Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Computer Vision - Image Fundamentals

WS 2019/2020

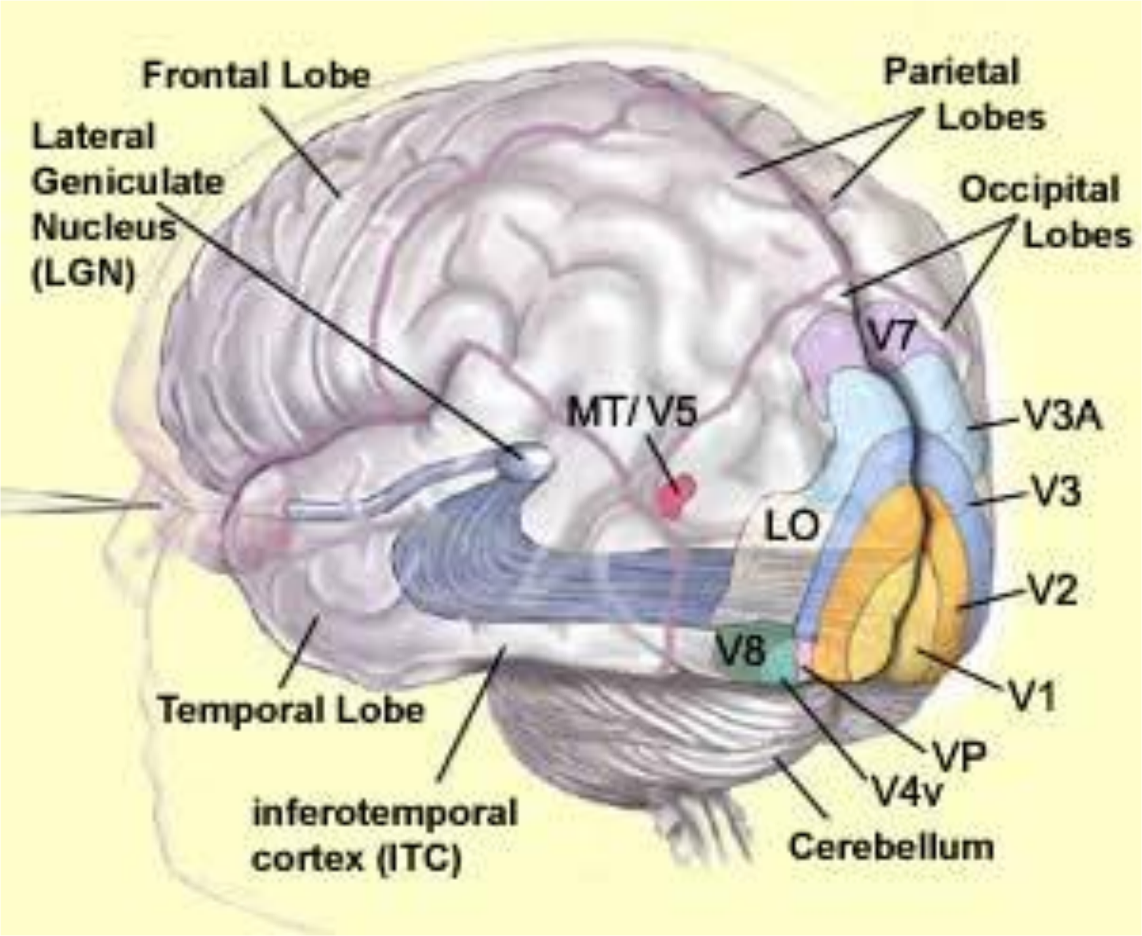
Prof. Dr. Simone Frintrop

Computer Vision Group, Department of Informatics
University of Hamburg, Germany

Content

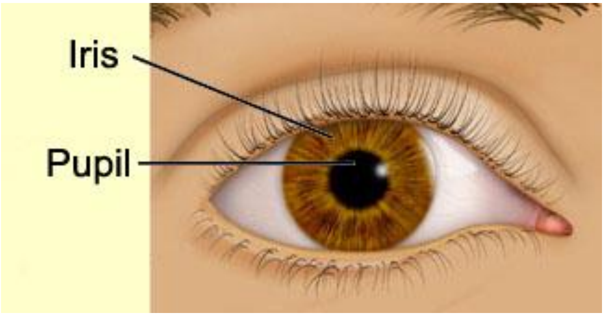
- The human visual system
- The electromagnetic spectrum as source of images
- Image acquisition and digitization
- Image representations
- Frequencies in images
- Noise in images
- Image processing domains (spatial vs transform domain)
- Color Models

