University of Osnabrück, Institute of Cognitive Science

# Computer Vision

# 2 – Image Acquisition

WS 2019 / 2020

Gunther Heidemann

# Survey

- Aim: Understanding how a scene is mapped onto a CCD
- Abstraction: Pin hole camera
- CCD-camera
- Human eye
- Color
- Image representation
- Loss of information
- Other types of image acquisition

# First known photo: 1822



[FP]

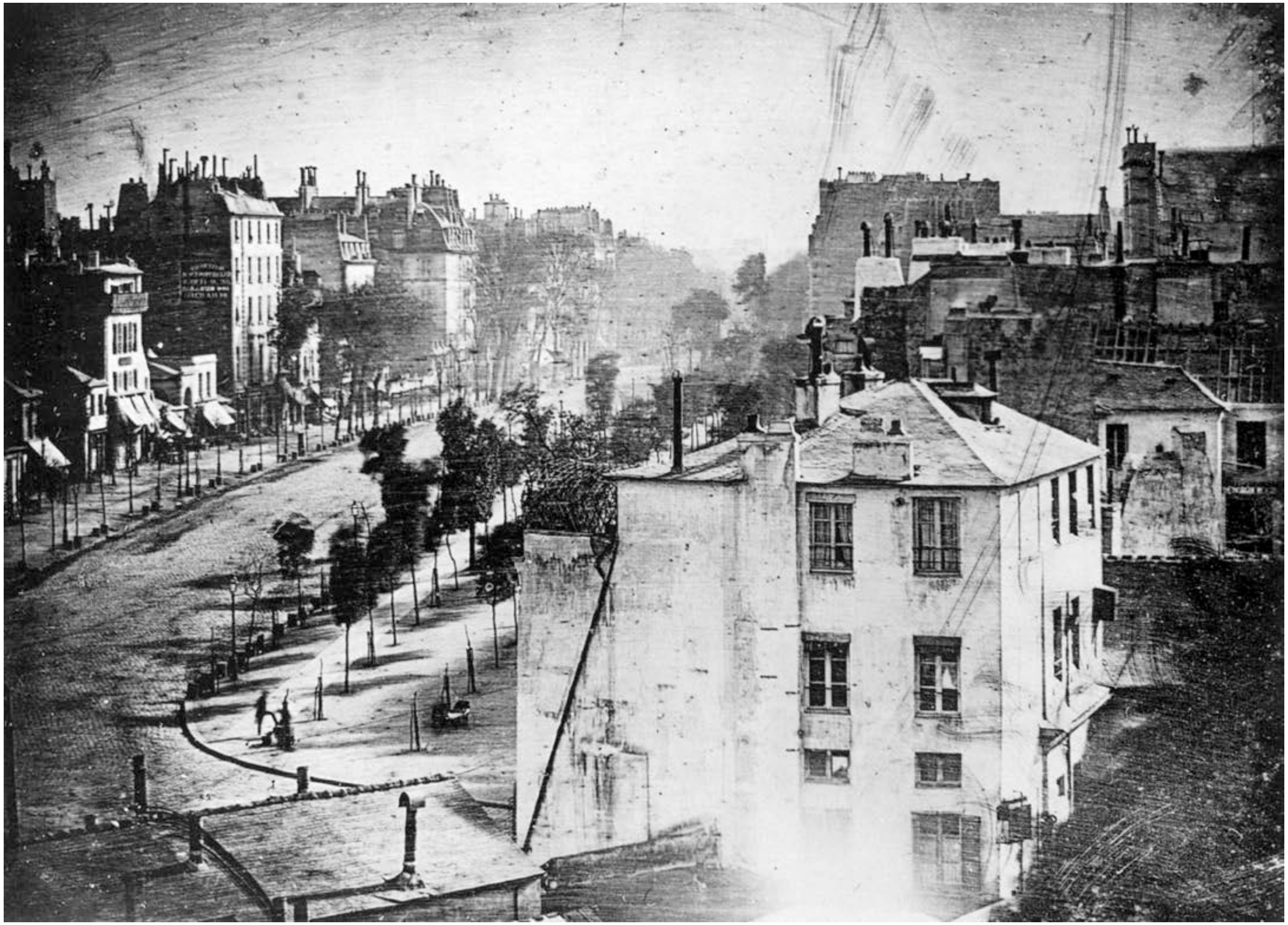
*La table servie*, Joseph Nicephore Niepce, 1822.

Heliography:

Metal with coating of bitumen, which hardens according to exposure to light.

# Daguerreotype: 1838





Louis Daguerre, 1838:

Copper plate with photo sensitive coating of silver + iodine vapor, developed using mercury vapor

Patent 1839

# Milestones of image acquisition and reproduction



- 1822: First photo
- 1839: Daguerreotype
- 1889: Roll film (George Eastman)
- 1895: Cinema (Brothers Lumière)
- 1908: Color photography (Brothers Lumière)
- 1920: Television (Baird, Farnsworth, Zworykin)
- TV in Germany:
  - 1939 about 500 private “Volksfernseher”
  - BRD starts TV 1952
  - ZDF 1963
- 1968-1973: Development from first image sensor based on discrete photo diodes to first commercial CCD camera of 100x100 pixels

# Pinhole camera



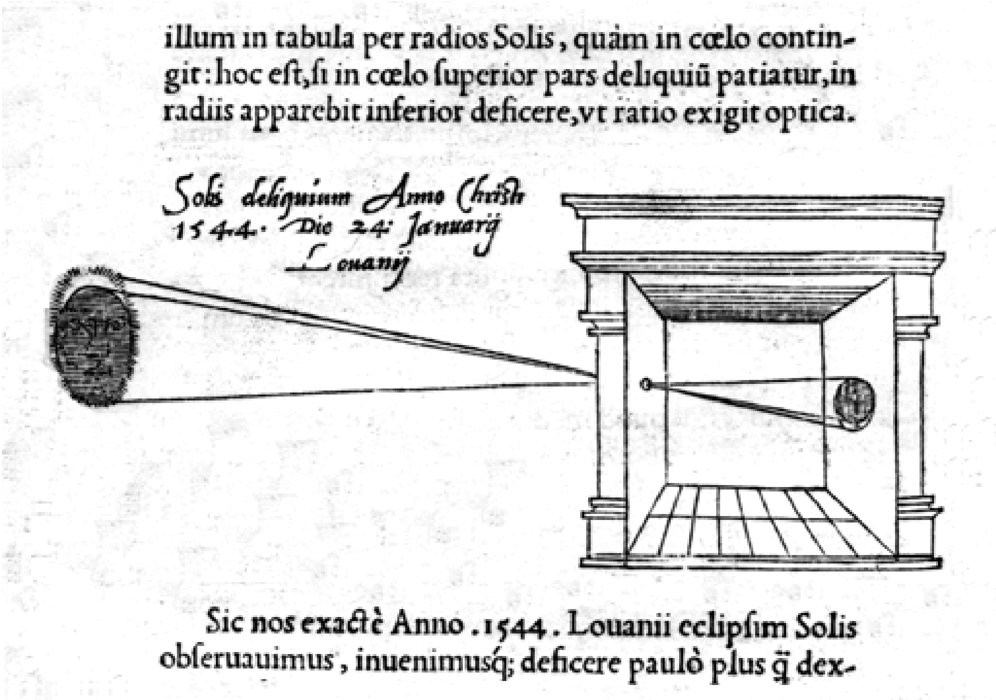
- Aristoteles:

  First description

- Da Vinci:

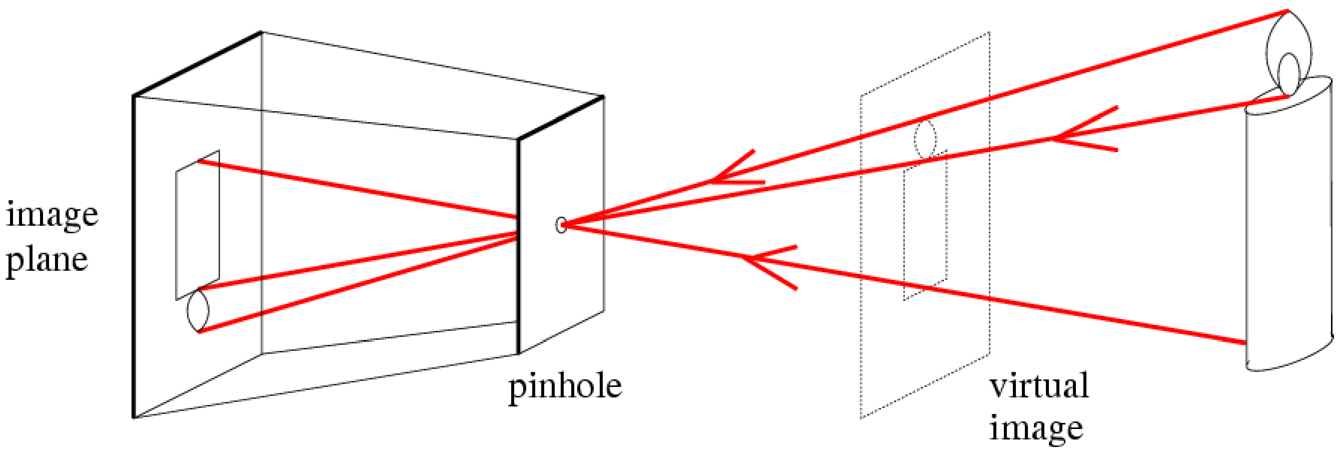
  Description, compares pinhole camera and eye, explains why images appear upside down (so he knew the idea of beams of light)

