.

Depth illusion - size constancy



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

Depth illusion - size constancy

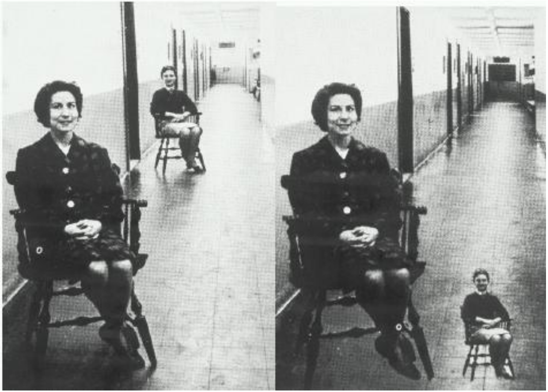
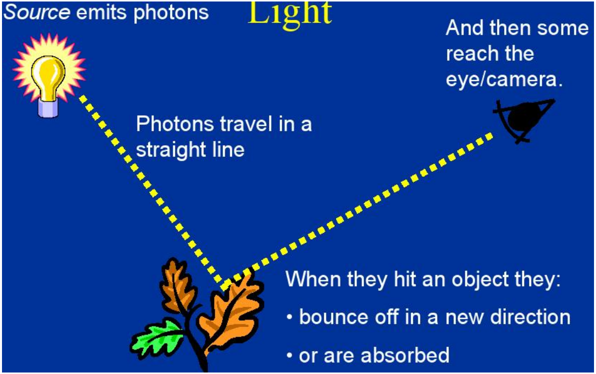
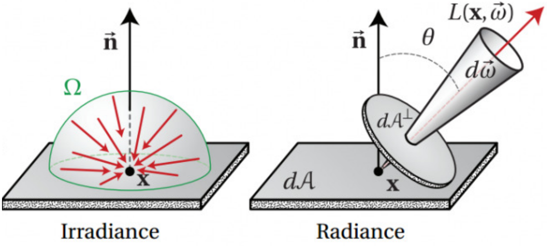


Image formation is a physical process that captures scene illumination through a lens system and relates the measured energy to a signal.



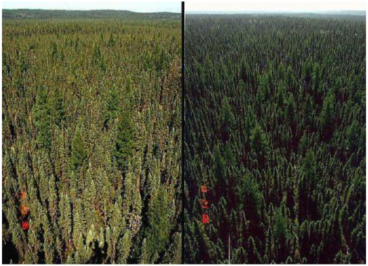
Basic concepts

- 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

Light source examples

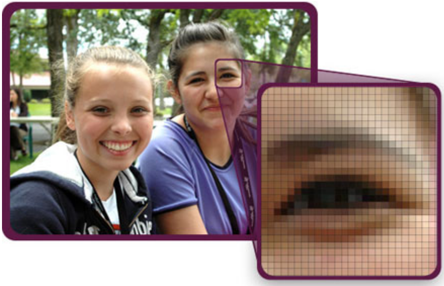


Left: Forest image (left): sun behind observer, (right): sun opposite observer
Right: Field with rough surface (left): sun behind observer, (right): sun opposite observer.

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,

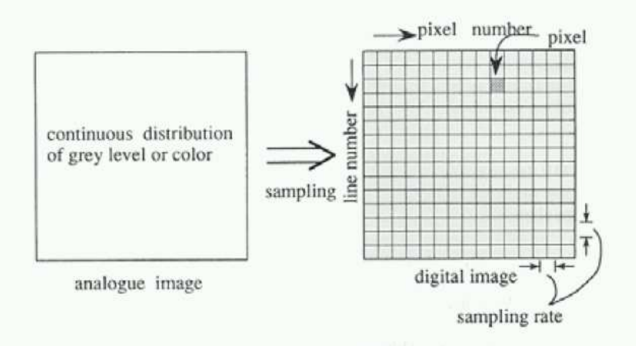
  - aspect ratio of the sensor (xsize/ysize),

  - number of sensor elements in total,

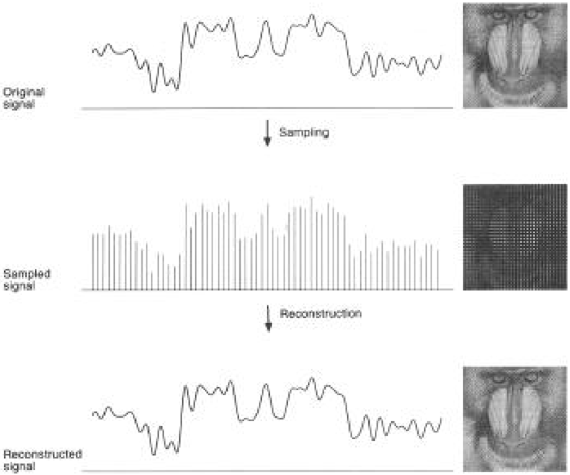
  - image center of sensor chip relative to the lens system.

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



Sampling and quantization

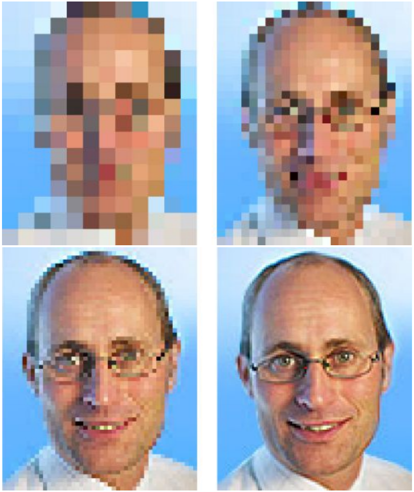


If sampling rate is high enough. original image can (at least in theory) be perfectly reconstructed.

Sampling and quantization

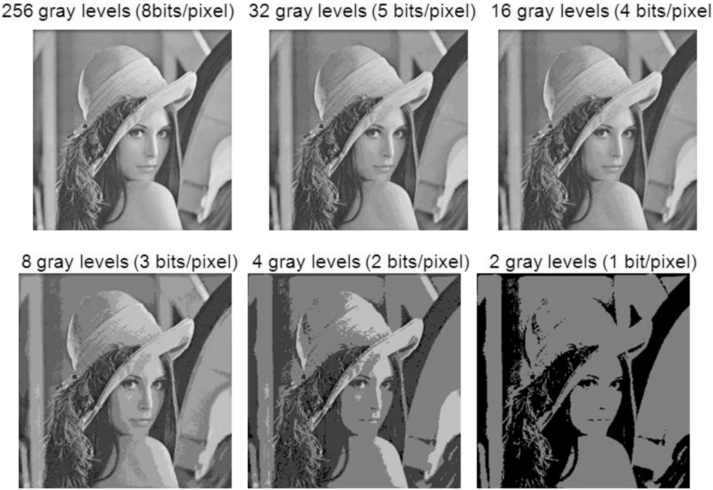
- Quantization: Assigning integer values to pixels (sampling an amplitude of a function).
- Quantization error: Difference between the real value and assigned one.
- Saturation: When the physical value moves outside the allocated range, then it is represented by the end of range value.

Different image resolutions



Sampling due to limited spatial and temporal resolution.

Different number of grey levels



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)