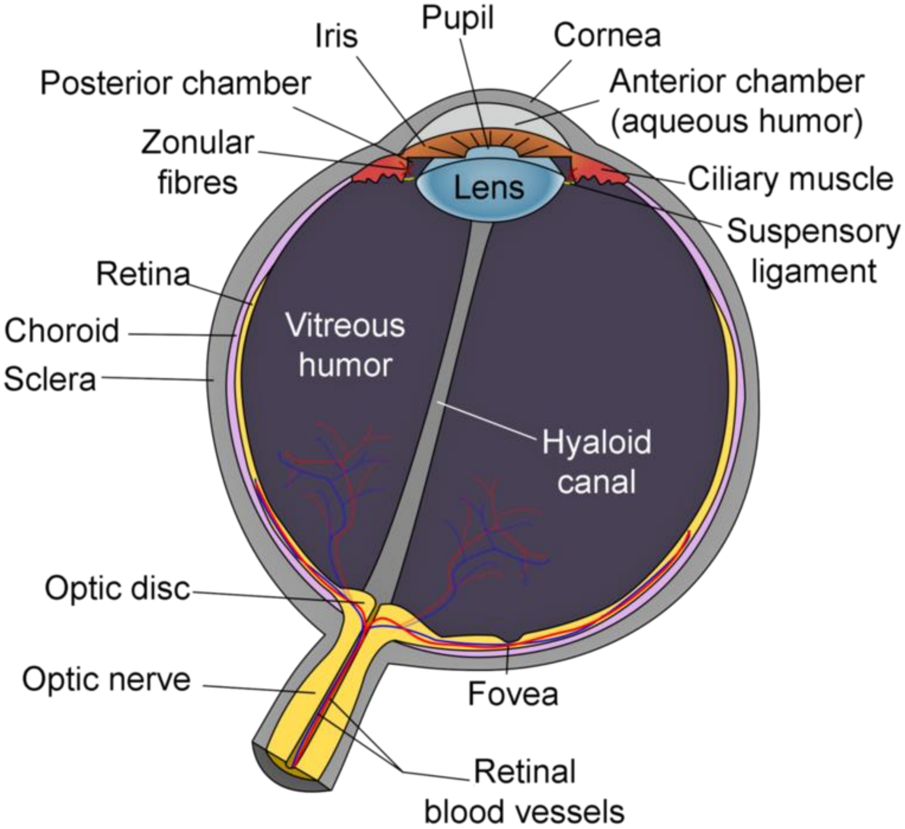
The Human Visual System



[http://thebrain.mcgill.ca/]

The Human Eye

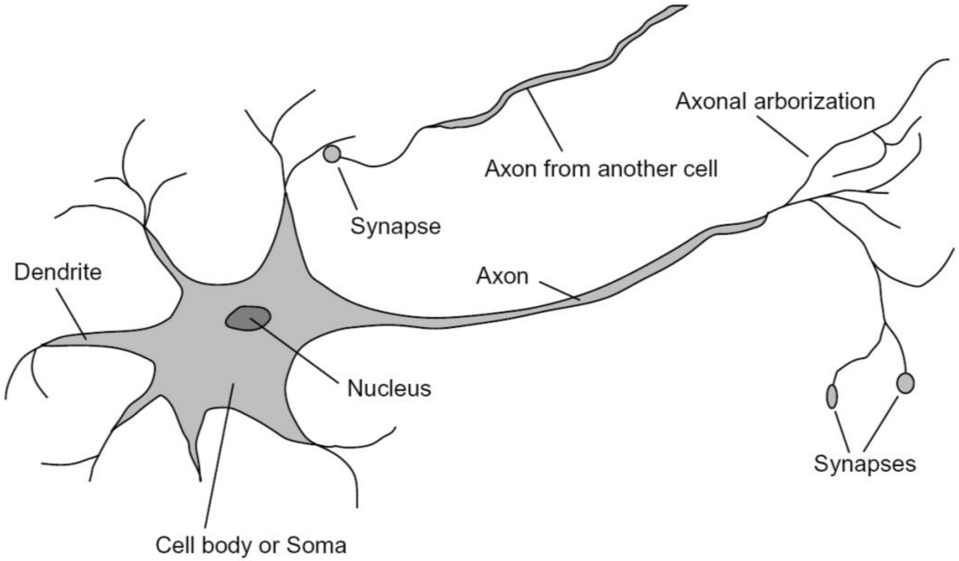




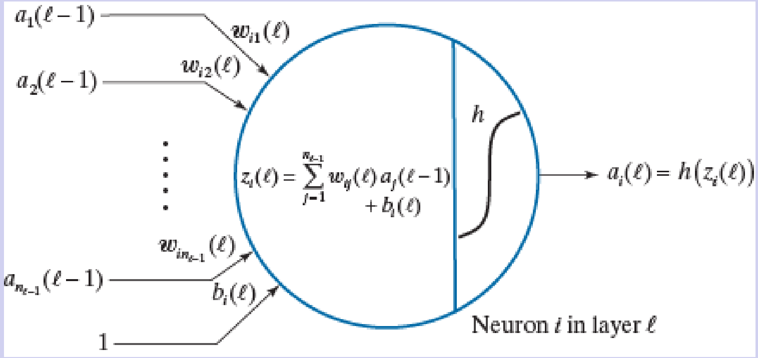
[Images: http://thebrain.mcgill.ca; http://en.wikipedia.org/wiki/Visual_system]