- We use the pinhole camera as an abstract camera model



# Pinhole camera

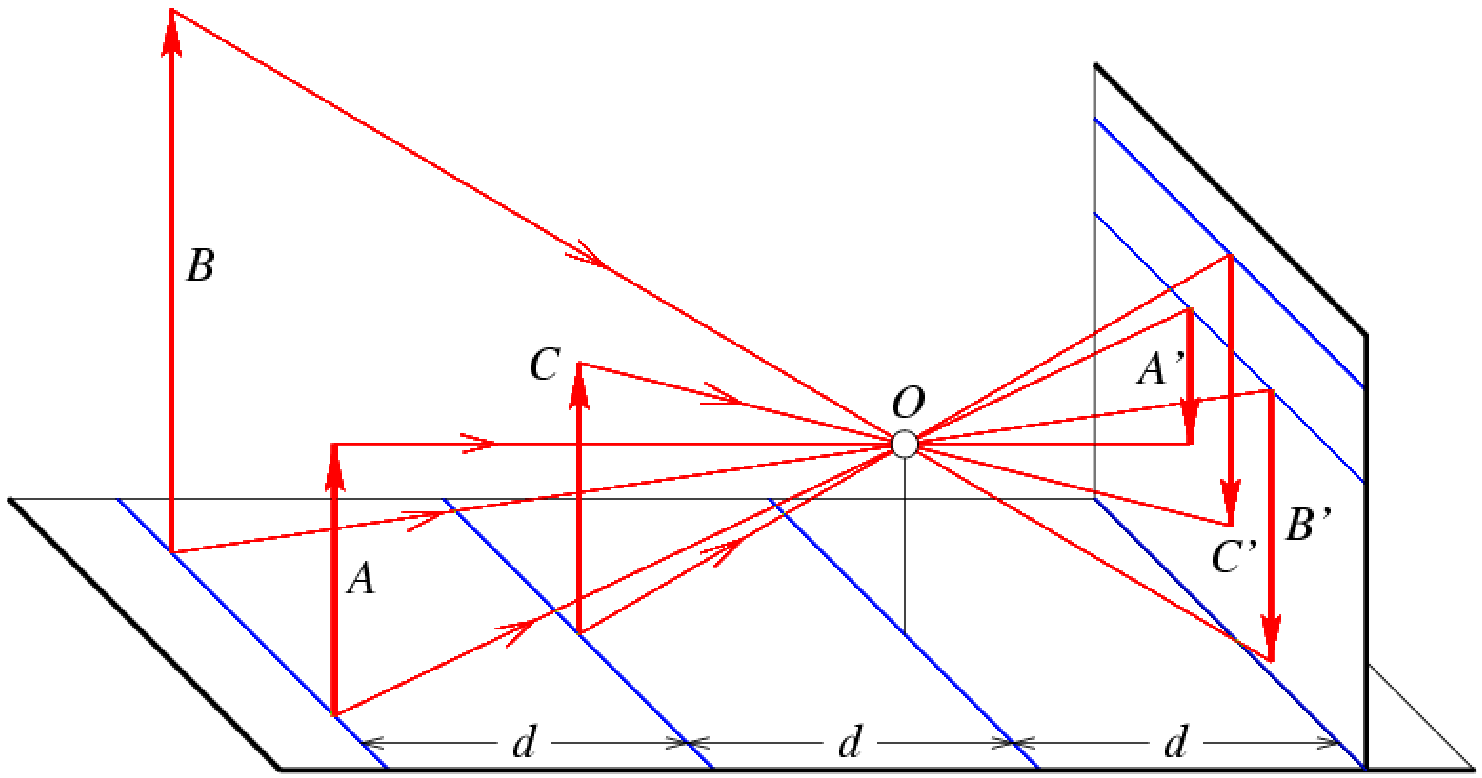
- Pinhole camera: Box with a small hole
- Object reflects light, light goes through the hole and hits the screen (opposite part of the box). The screen is the “image plane”.
- Why do we need the hole?
  - Without hole: Light reflected by *each* point of the object hits *each* point on the screen.
  - With hole: *Each* point of the object reflects light to *one* point of the screen.




[FP]

# Pinhole camera



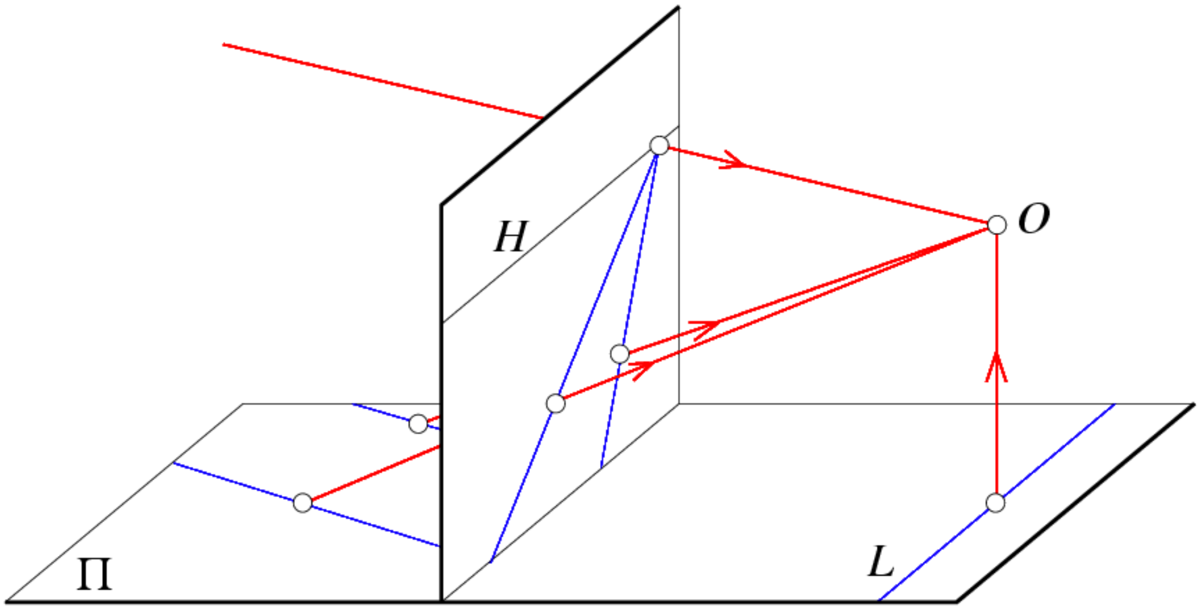



[FP]

The size of an object in the image plane is inverse proportional to the distance of the real object to the hole. *O* denotes the hole.

# Pinhole camera



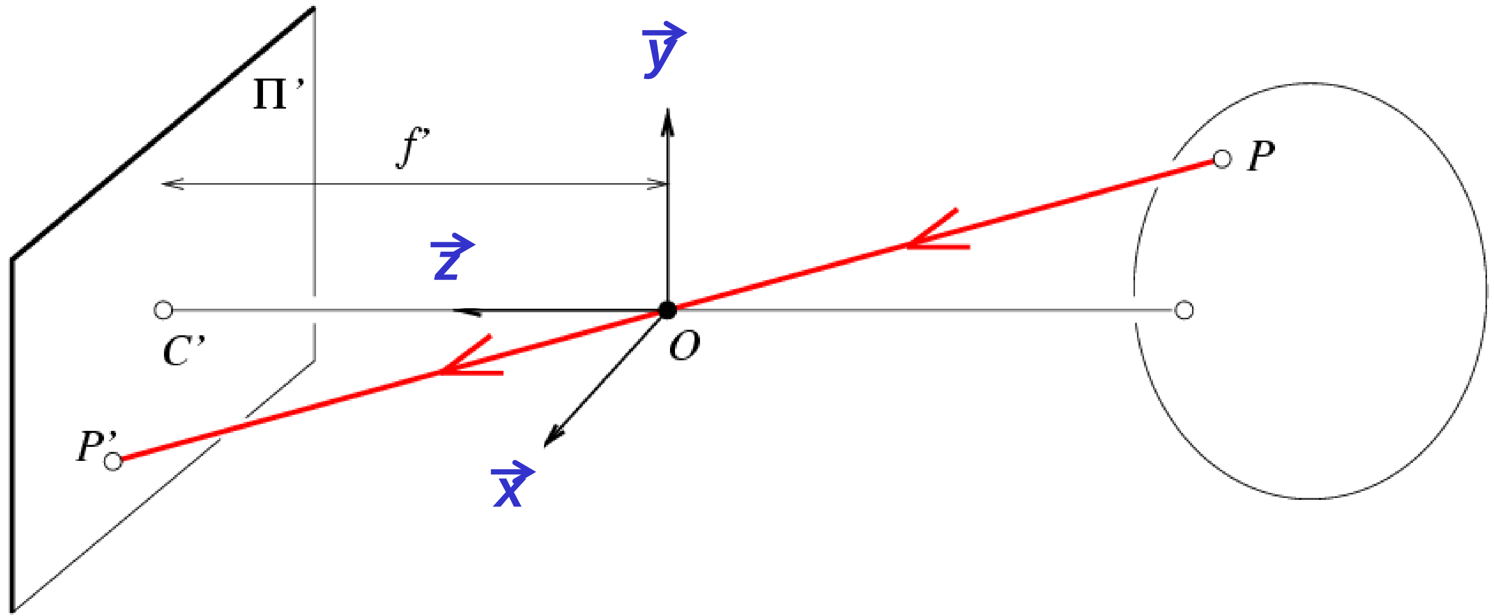



[FP]

- Alternative: Image plane (film plane) *H* is depicted *in front* of *O.*
- Parallel lines (left, in Π) are mapped onto lines which meet in a common point, called the *vanishing point* for this direction.
- Non-parallel lines end in different vanishing points.
- For all lines in a plane Π, the vanishing points are collinear. This line is called the *horizon H* for Π.

# Pinhole camera






[FP]

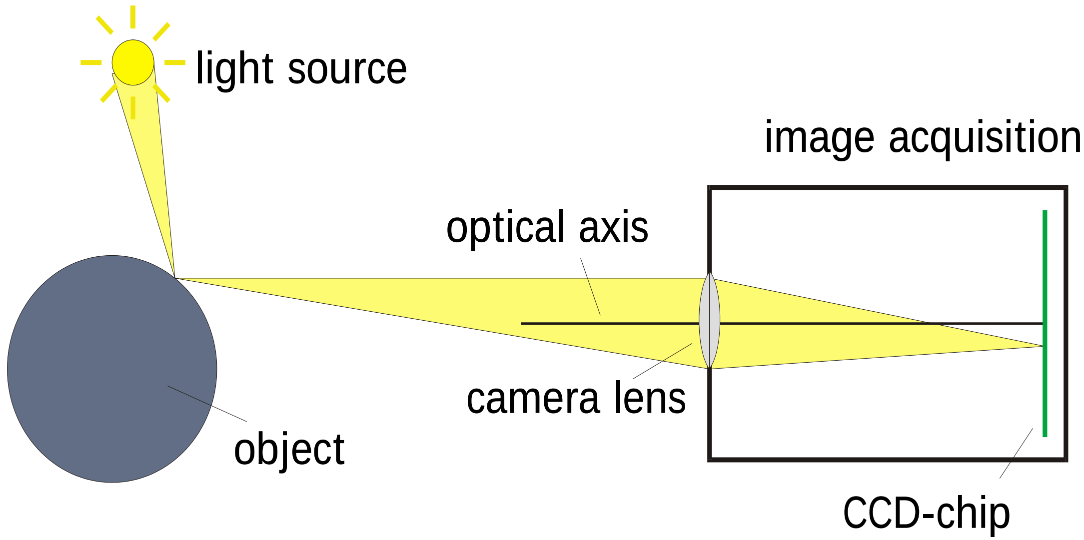
A point of an object in (3d-) space $P = (x, y, z)$ is mapped onto the (2d-) point $P'(x',y')$ in $\Pi'$ :

$$x' = f' \, x/z, \quad y' = f' \, y/z$$

Note in this coordinate system $z<0$ always holds for all objects visible to the camera.

