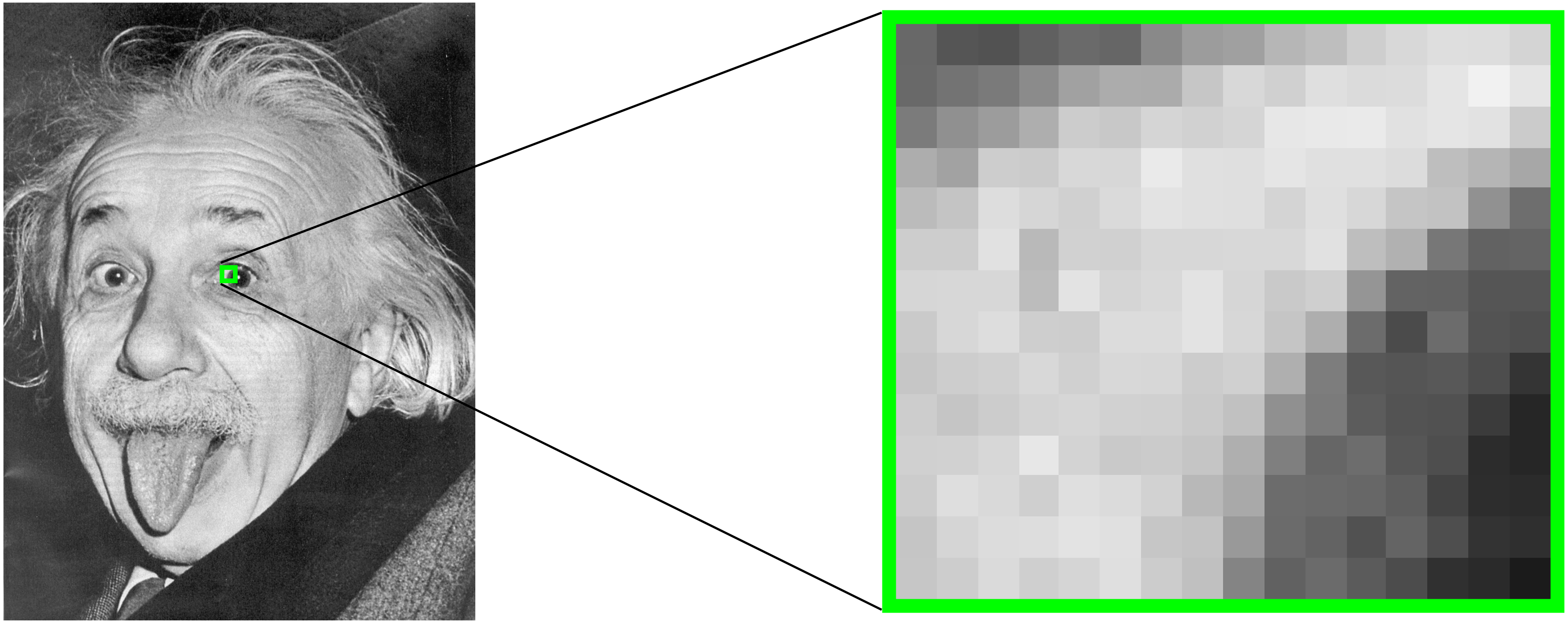
Image Processing and Computer Graphics

Image Processing

Class 1

Introduction and image basics

What is an image?



Author: Daniel Cremers

- Digital image: regular grid I_{ij} of intensity values

Coordinates Value of int.

- Continuous function $I : (\Omega \subset \mathbb{R}^2) \to \mathbb{R} \qquad (x, y) \mapsto I(x, y)$

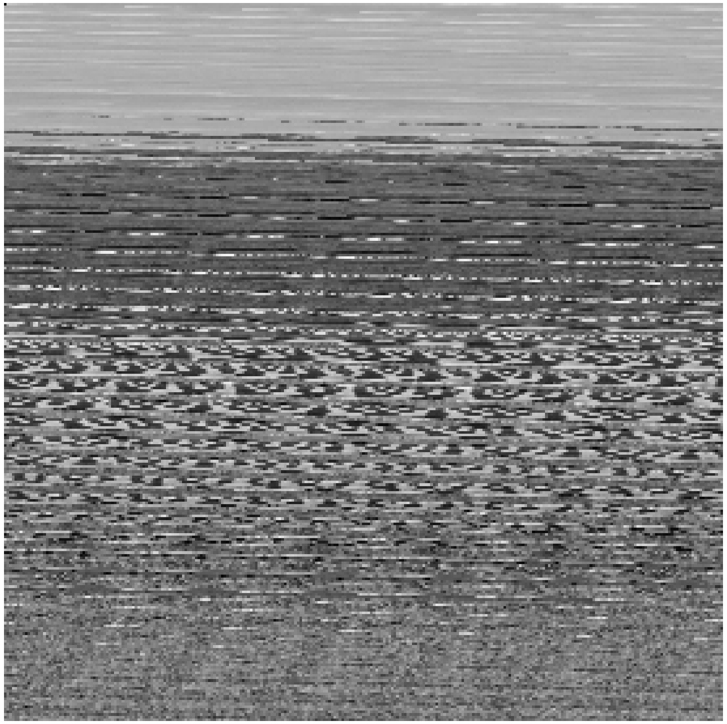
Why is computer vision difficult?

- Vision is a natural and easy task for humans (and many animals)
- This is not for free: ~50% of the primate's cortex deals with the processing of visual information (Felleman-van Essen 1991)
- Matching human visual capabilities means to solve a large part of the AI problem
- Images (together with language) are most popular in machine learning research

Importance of the spatial arrangement

- Image content is defined by the spatial arrangement of intensities
- It is not sufficient to treat images as vectors and to analyze these vectors