Neurons

- Basic units of the brain: neurons
- Neuron: An electrically excitable cell that processes and transmits information
- Neurons can connect to each other to form neural networks



Outlook: computational models of such neurons form the basis of neural networks (deep learning), see lecture on classification

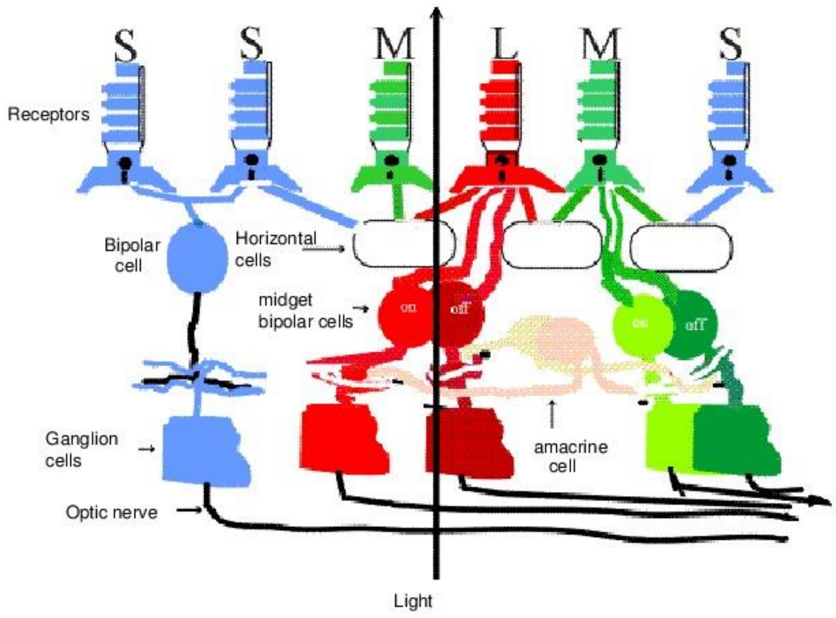


[Gonzales/Woods 2017]

Retina

The retina contains several cells (neurons) which process the incoming light:

- Photoreceptors (rods/cones)
- Bipolar, horizontal, amacrine cells (less important for CV)
- Ganglion cells

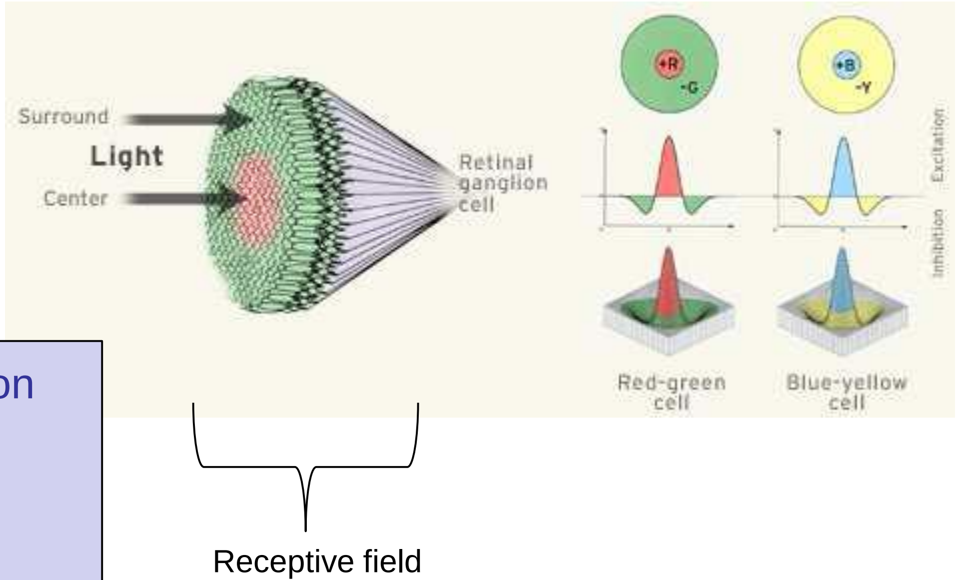


The retina

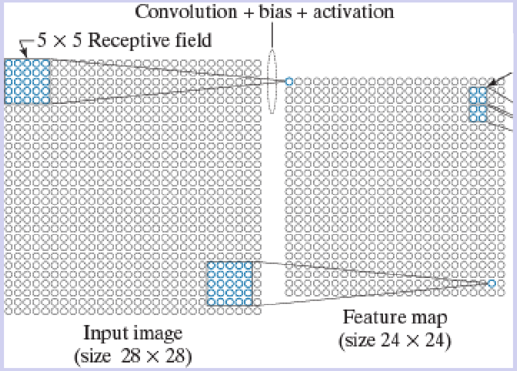
[Image: Kaiser 1996]