# CCD-camera

- For a sharp image a small pinhole is required
- Small pinholes lead to dark images
- We need lenses or lens systems for sharp *and* bright images
- In a real camera a film or CCD replaces the screen.




[T]

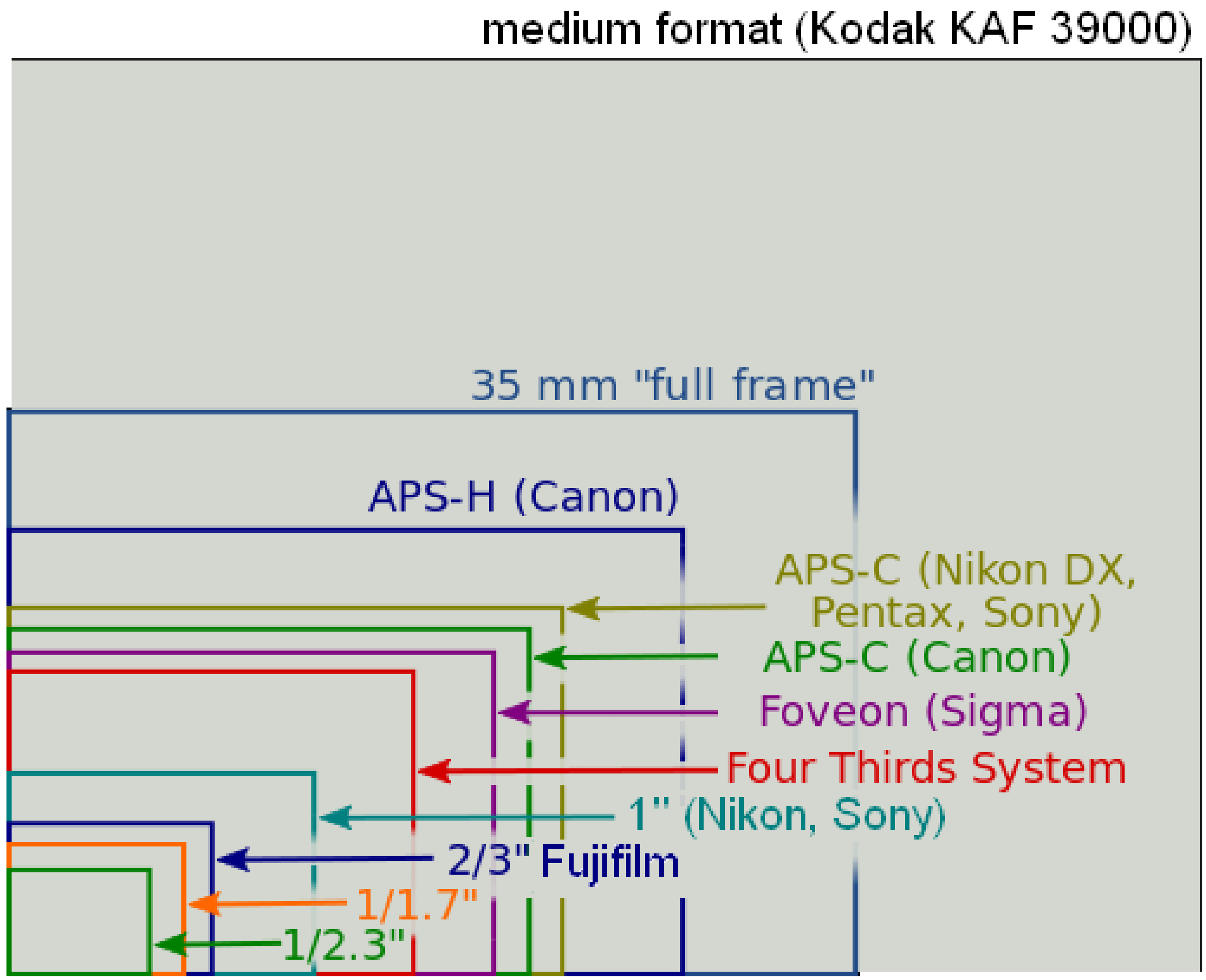
# CCD-camera




[J]

CCD-chip of an industrial camera

# Common image sensor formats

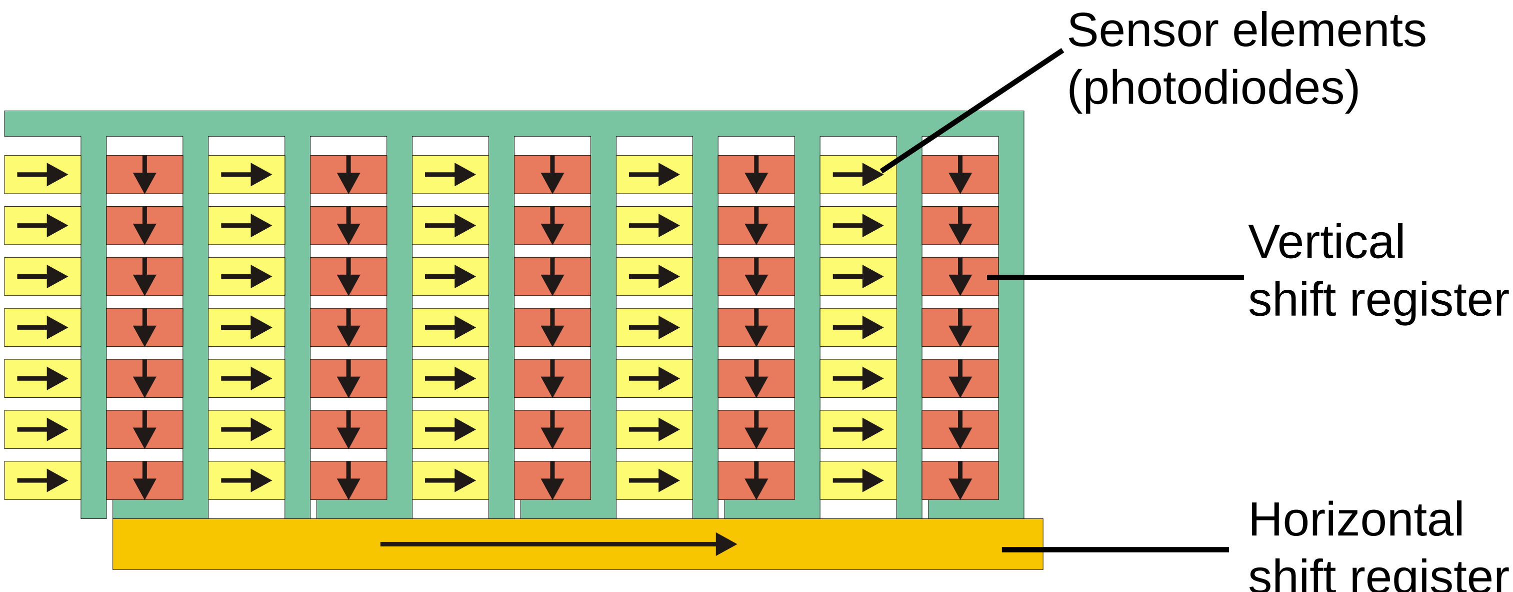



[W1]

# CCD-Chip



- CCD-chip: Light sensitive ***C****harge* ***C****oupled* ***D****evice*
- Typical pixel size 5−20µm



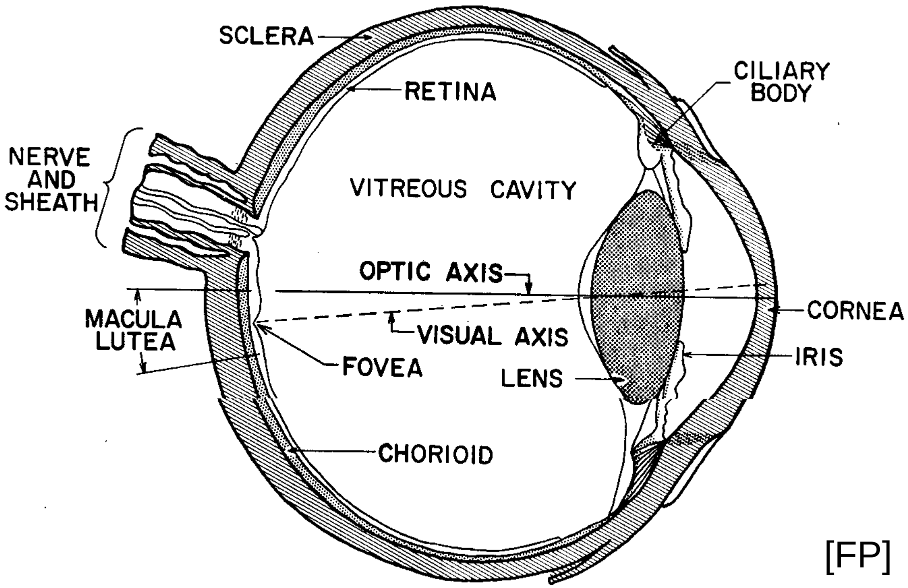

CCD Chip (interline transfer architecture)

[T]

# Human eye



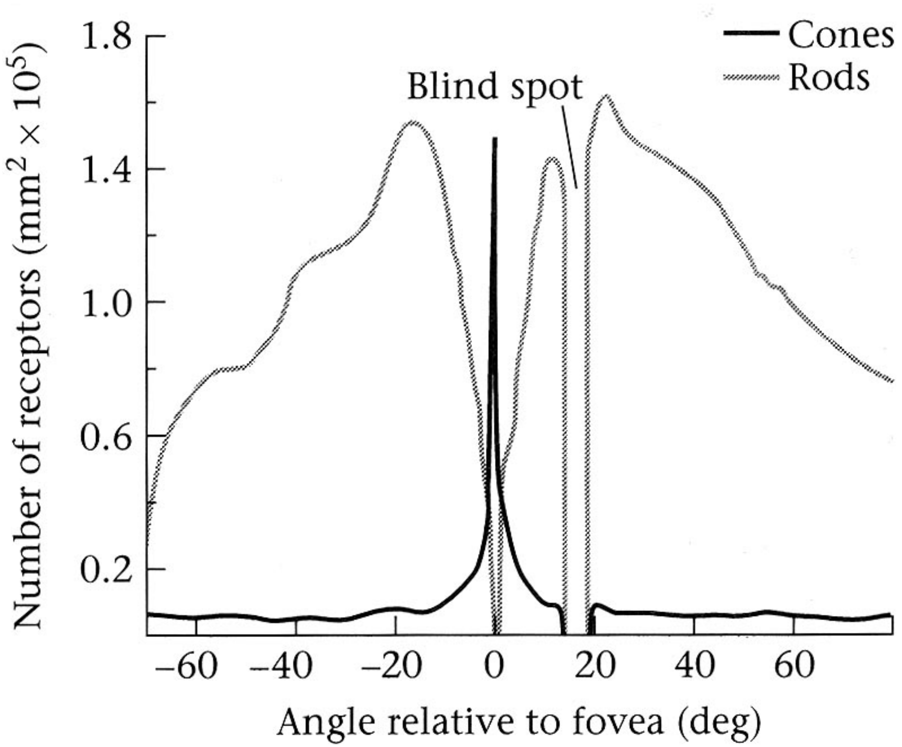
- Focusing by changing the shape of the lens
- Iris controls brightness
- Retina contains two types of receptors:
  - Cones
  - Rods for low light
- Fovea: Small region of high resolution, contains mostly cones, color vision
- Optic nerve: 1 million fibers