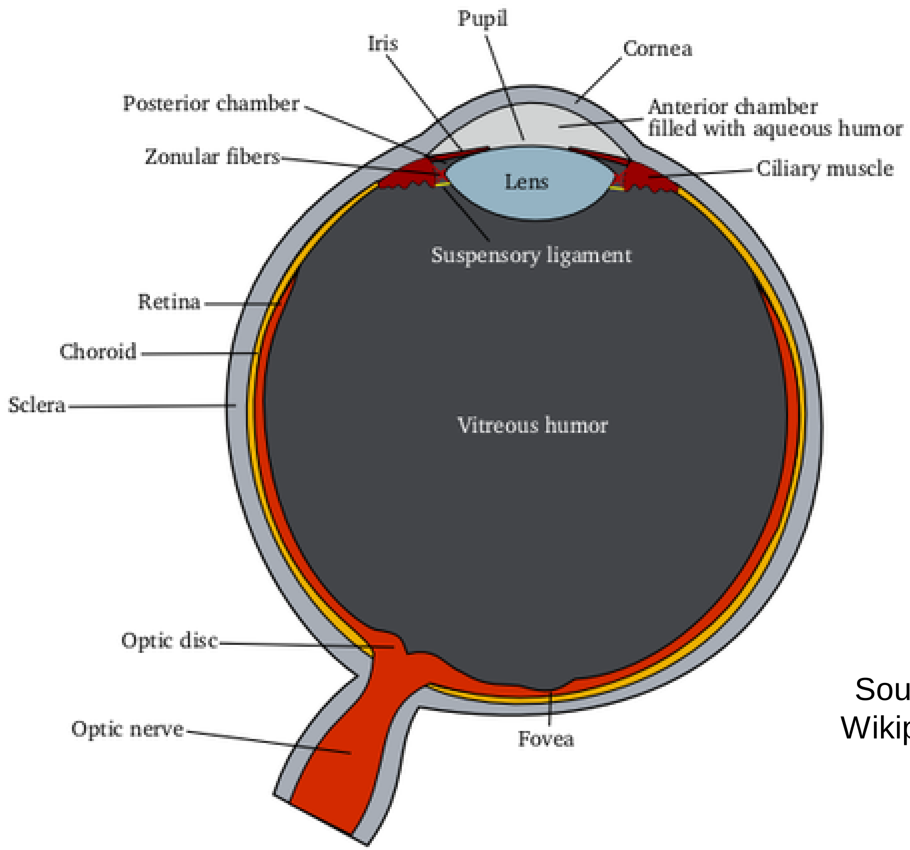
Zebra image



Same image with a different row length

Ambiguities resolved by context

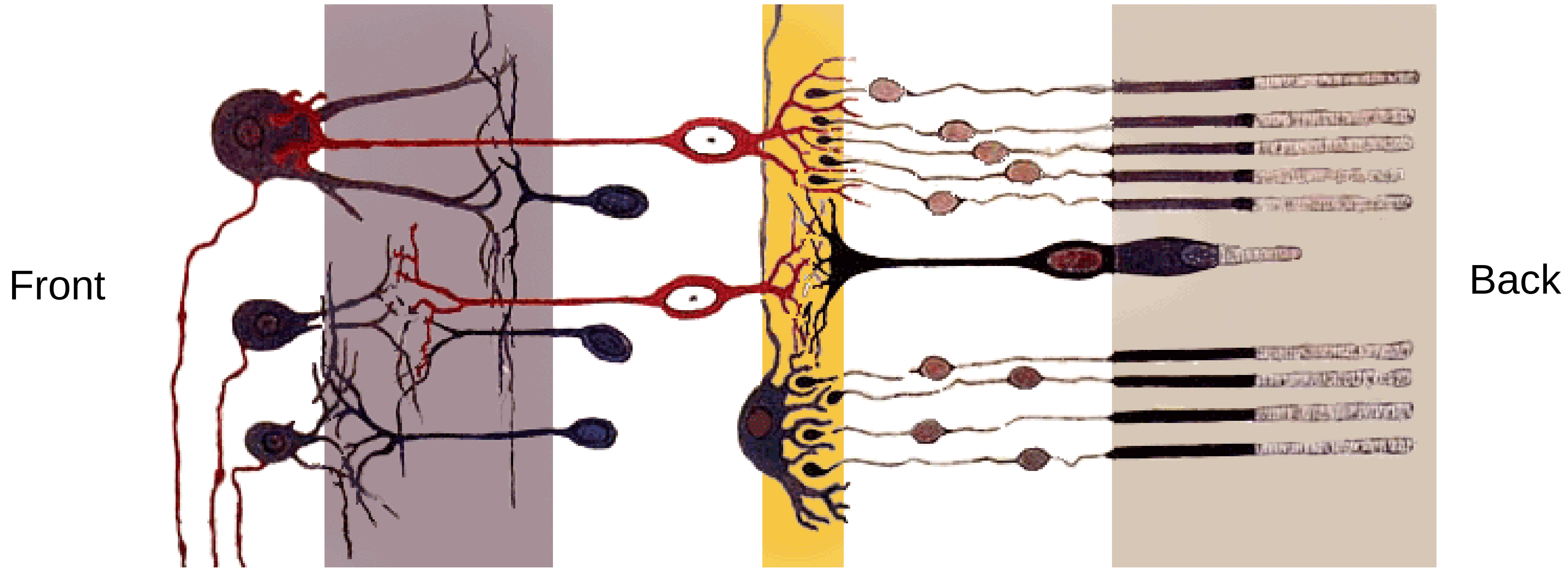
The human eye



Source: Wikipedia

- Looks crude, but it's very well optimized
- Especially the ability to adapt to changing light intensities is noteworthy