Retina: Receptive field

- Receptive field (RF) of cell: spatial area in visual field that has influence on the cell output



Outlook: „receptive field“ is a common term in artificial neural networks to denote the input area for a neuron



[Image http://www.webexhibits.org/colorart/ganglion.html]

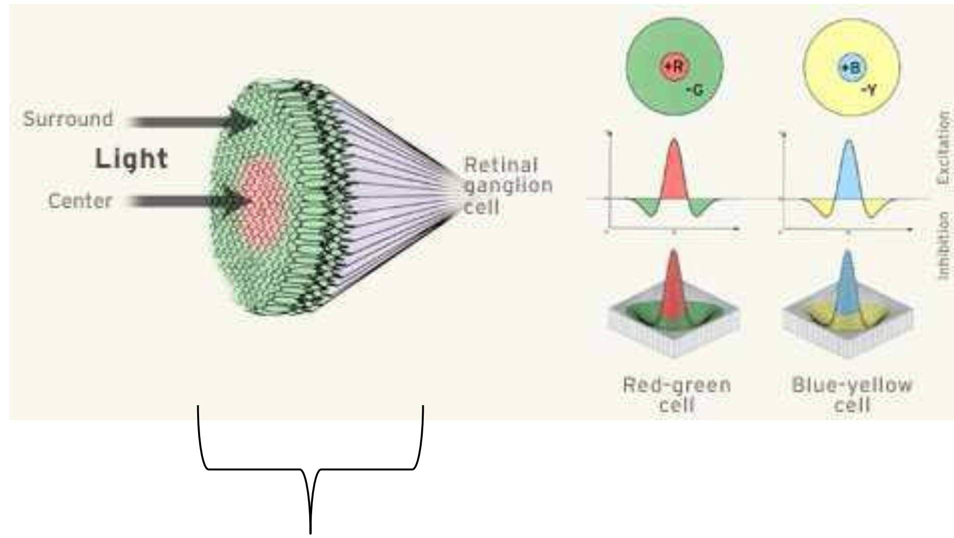
Retina: Ganglion cells

Retinal ganglion are especially important for CV, they can be modelled well by digital filters

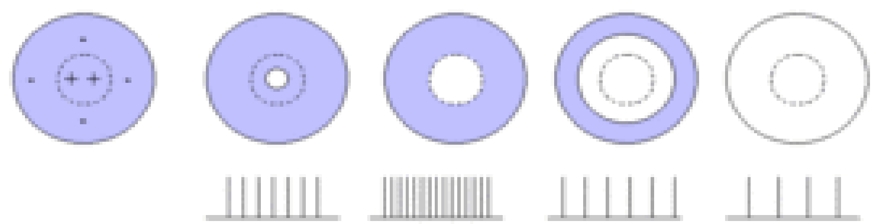
RFs are circular and exist in two types:

- On-center cells: respond excitatorily to light at the center
- Off-center cells: respond inhibitorily to light at the center

The surround has always the opposite characteristics



Receptive field



[Images: http://www.webexhibits.org/colorart/ganglion.html; https://de.wikipedia.org/wiki/Rezeptives_Feld]

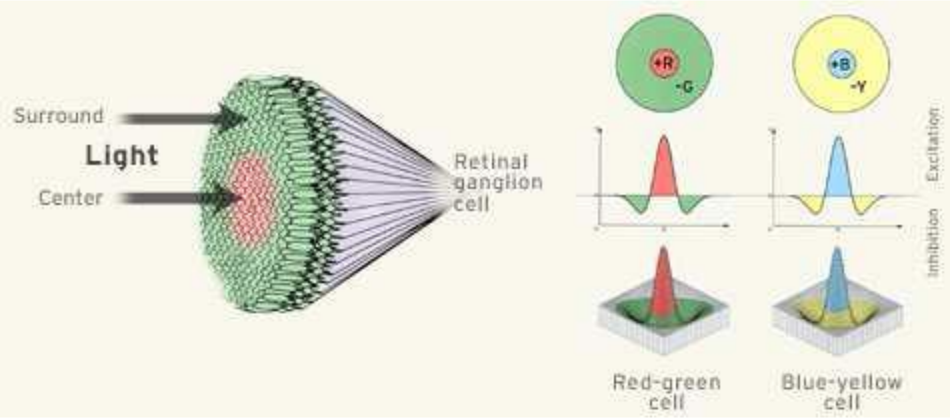
Retina: Ganglion cells

Retinal ganglion are especially important for CV, they can be modelled well by digital filters

RFs are circular and exist in two types:

- On-center cells: respond excitatorily to light at the center
- Off-center cells: respond inhibitorily to light at the center