[FP]

# Human eye





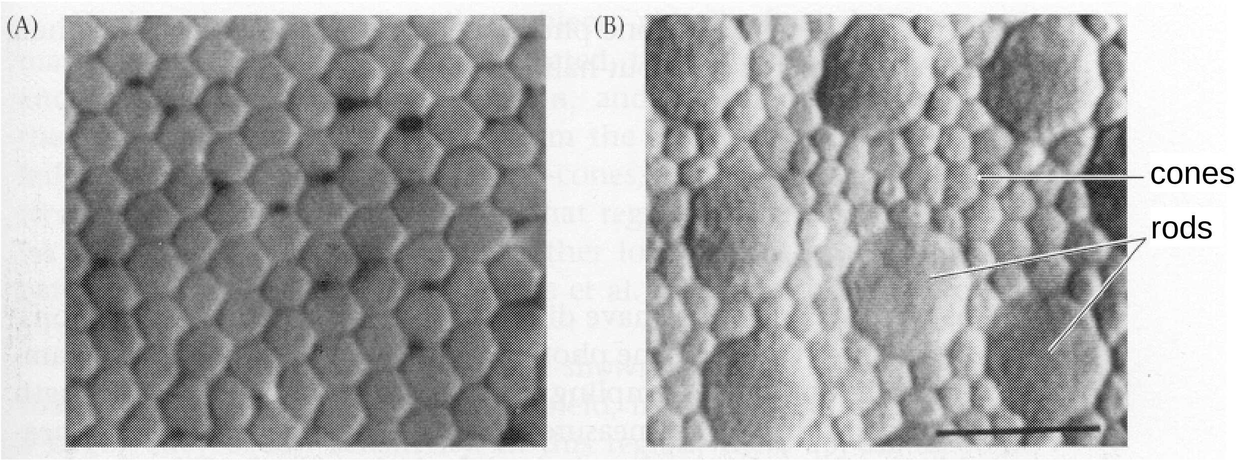

B. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., (1995).

Distribution of cones (fovea) and rods

Blind spot: Spot where optic nerve emerges (papilla)

# Receptors of the eye






B. Wandell, *Foundations of Vision*,
Sinauer Associates, Inc., (1995).

Left:  Cones in the fovea

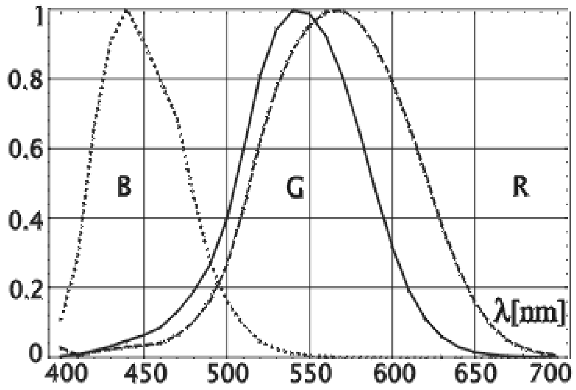
Right: Cones and rods in the periphery

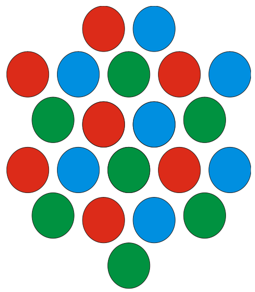
# Color



- Visible wavelengths: About 400-750 nm
- Eye contains three types of receptors with different spectral sensitivities
- Arranged side by side in the retina
- So we reduce the incoming spectrum to just three stimuli:

  $\infty$ - dim. input → 3-dim. representation



[T]

# CCD-camera: RGB color model



Each color is represented
by a triple (red, green, blue)

[T]

# CCD-camera: Color

- Problem: We need to record three color values for each image point
- Two construction principles:

1. 1-chip camera:
   - The sensor elements for R, G and B are arranged side by side on a single chip.
   - Colors are separated by a filter overlaid to the sensors
2. 3-chip camera:
   - A prism separates light of different wavelengths and directs it onto three chips

# 1-Chip camera

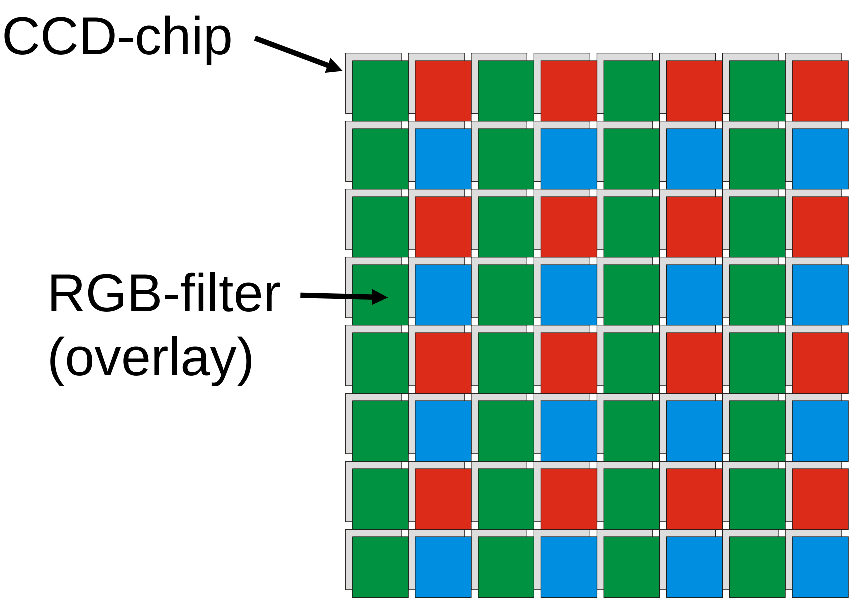


Pros:

- Compact
- Cheap

Cons:

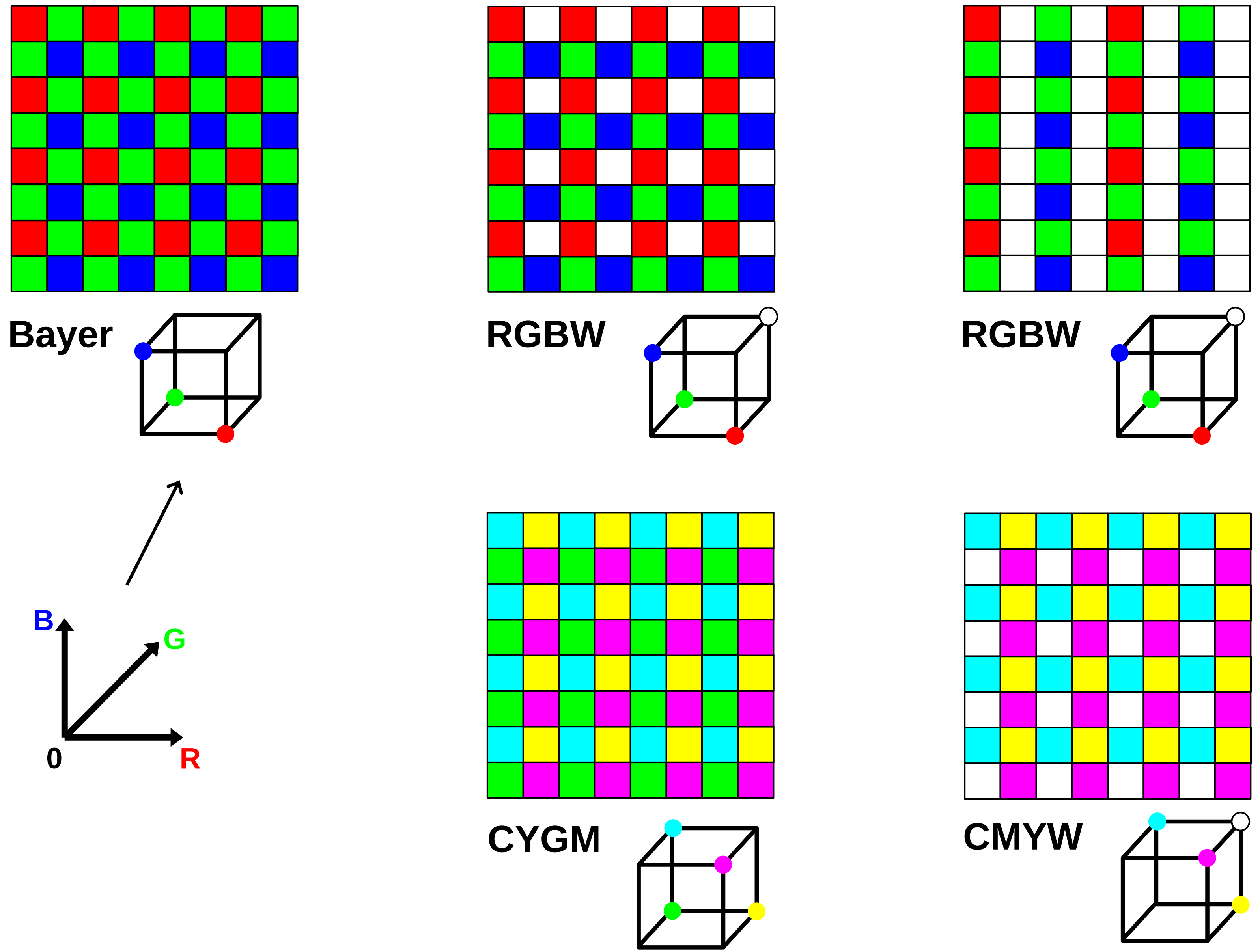
- Limits resolution
- R, G and B are not recorded at exactly the same location




[T]

# Different filter layouts



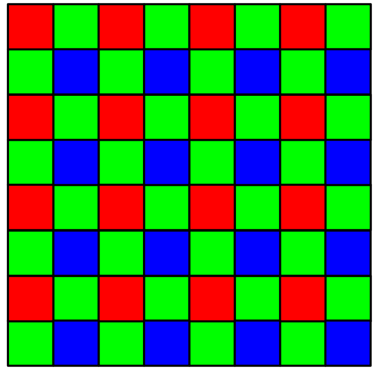



[H]