The retina



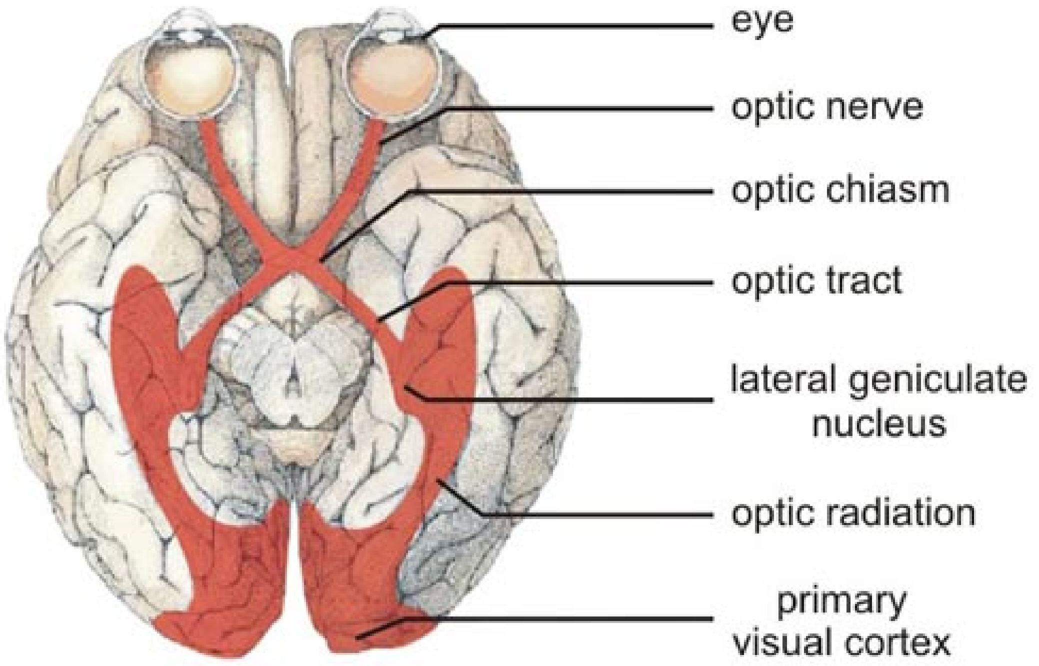
Source: Wikipedia

- Cones (color vision) and rods (gray, high sensitivity)
- Most cones near the optic center
- A lot of analog preprocessing is going on here (mainly smoothing and edge detection)
- Sharp image only in the center (very different from digital images)

What do you think?

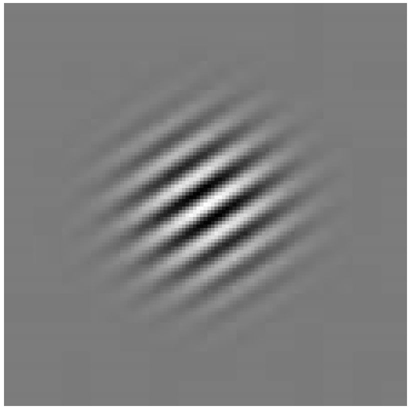
- The retina significantly reduces the amount of data before it is sent to the brain. In the visual cortex the data is then expanded again.

  Can you imagine why?

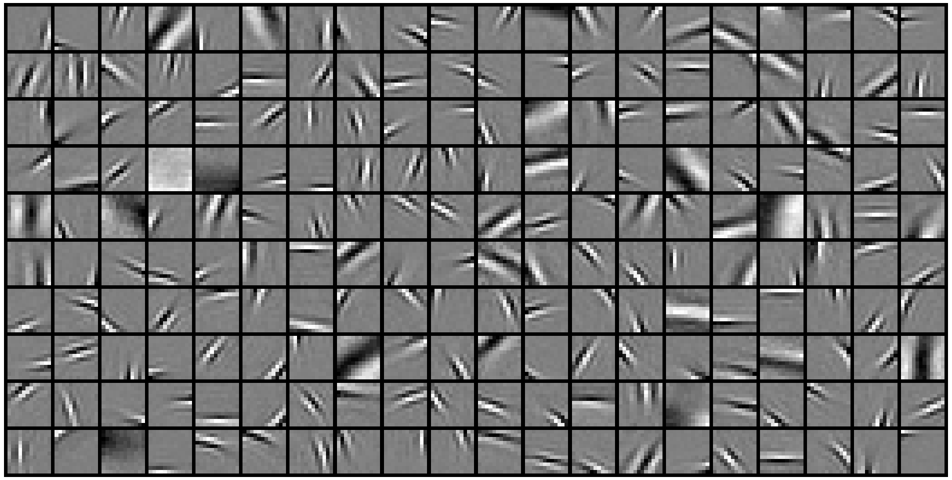


The primary visual cortex (V1)

- The brain's representation is "digital" – spiking or not spiking
- Detailed information coded in spike rate and timing
- Orientation selective cells
  - Selective response to stripes of certain orientation and scale
  - Correspond to Gabor filters
  - Most basic features in an image
  - Found by Hubel-Wiesel 1959 (in cat, later in primates)
- Sparse coding
  - Over-complete code (many different basis functions)
  - Only few are active when presenting a certain signal

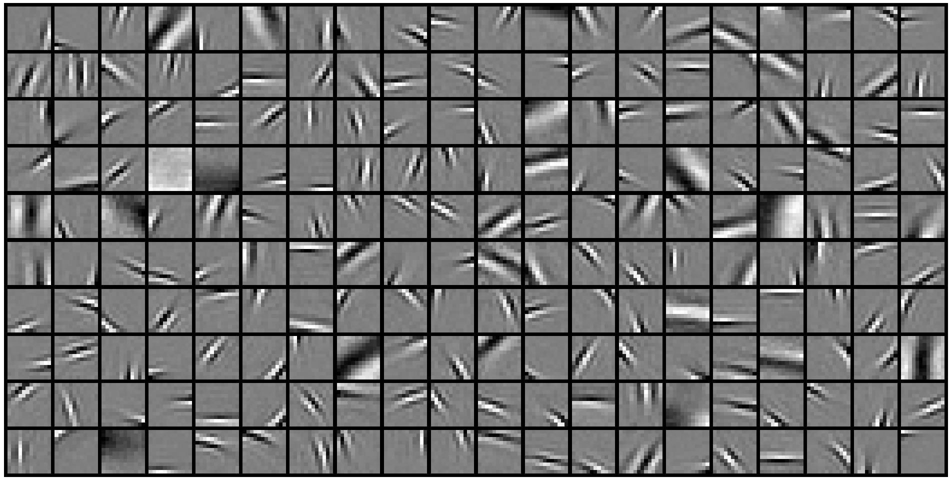


Gabor filter

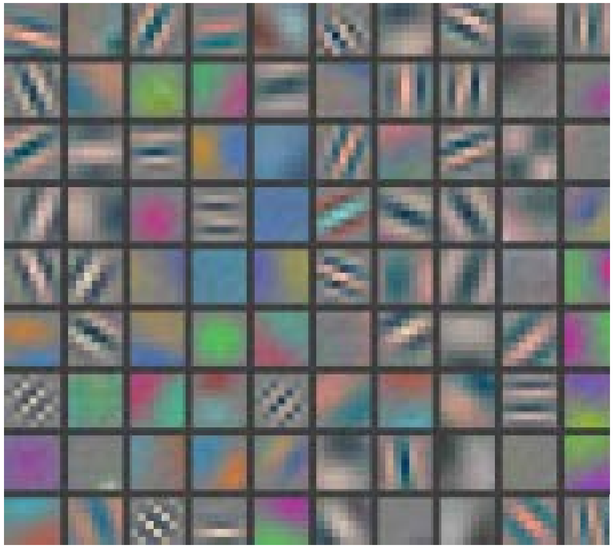


Basis functions of sparse coding model (Author: Bruno Olshausen)