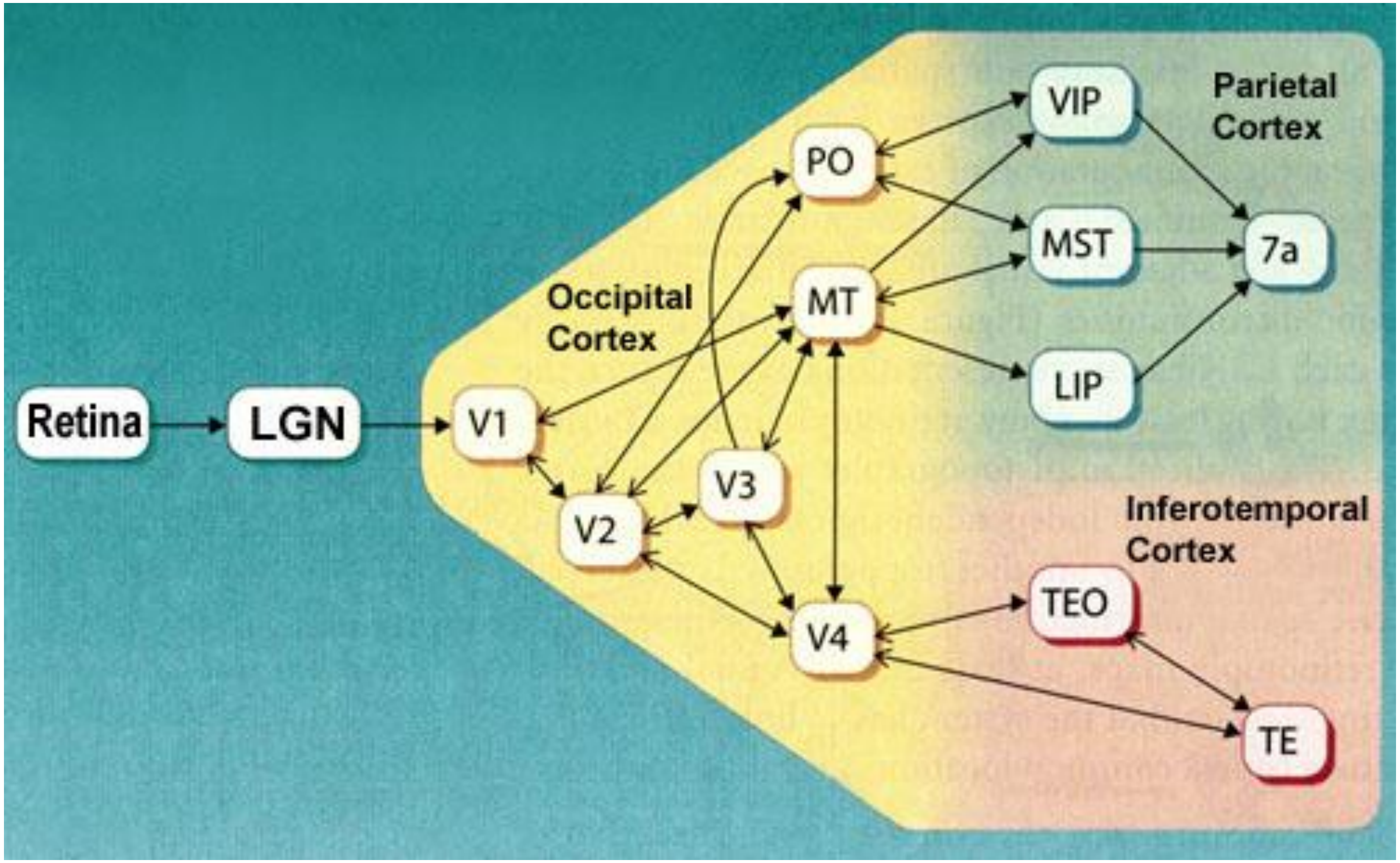
The surround has always the opposite characteristics



Outlook:
Cell responses can be modeled with *digital filters*.
Ganglion cells can be modelled by *Difference of Gaussian filters*

[Images: http://www.webexhibits.org/colorart/ganglion.html; https://de.wikipedia.org/wiki/Rezeptives_Feld]