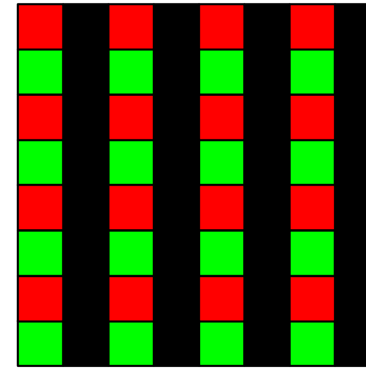
# Problems of interpretation



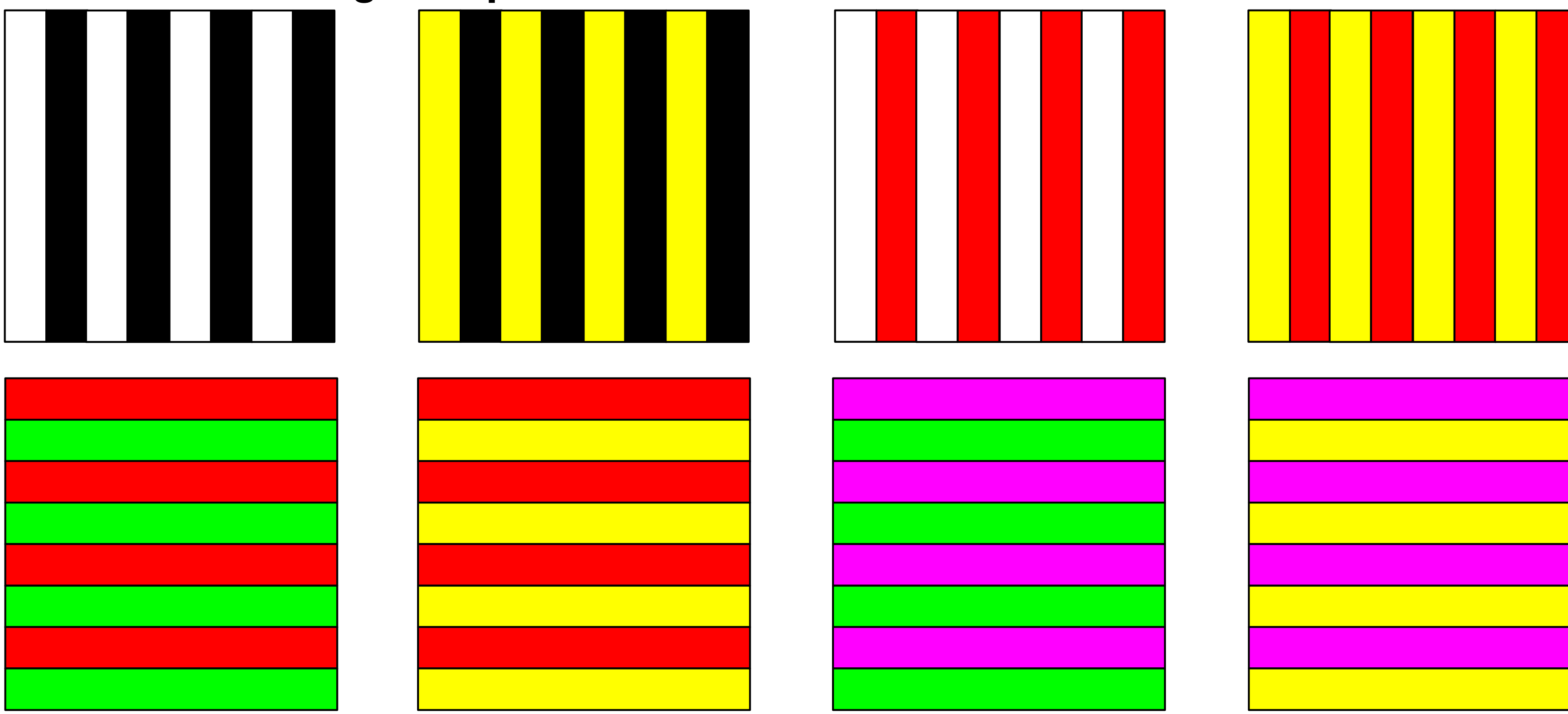
**If a Bayer-Filter**



**has this activation**



**then the following interpretations are consistent with the activation:**



# 3-chip camera

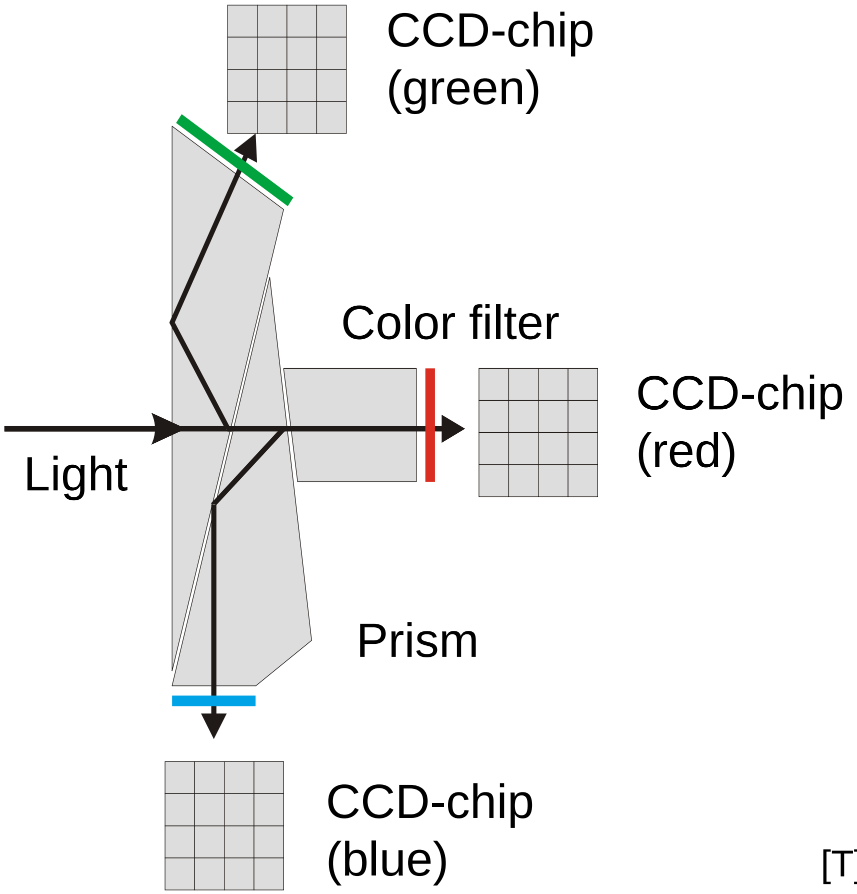


Pros:

- Better spatial resolution
- R,G,B are recorded at the same location

Cons:

- Bigger
- More expensive




[T]

# Image acquisition



- Photodiodes provide analog signal
- Camera digitizes the signal
- Transmission from CCD-Chip to the computer:
  - Analog (outdated, but still in use):
    - Camera converts digital representation to an analog signal according to the video standard
    - Computer re-digitalizes the video signal
    - Loss of quality, but still in use for compatibility
  - Digital (better)
- Transmission to
  - computer memory or
  - specialized image processing card

# Image representation

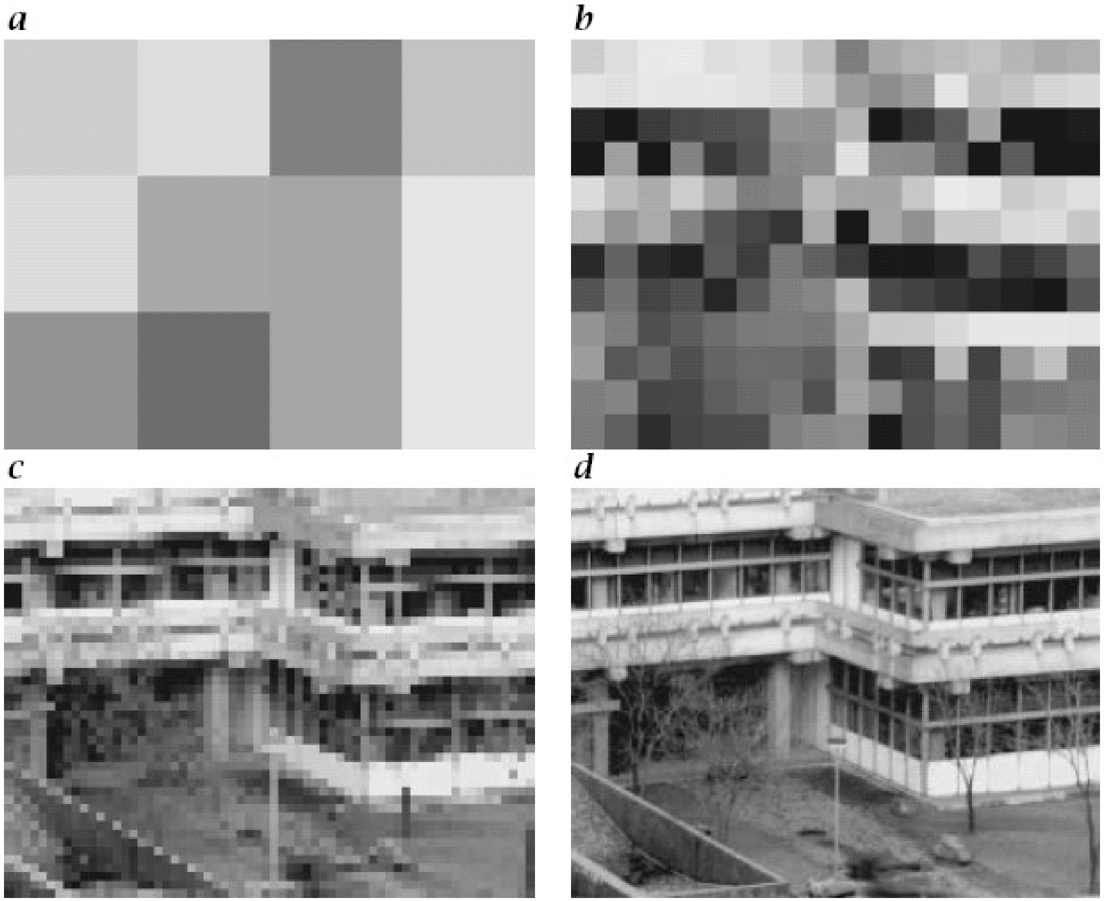


Threefold discretisation of the input

1. spatial (pixels)
2. color (or intensity only)
3. temporal (frames)