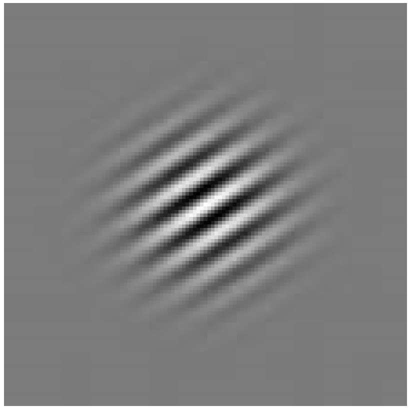
It's well understood how V1 works



Basis functions of sparse coding model
(Author: Bruno Olshausen)



First layer filters of a deep network trained on image classification
(Author: Matthew Zeiler)

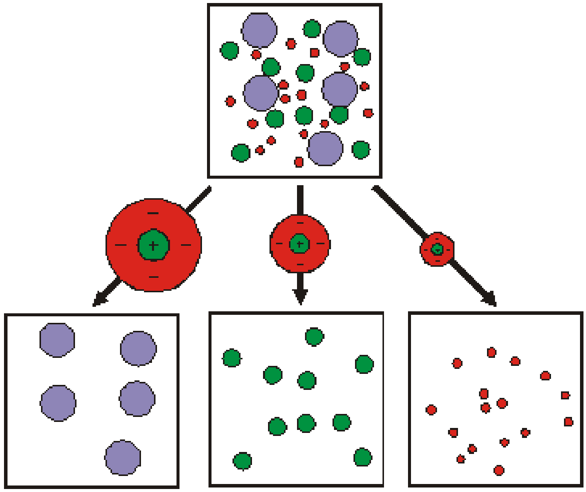


Gabor filter

Receptive fields and integration of information

- In lower cortical areas (especially V1), neurons respond only to signals in a very small local area (**receptive field**)

- In higher cortical areas, the receptive fields get larger and larger due to integration of information from lower areas



Source: Johannes Zanker

- How this integration works, how it is learned, is interesting in both neuroscience and computer science

- Contemporary machine learning follows this concept (deep learning)

Course overview

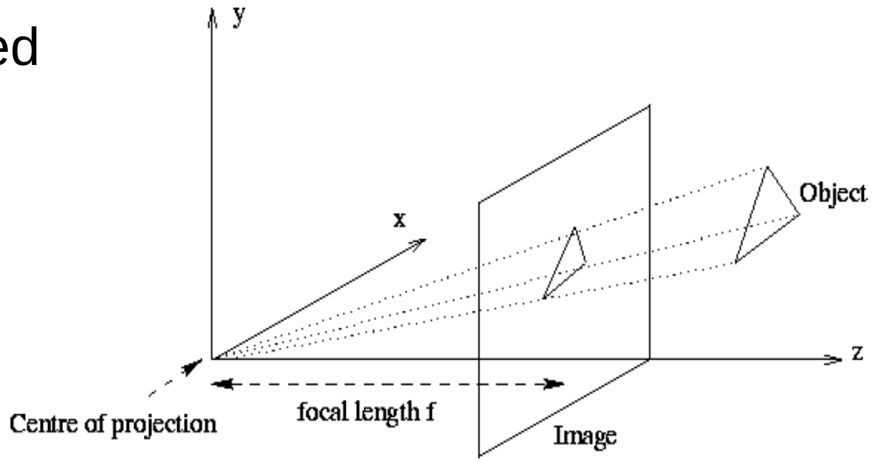
- Class 1: Introduction and image basics
- Class 2: Noise, basic operations and filters
- Class 3: Energy minimization
- Class 4: Variational methods
- Class 5: Motion estimation
- Class 6: Local descriptors and feature learning
- Class 7: 3D shape
- Class 8: Recognition
- Class 9: Segmentation

Imaging model: pinhole camera

- Objects points (X, Y, Z) are projected to image points (x, y) by

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z}$$

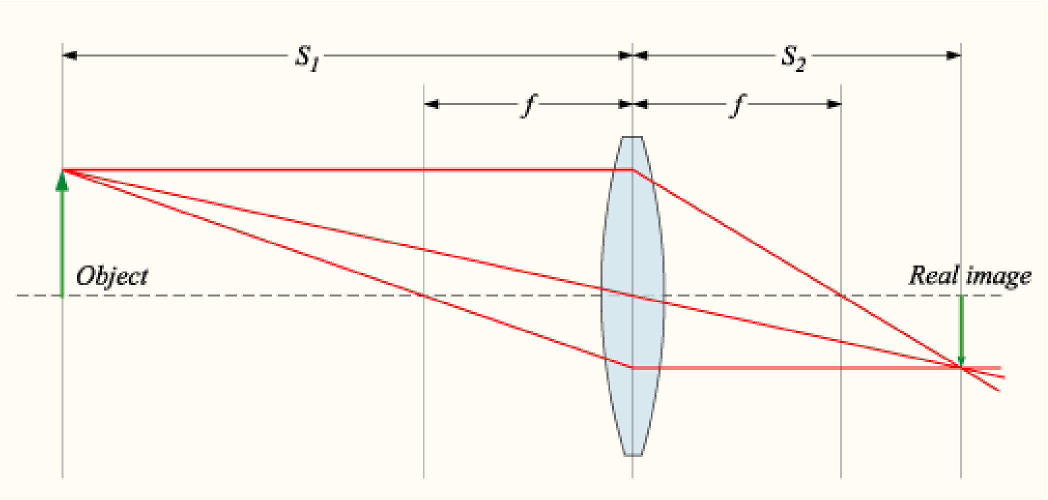
- Simplified camera model



- Practical problem: sharpness vs. light intensity

- Ratio steered by **aperture** (size of the hole)

Optical camera

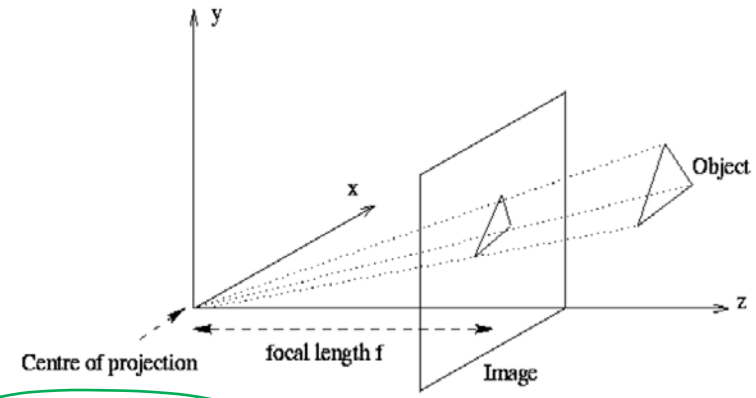


- Light focusing (solves problem of pinhole camera)
- Large aperture and sharpness possible at a certain depth
- Thin lenses: $$\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f}$$
- Wide lenses: focal length depends on orientation and color
- Computer vision practice: pinhole camera + correction of lens effects
- Some exceptions, where the effect of lenses is more important:
  - Shape from defocus
  - Image analysis in microscopy
  - Wide-angle cameras