The Human Visual System



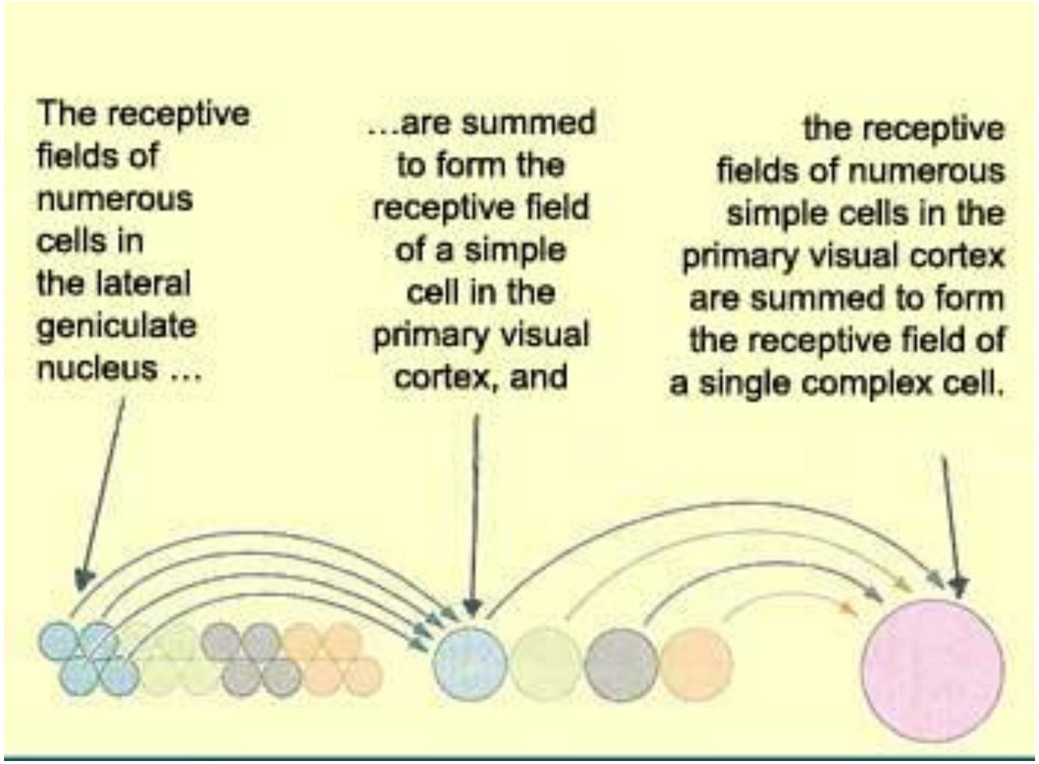
The visual cortex consists of

- **The primary visual cortex (V1)**
- extrastriate visual cortical areas such as **V2**, **V3**, **V4**, and **V5**

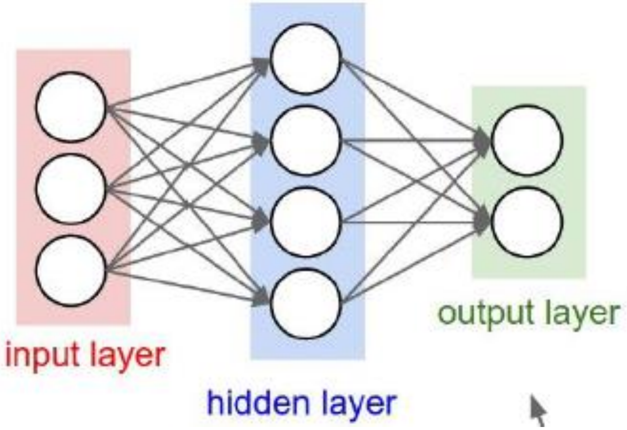
[http://thebrain.mcgill.ca/]

Cells in V1

Hierarchy of cells:



Outlook: this hierarchy is modeled in neural networks (see lecture on classification)

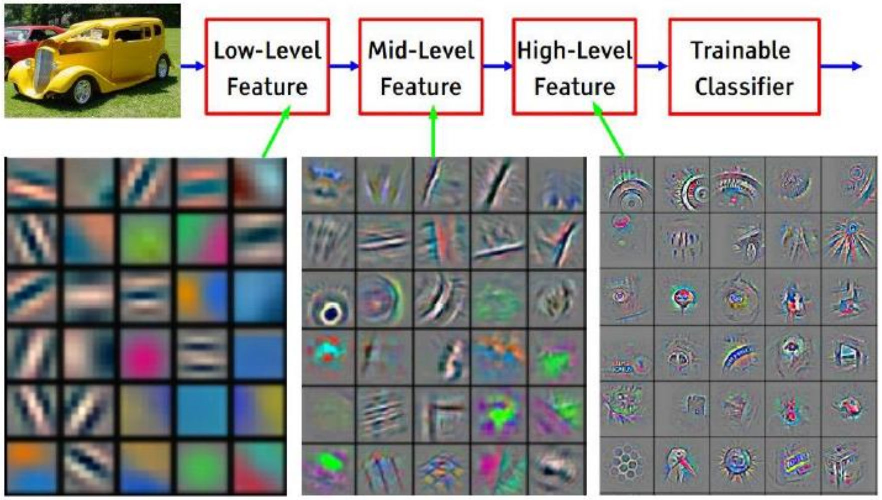


[Image: http://thebrain.mcgill.ca/flash/i/i_02/i_02_cl/i_02_cl_vis/i_02_cl_vis.html#2]