# Image representation: Spatial resolution

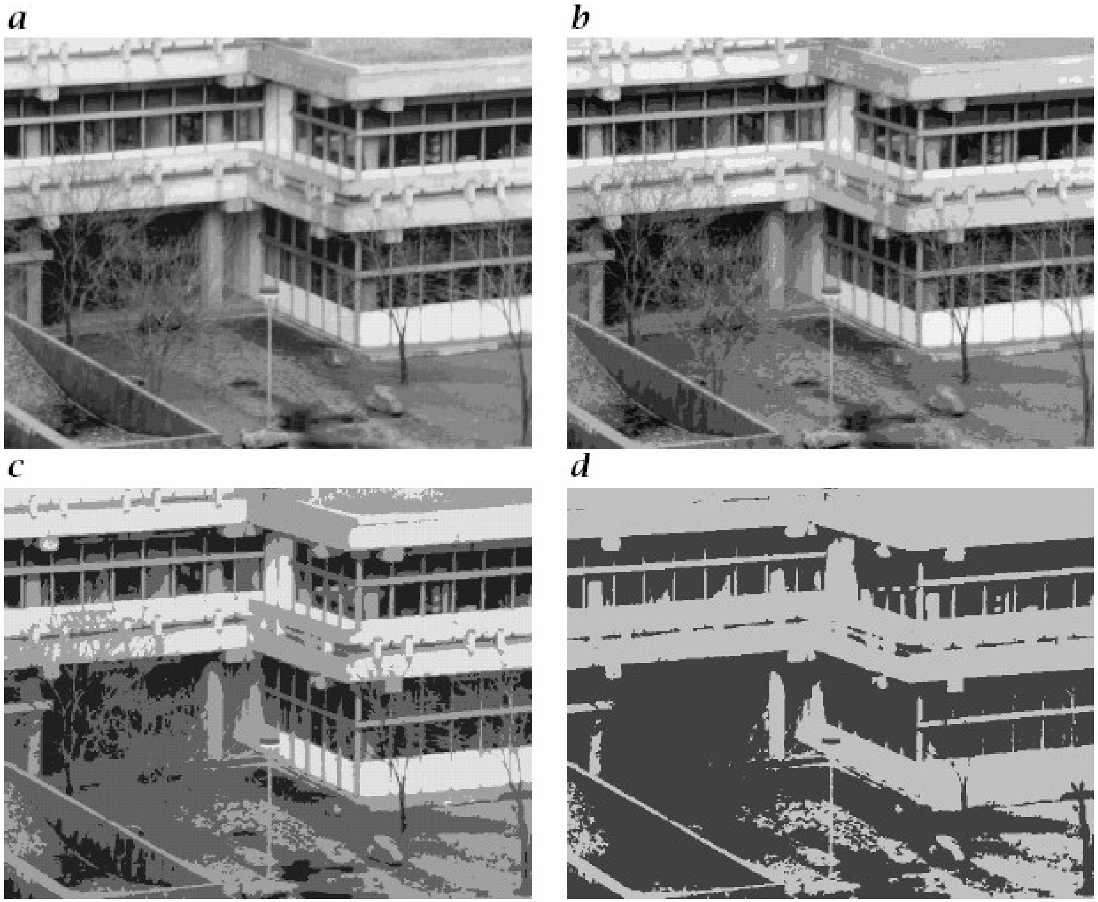





(a): 4x3, (b): 16x12, (c): 64x48, (d): 256x192 [J]

# Image representation: Contrast resolution





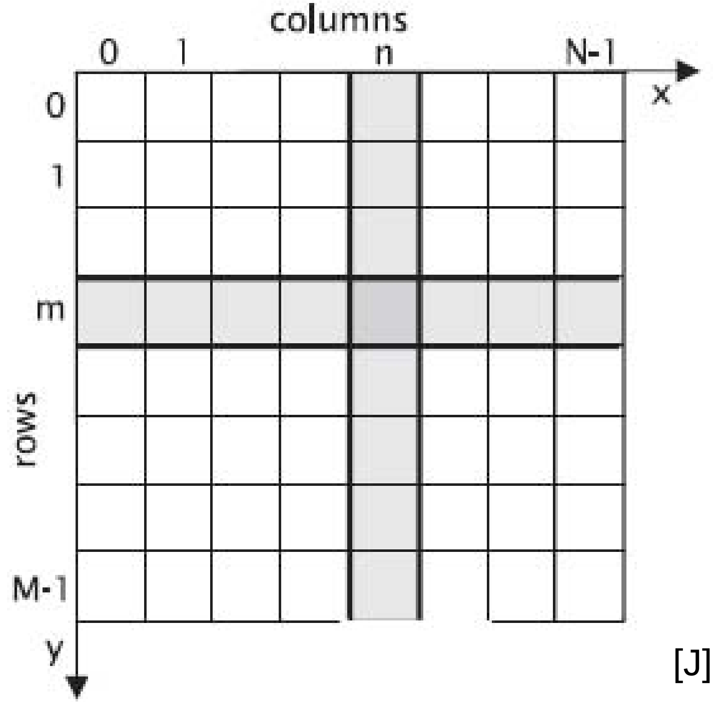

# gray values: (a): 16, (b): 8, (c): 4, (d): 2. [J]

# Image representation: Pixel



- Sensor is a 2d-array of photodiodes which map the intensity g(*x, y*) onto a matrix (spatial discretisation)

- Commonly, pixels have square shape ⇒ Definition of neighborhood is a problem

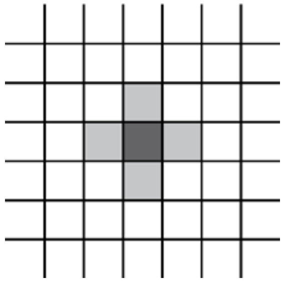


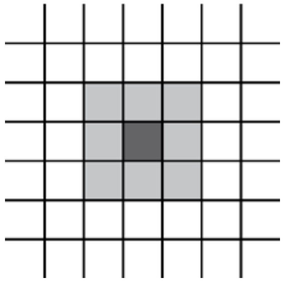

[J]

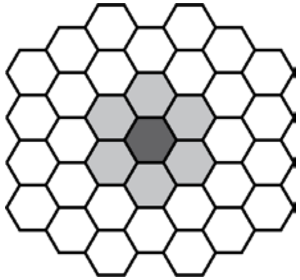
# Definition of neighborhood



- For square pixels two definitions of neighborhood present themselves:
  - 4-neighborhood: Only pixels with a common edge are neighbors.
  - 8-neighborhood: In addition, pixels with a common corner are neighbors.
- Neighborhood is well defined for hexagonal pixels.





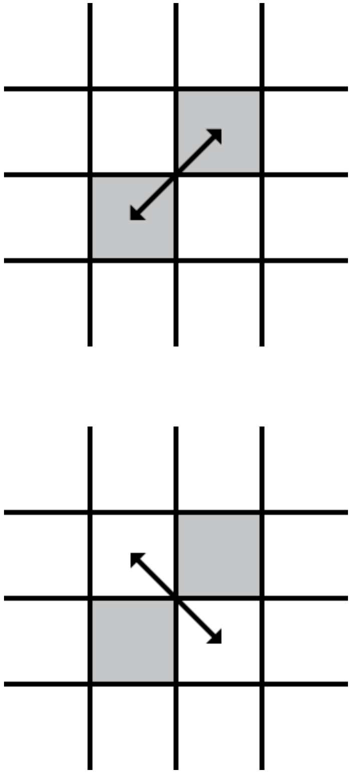


[H]

# Connectivity of objects



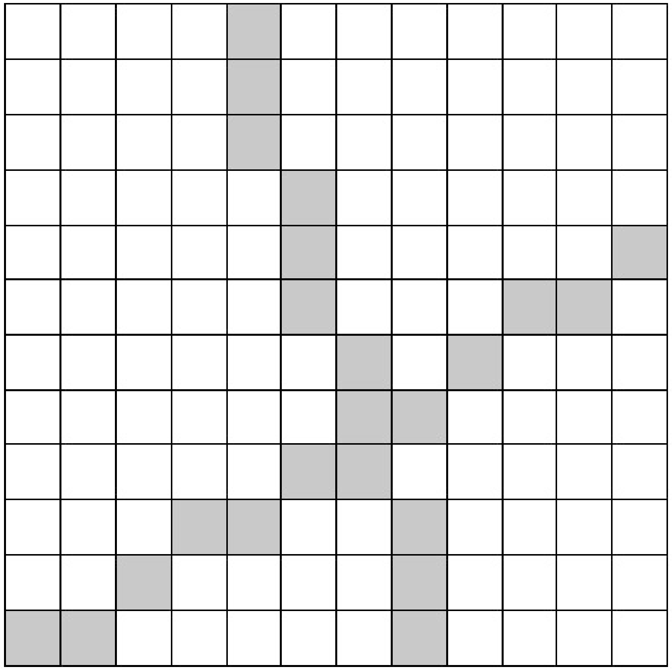
- Object is connected if each of its pixels has at least one neighboring pixel which is also part of the object.
- Problem: We have to use different neighborhood definitions for fore- and background.
- E.g., let the gray pixels denote the object (foreground). Assume to obtain a connected object an 8-neighborhood is mandatory. Then a 4-neighborhood must be assumed for the background (white pixels). Otherwise a connected background would extend across a connected object.
- By contrast, if the background is to be connected, we must assume an 8-neighborhood for the background and a 4-neighborhood for the foreground. So the object is no longer connected.
- Using a hexagonal grid the problem does not arise.