Fisheye effect of wide-angle cameras

Image representation: gray value images

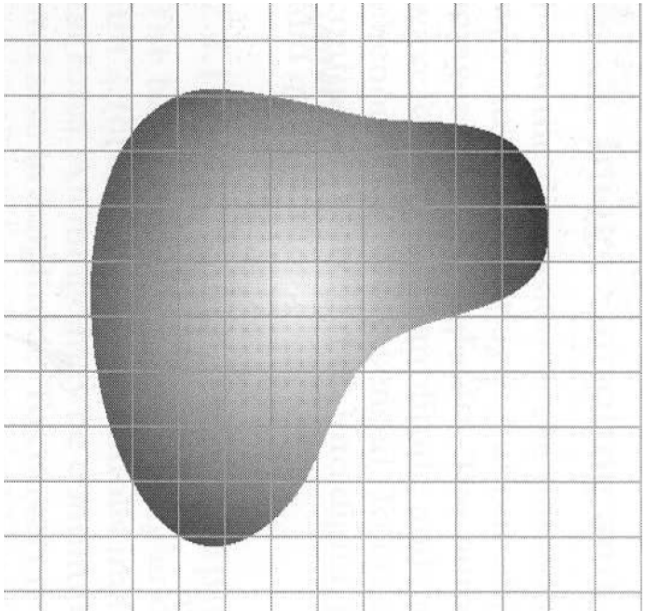
- Continuous 3D world projects light intensities to a 2D plane



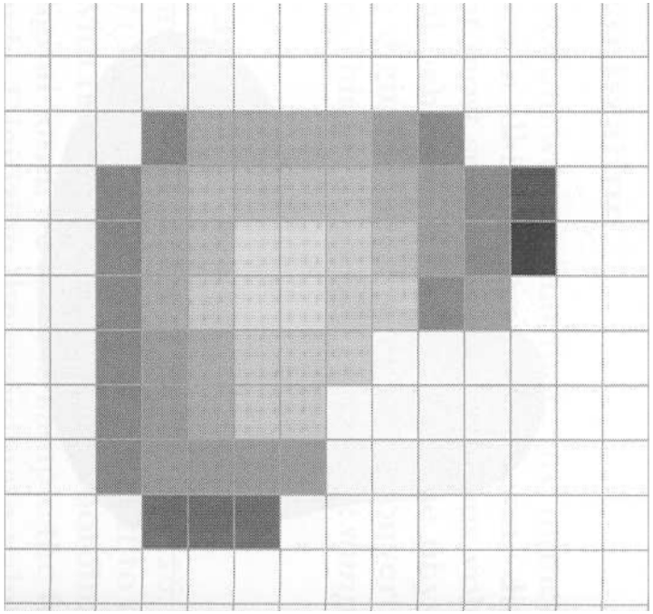
- Image is a continuous function

$$I : (\Omega \subset \mathbb{R}^2) \to \mathbb{R}$$

- Ω is called **image domain**; it is usually rectangular



Continuous image

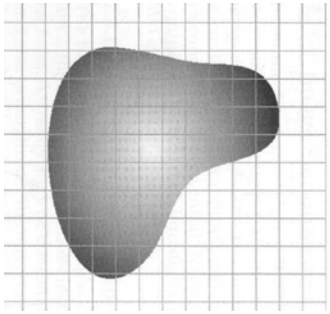


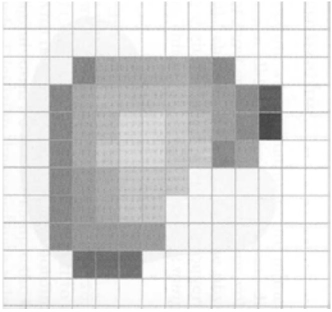
Discrete, sampled image

Sampling

- In digital images, intensities are only given on a pixel grid (e.g. grid of a CCD chip)