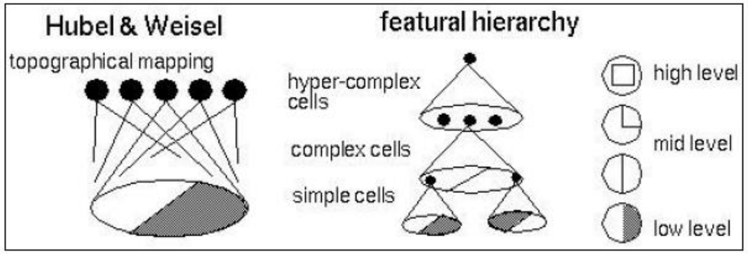
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Feature Hierarchy

If we visualize patterns that neurons in artificial neural networks respond to, we observe similar patterns as the ones that human neurons respond to:

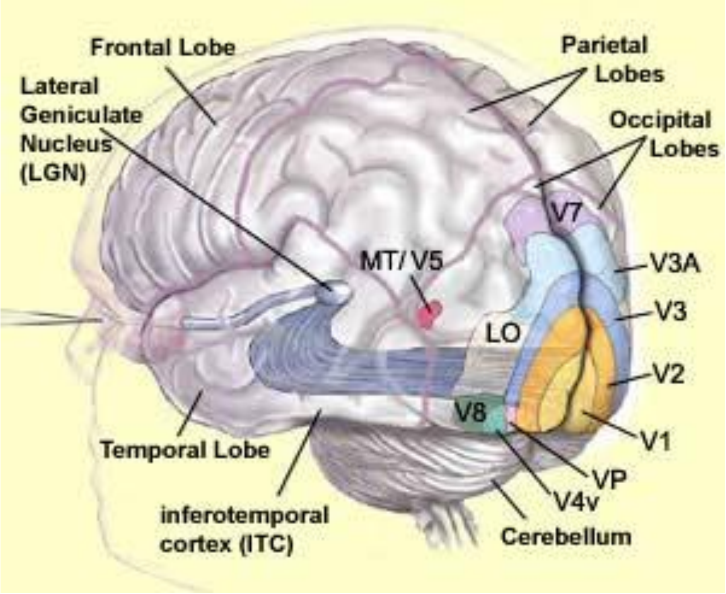


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]



Some important facts

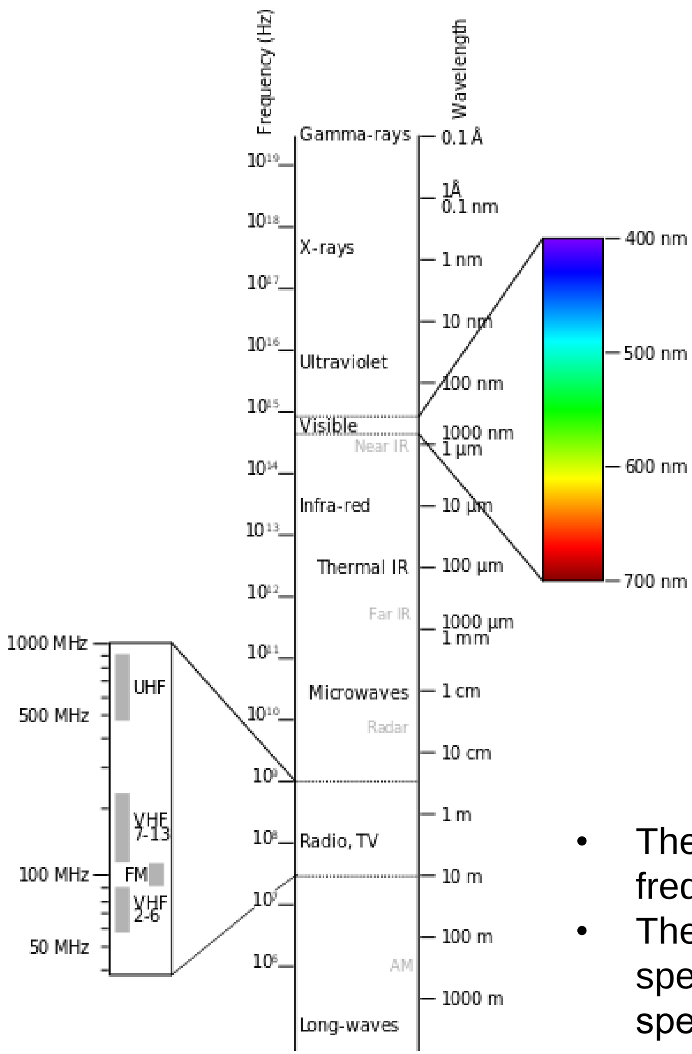
- Up to V1, the path of visual information is mainly serial (eye, LGN, V1)
- The processing after V1 is highly parallel
- Higher brain areas are functionally segregated (motion processing, color processing, face detection, …)
- The further you get away from the eye, the more complex the behavior of single cells (“grandmother neuron”)