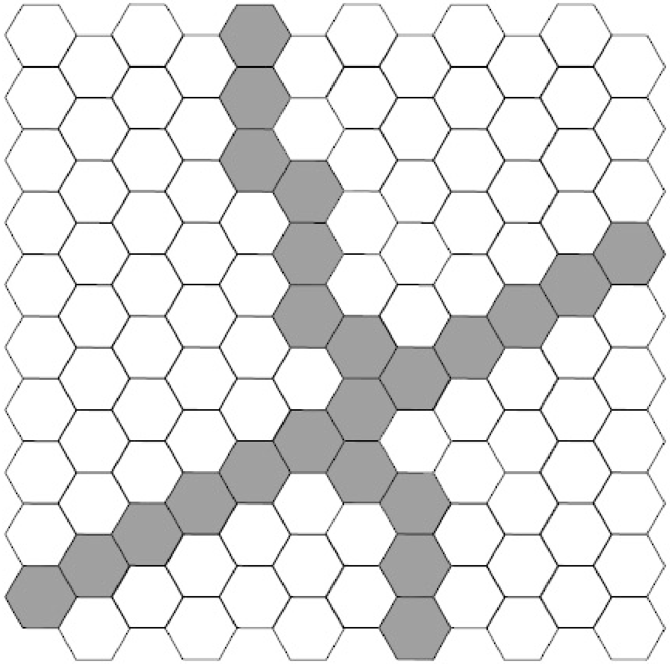
[H]

# Connectivity





Two intersecting lines
on a square grid



Two intersecting lines
on a hexagonal grid

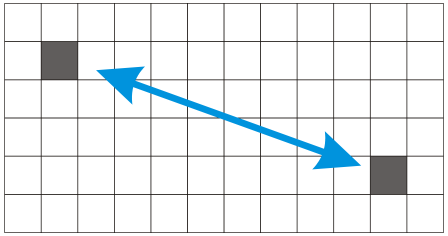
[H]

# Distance measures

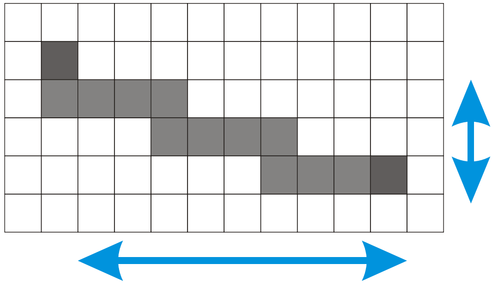
$$D_e(x_1, y_1; x_2, y_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$ Euclidean distance

$$D_4(x_1, y_1; x_2, y_2) = |x_1 - x_2| + |y_1 - y_2|$$ Cityblock distance

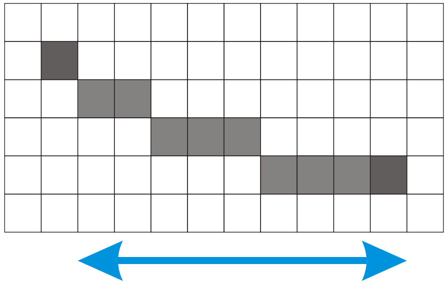
$$D_8(x_1, y_1; x_2, y_2) = \max(|x_1 - x_2|, |y_1 - y_2|)$$ Chessboard distance



Euclidean distance



Cityblock distance



Chessboard distance [T]

$D_4$ / $D_8$ is the smallest distance between two pixels under 4- / 8-neighborhood

# Loss of information

Two types of loss of information:

- Deterministic loss
- Stochastic loss

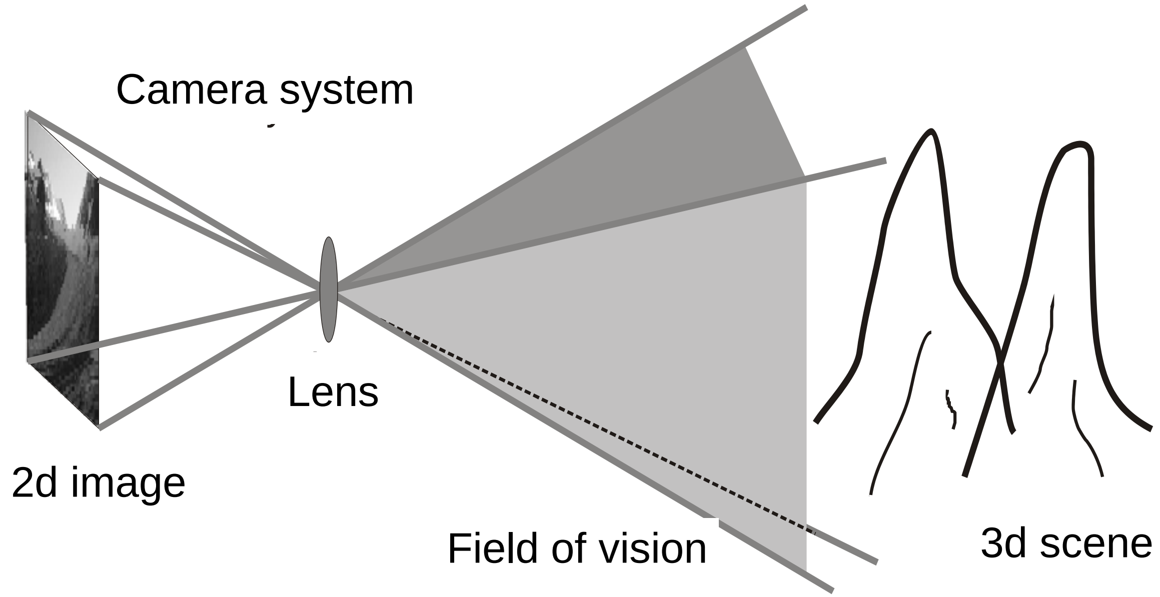
Most important causes:

- Projection and sampling
- Bad camera parameters:  Over/underexposure, focus
- Motion blur
- Noise

# Loss of information by projection



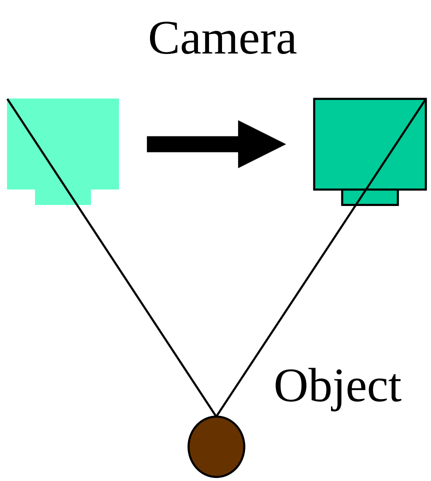
- Points outside the field of vision ("pyramid of sight") are lost
- Occluded points are lost
- Depth information is lost

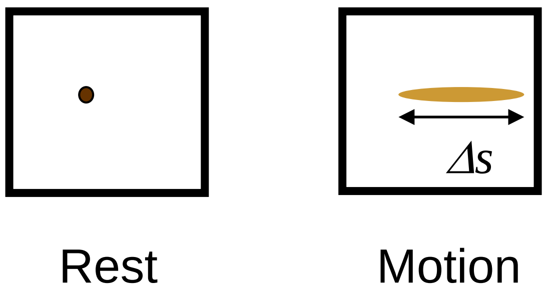



[T]

# Loss of information by motion blur

- Let camera move by $\Delta s$ during exposure time $\Delta t$.
- Thus a point $p$ of the object is mapped onto several points on the CCD-chip.
- Let the intensity of the point be $h$ for a still camera
- For a moving camera the intensity is only $h/\Delta s$.
- If $\Delta s$ is identical for all points on the CCD-chip, motion blur can be described by a convolution (linear filtering, later).



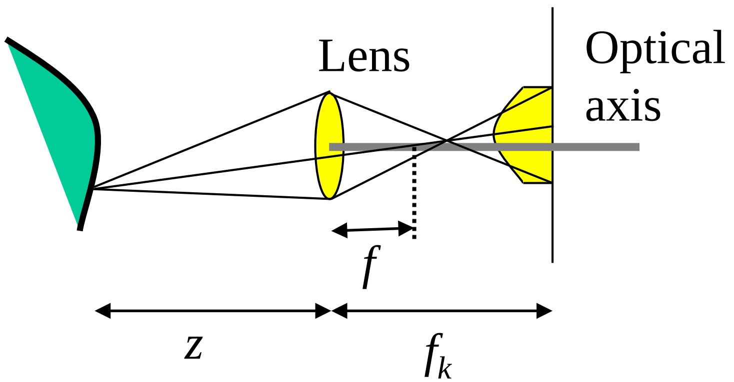




[T]

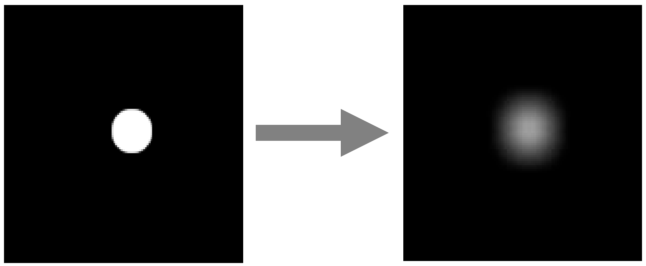
# Loss of information by focus blur



- Blur depends on the distance *z* between lens system and object and parameters of the lens system.



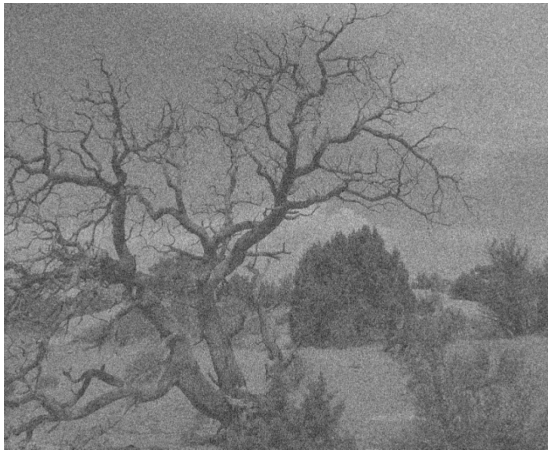

- Focus blur can be detected using a point like test object.



[T]

# Stochastic loss of information

- Noise: non deterministic
- Types of noise:
  - Quantum noise
  - Impulse noise
- We need the probability that a pixel is disrupted in a certain way



[T]

# Gaussian noise



- Quantum noise: Photons from an object hit the CCD chip only with a certain probability, in addition, there is a spatial Possion-distributed uncertainty.
- Image $g(x,y)$ is composed of two components:
  a. Photons that follow – with sufficient accuracy – a classical trajectory (straight line) from the object to the image: $f(x,y)$.
  b. Photons where quantum effects can not be neglected: Noise $\eta(x,y)$.

$$g(x,y) = f(x,y) + \eta(x,y)$$

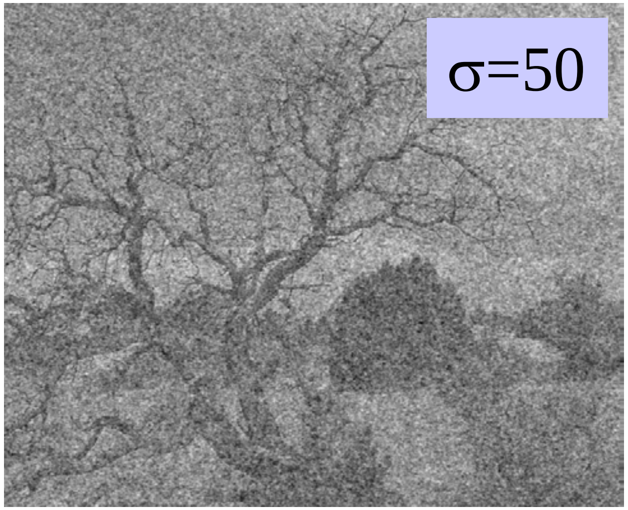
- Approximation: The disruption of the “classical” $f(x,y)$ by quantum effects is spatially uncorrelated (on the CCD chip) and can be described as a Gaussian distribution