$$\{I_{ij} | i = 1, ..., N; j = 1, ..., M\}$$



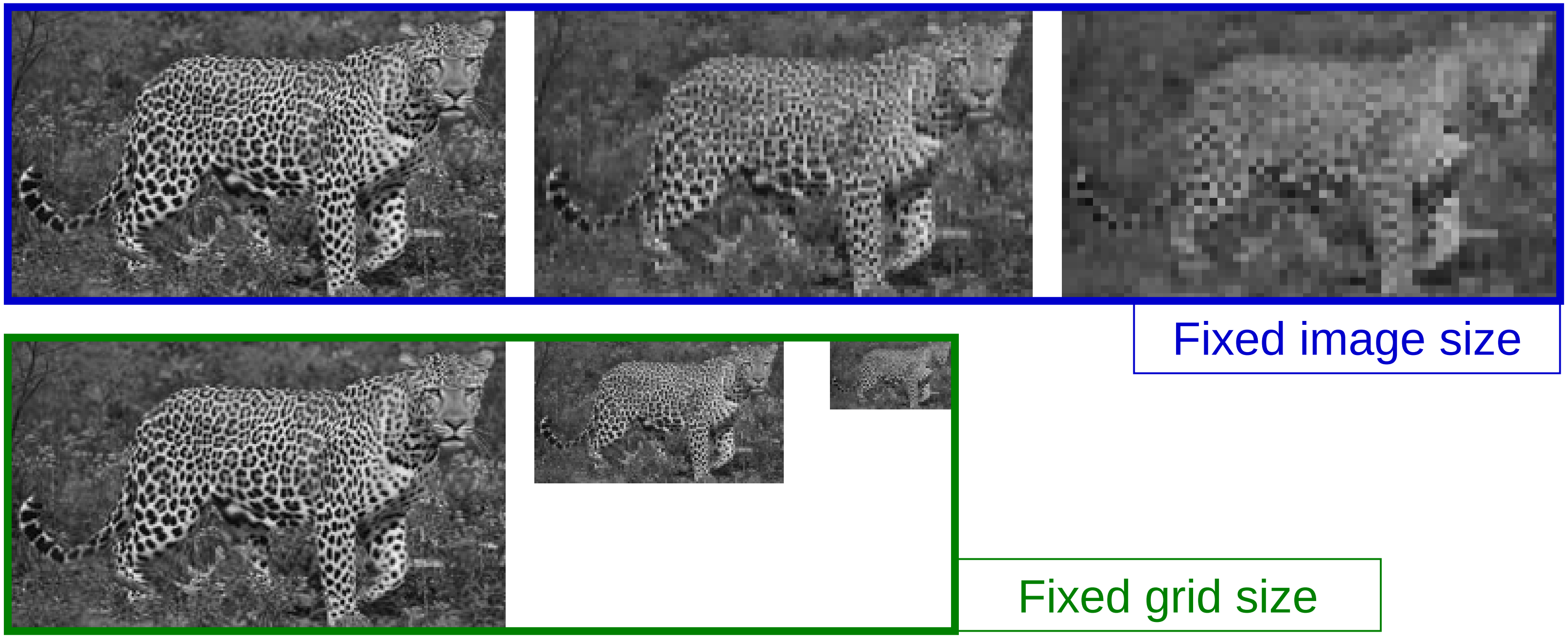


- Discretization of the image domain
- Grid points are called **pixels** (picture elements)
- Grid is usually a rectangular point grid with equal spacing
- **Grid size** h defines the spacing of pixels
- Often same spacing in all directions (square pixels): $h = h_x = h_y$
- If true spacing not known → grid size of input image set to $h = 1$

Image resolution, downsampling, upsampling

- Given a certain image of a scene, the number of grid points to represent the discrete image is called the **image resolution**

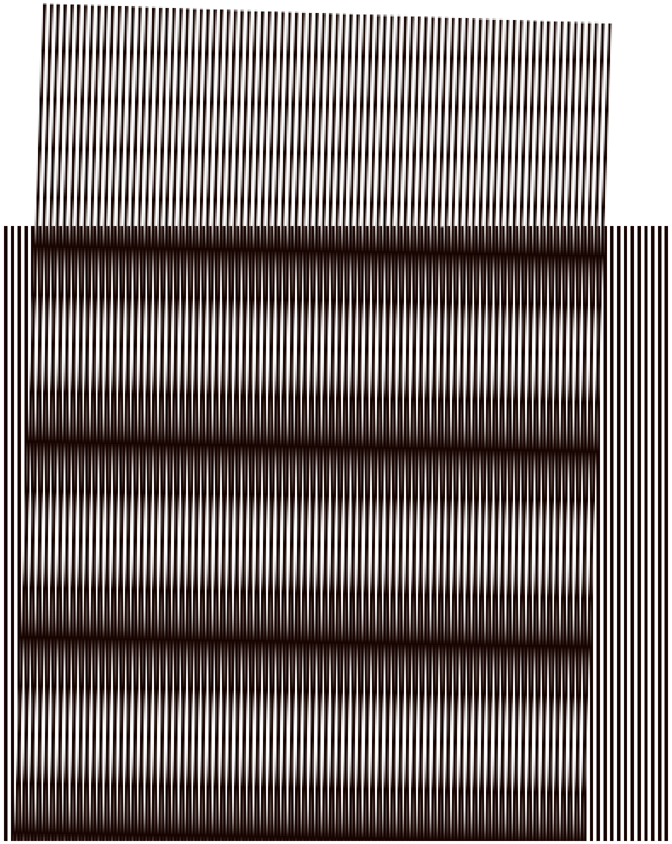
- Reducing the number of grid points is called **downsampling**



- Increasing the number of grid points is called **upsampling**

Aliasing and Moiré effect

- A function can be represented by its frequency components (Fourier transform)
- A discrete signal can only represent frequencies up to a certain limit → **Nyquist frequency**
- Ignorance of the Nyquist frequency leads to **aliasing** artifacts (example: straight lines become stepped)
- Sampling of periodic signals is a frequency modulation (multiplication of two periodic signals)
- It can lead to Moiré effects (a special aliasing artifact)
- Videos can exhibit temporal aliasing

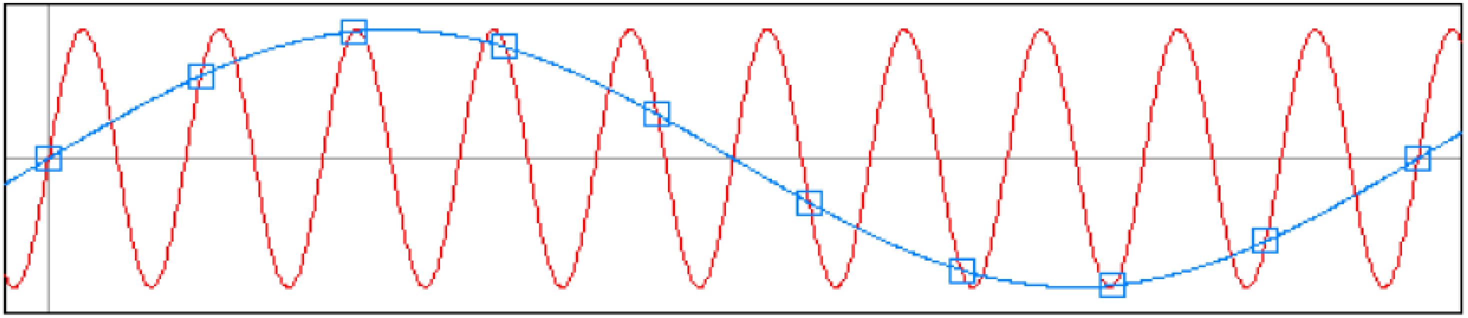


Moiré effect, Source: Wikipedia

Spinning wheel example for temporal aliasing

Nyquist-Shannon theorem

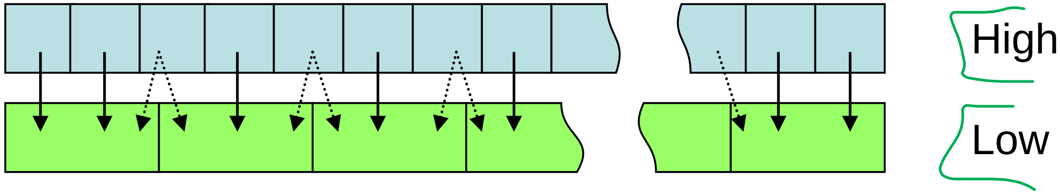
- Two different frequencies may lead to the same discrete signal



- An input signal can be reconstructed from samples in a unique way if the sampling rate is at least two times the bandwidth of the input signal
- Reduction of bandwidth (= maximum frequency) can be achieved by smoothing the signal
- Consequence: smooth your input image sufficiently before downsampling

Downsampling

- Decanting operator: ensures minimum smoothing necessary to avoid aliasing



- Pixels from high resolution image spill their intensity to the pixels of the low resolution image

- In case of overlap, intensity distribution to both cells according to the overlap ratio

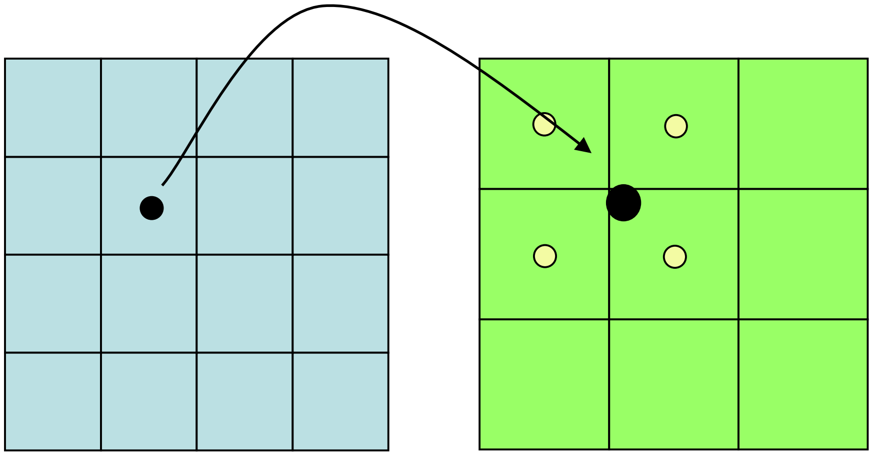
- Normalization of intensity value in the low resolution image by the downsampling factor

- Operator is **separable**: can be applied sequentially along all axes

Upsampling, bilinear interpolation

- Bilinear interpolation: weighted average of neighboring pixels

- Project fine grid point to available coarse grid



- Compute weighted average along x-axis

$$a_j = (1-\alpha) I_{i,j} + \alpha I_{i+1,j}$$

$$a_{j+1} = (1-\alpha) I_{i,j+1} + \alpha I_{i+1,j+1}$$

- Compute weighted average along y-axis

$$I_{k,l} = (1-\beta) a_j + \beta a_{j+1}$$

- General concept to retrieve values at points between grid points

- Can be extended to arbitrary dimensions (trilinear interpolation)

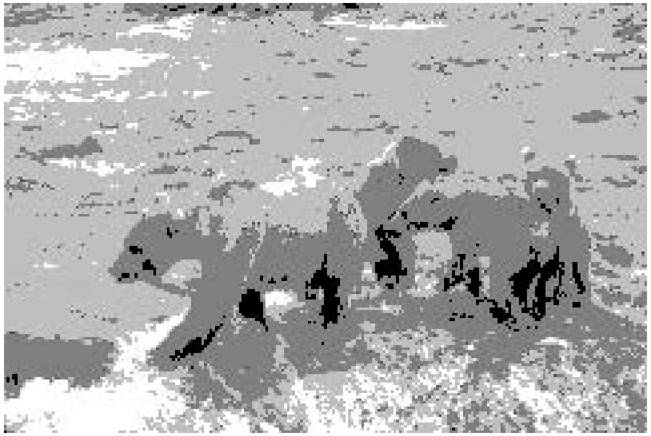
Quantization

- Discretization of the co-domain $\mathbb{R} \mapsto \{1, ..., N\}$
- Needed for representation in the computer (integer or float).
- Usual image formats have 256 gray scales → 8 bit per pixel (bpp)
  (often more in microscopy and industrial cameras)
- Humans can distinguish only 40 gray scales (and several thousand colors)
- Optimal quantization by clustering (e.g. k-means)
- We will usually assume $I(x,y) \in \mathbb{R},\ 0 \leq I(x,y) \leq 255$

Quantization

256 scales (8 bpp)

16 scales (4 bpp)



4 scales (2 bpp)

2 scales (binary image)

Types of images

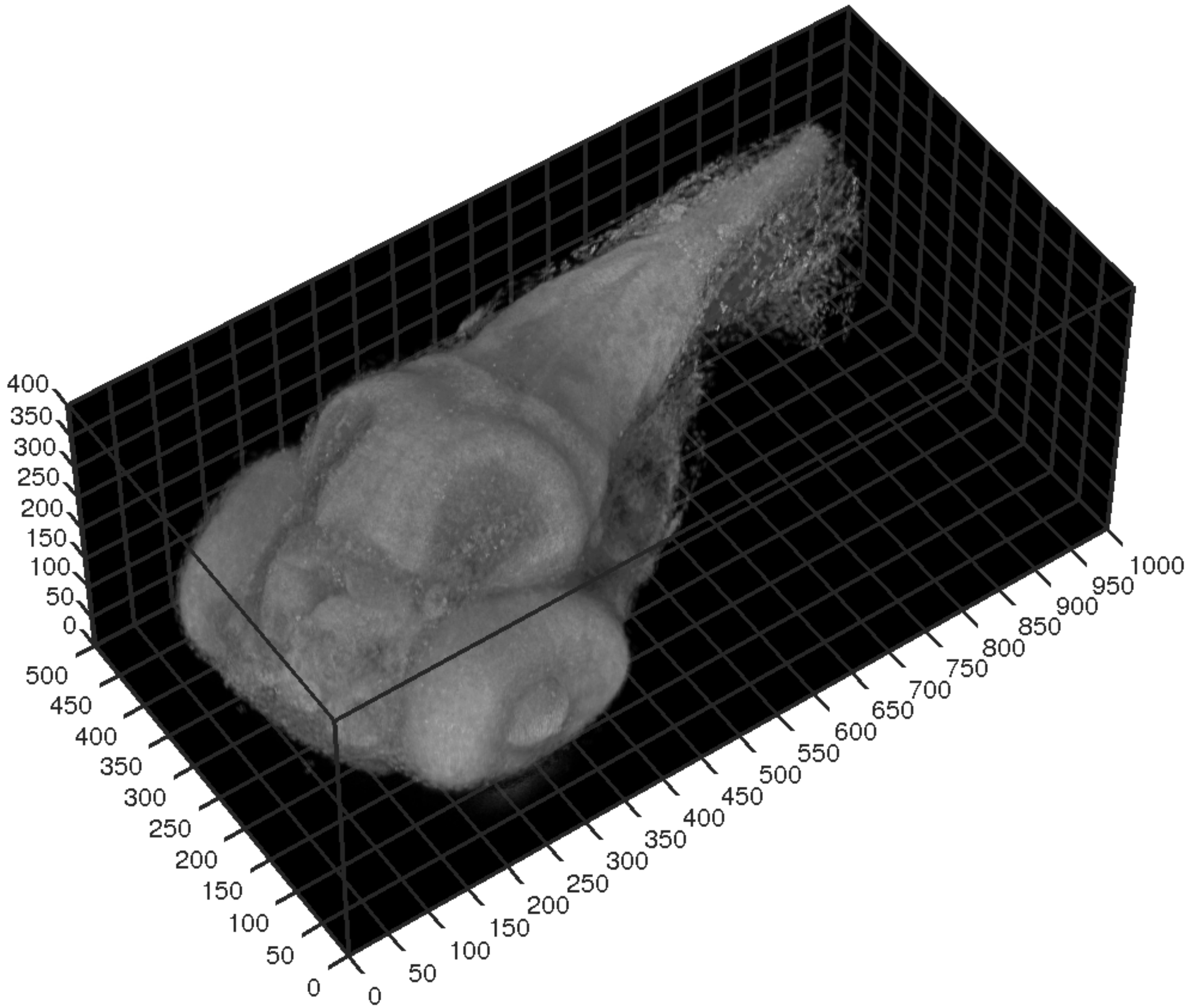
- Generalization of images with respect of
  - dimensionality of the image domain
  - dimensionality of the co-domain