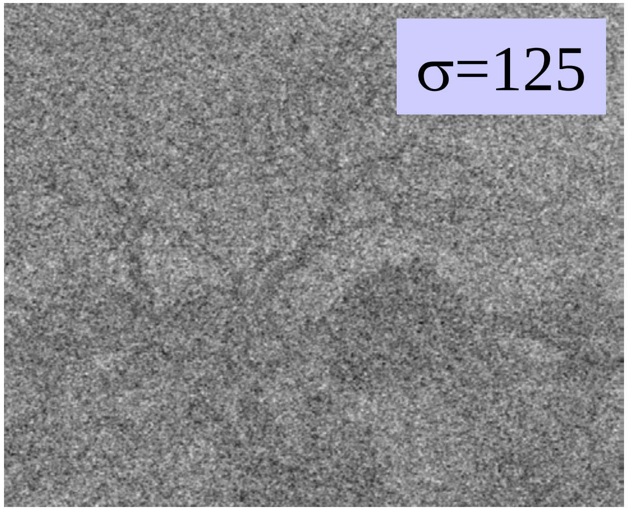
$$P(\eta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\left(\frac{\eta}{\sigma}\right)^2\right)$$

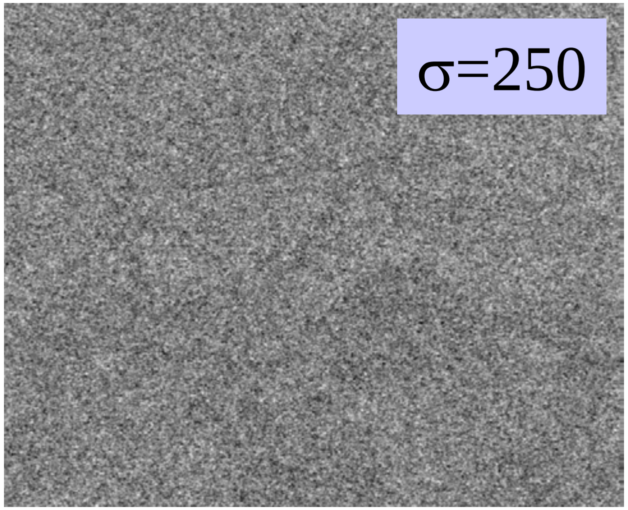
- Variance $\sigma^2$ reflects influence of the quantum noise.

# Gaussian noise






[T]

# Signal to noise ratio



- *SNR*: Signal to noise ratio is “strength” of the noise compared to “strength” of the signal.
- **Signal**: Difference between the gray values of object and background
- Alternative definition if object is unknown or undefined:
  - Background = Gray value 0
  - Object = Maximum or average gray value
- **Noise** is given by standard deviation $\sigma$.
- Thus we have two definitions (for an $M$x$N$ image):

$$SNR_{max}(f) \quad = \quad f_{max} / \sigma,$$

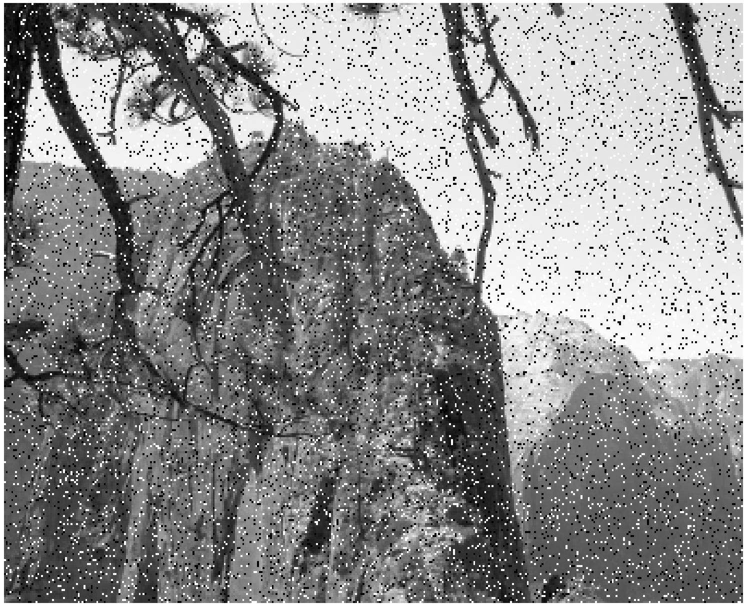
$$SNR_{avg}(f) \quad = \quad 1 / MN \cdot \sum_{m=0\ldots M-1} \sum_{n=0\ldots N-1} f(m,n) / \sigma.$$

# Impulse noise



- Isolated pixels are disrupted.
- Disruption is maximal but independent of the correct gray value

  (i.e. the disrupted pixel is either black or white – “salt-and-pepper noise”)



[T]

# Other Image Acquisition Systems

# Other image acquisition systems



- The results of non-standard image acquisition can often still be represented by gray value images.
- In this case the gray value has an interpretation other than luminance. Examples:
  - X-ray images
  - Reconstructed layers of 3d scans
  - Reflection time images
  - Depth images

# X-ray images



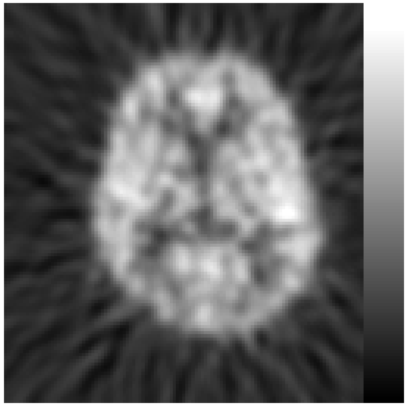
[T]

- Monochrome
- Semantics: Projection of X-ray absorption
- High resolution (> 4 megapixel)
- Range of values:  8–12 Bit
- Problem:  Lacking depth information

# Reconstructed layers of 3d scans

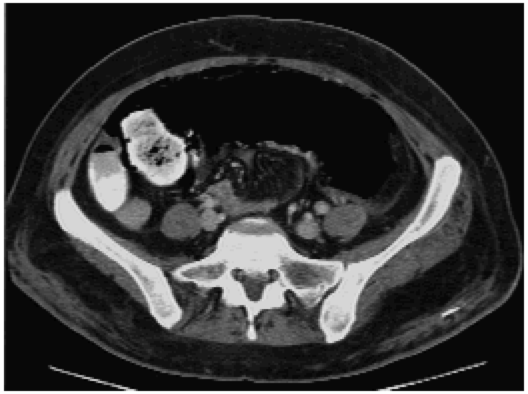


- One layer of 3d scan
- Monochrome
- Semantics: depends
- Small to medium resolution (0.02-0.25 megapixel)
- Range of values: 8-12 Bit
- Problems: Artifacts, noise



PET (positron emission tomography) shows concentration of injected radioactive isotope (tracer) attached to a biologically active molecule. Indicates e.g. glucose uptake.

MRI (magnetic resonance imaging) provides good contrast between different soft tissues of the body.

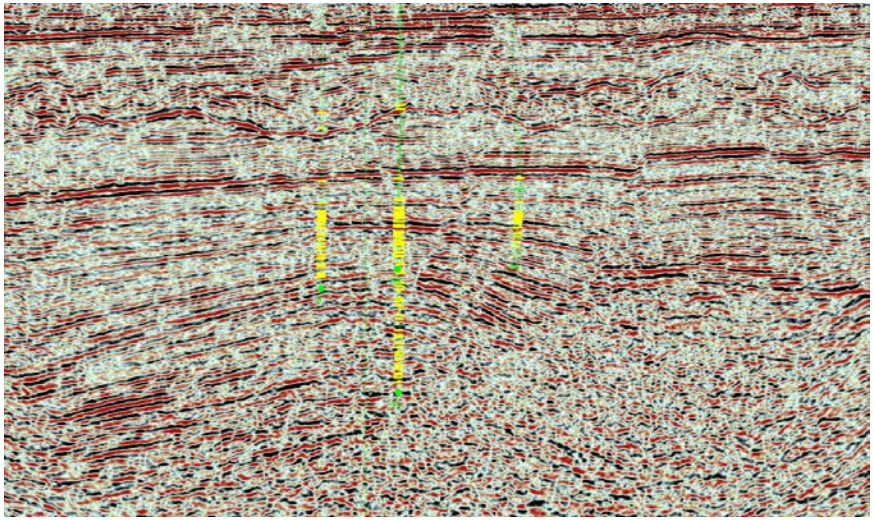


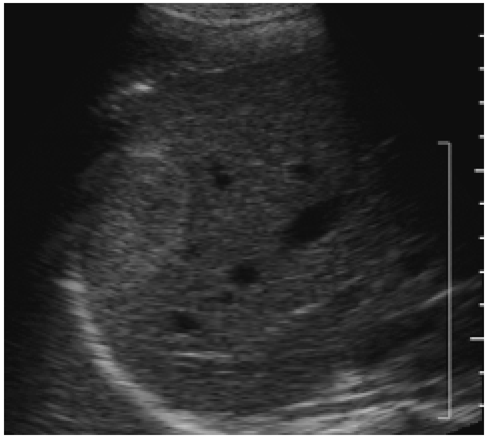
[T]

CT (X-ray computed tomography), computed from 2d X-rays, shows absorption

# Reflection time images







[T]

- Monochrome
- Semantics: Time to receive an echo
- Varying resolution
- Range of values: 8 Bit
- Problem: Artifacts

# Depth maps



[T]

- Monochrome
- Semantics: Distance to camera
- Resolution: Equal to digital camera
- Range of values: 8-16 Bit
- Problem:  Depth errors.

# Literature and sources


[T] Klaus D. Tönnies, *Grundlagen der Bildverarbeitung*, Pearson Studium, 2005.

[J] Bernd Jähne, *Digitale Bildverarbeitung,* Springer, 2005.

[FP] David Forsyth, Jean Ponce, *Computer Vision: A Modern Approach,* Prentice Hall, 2002.

[SS] Linda G. Shapiro, George C. Stockman, *Computer Vision,* Prentice Hall, 2001.

[BK] Henning Bässmann, Jutta Kreyss, *Bildverarbeitung Ad Oculos*, Springer, 2004.

[L] David Lowe, Slides, http://www.cs.ubc.ca/~lowe/425/.

[A] *Artexplosion Explosion® Photo Gallery*, Nova Development Corporation, 23801 Calabasas Road, Suite 2005 Calabasas, California 91302-1547, USA.

[C] Corel GALLERY™ Magic 65000, Corel Corporation, 1600 Carling Ave., Ottawa, Ontario, Canada K1Z 8R7.

[W1] From www.wikipedia.org, license CC BY-SA 3.0, 3.2.2016.

[H] Copyright Gunther Heidemann, 2008.