- "Standard" images: image domain is two-dimensional, co-domain one-dimensional (gray scale)

- Other dimensionalities of the image domain:
  - 1D signals
  - 3D images (volumetric images, image sequences)
  - 4D images (sequence of volumetric images)

- Other co-domains
  - Vectors (e.g. color images)
  - Matrices

Volumetric 3D data



Volumetric dataset obtained with a confocal microscope showing a zebrafish larvae

Image sequences

Color images

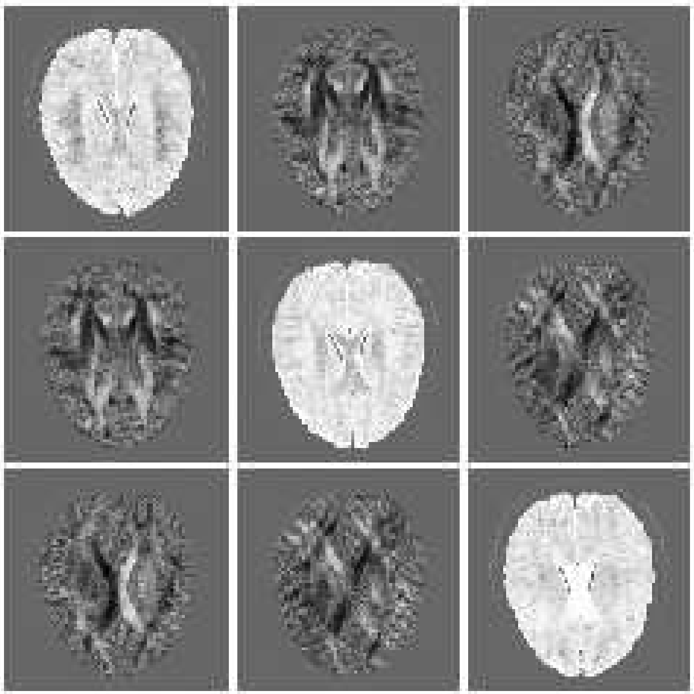
Color image

Red channel

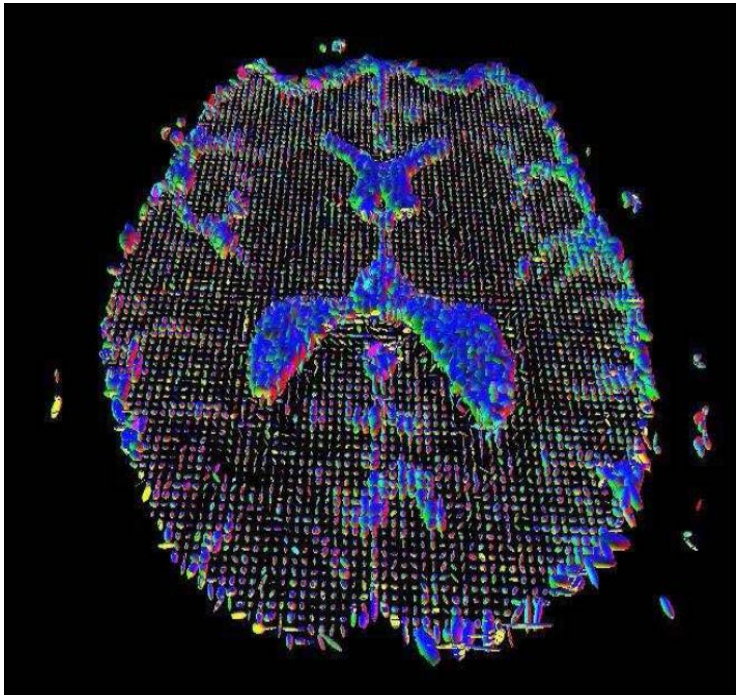
Green channel

Blue channel

Matrix-valued images



Matrix channels of DT-MRI
Weinstein-Kindlmann-Lundberg 1999



3D visualization via ellipsoids
Data: Anna Villanova, BMT, Eindhoven
Visualization: MIA group, Saarland University

- Diffusion tensor MRI: flow preferences of water molecules
- Each voxel (volume element) comprises a 3x3 matrix

Summary

- Image processing, especially computer vision, is a core playground for machine learning.
- The human visual system shows great performance but is only partially understood
- Images are projections of the real, continuous world and therefore continuous functions
- Digital images are discrete approximations of these functions
- (Down)sampling of images requires smoothing/averaging to avoid aliasing artifacts
- Everything with a regular grid structure can be regarded as image

Course webpage, exercises

- These slides are available at:
  http://lmb.informatik.uni-freiburg.de/lectures (user: open, passwd: thebox)

- Refer to the website also for the exercises

- You can also find some recordings from earlier semesters, but the course is being reworked this semester
  → directory name “newslides” in the link indicates up-to-date content

- Programming language: C vs. Python