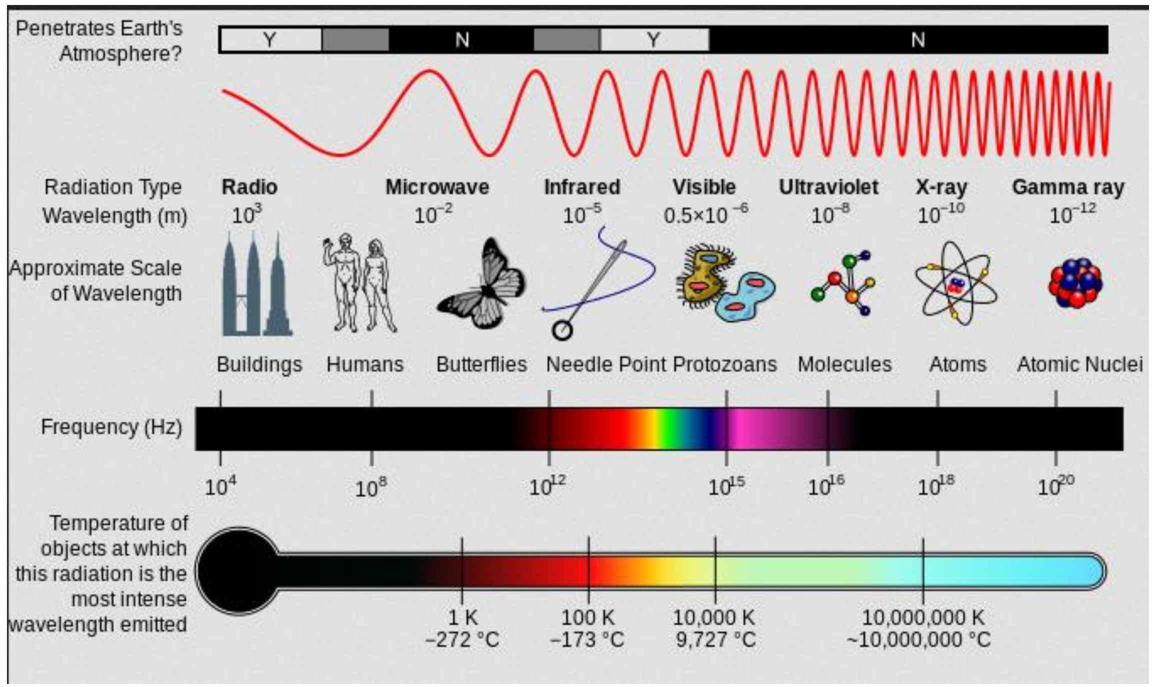


- More details on human visual system and their implications for computer vision in summer term lecture “Computer Vision 2”

[http://thebrain.mcgill.ca/]

The Electromagnetic Spectrum





- The **electromagnetic spectrum** is the collective term for all known frequencies and their wavelengths of the photons.
- The visible spectrum is a very small portion of the electromagnetic spectrum. Images can also come from other parts than the visible spectrum…

[Wikipedia: Electromagnetic spectrum]

The Visible Spectrum

- Most images we are interested in come from the visible spectrum and are captured with ordinary cameras



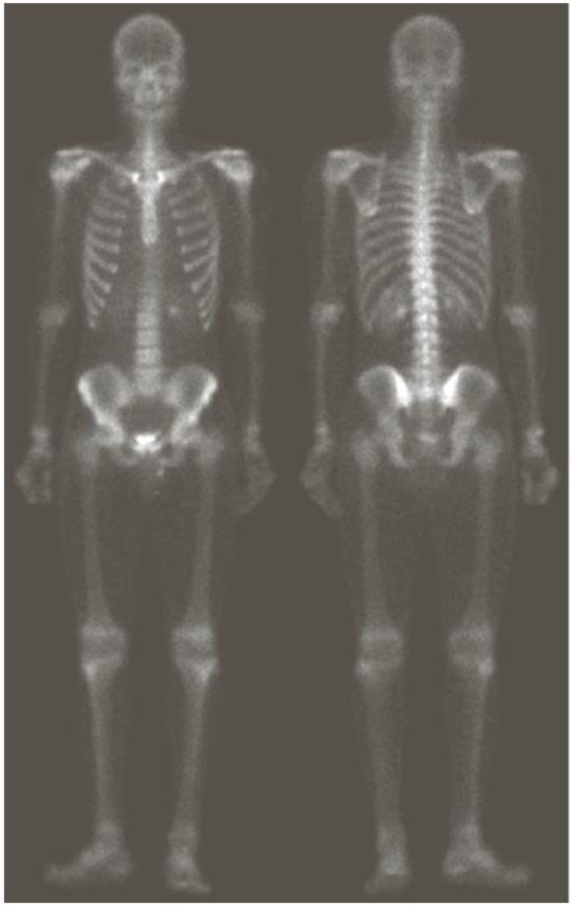


- But there are plenty of other image sources as well…

[Images from ImageNet Dataset]

Image Sources

Bone scan from gamma-ray imaging:



[Gonzalez/Woods]

Image Sources

X-ray imaging:

Chest X-ray

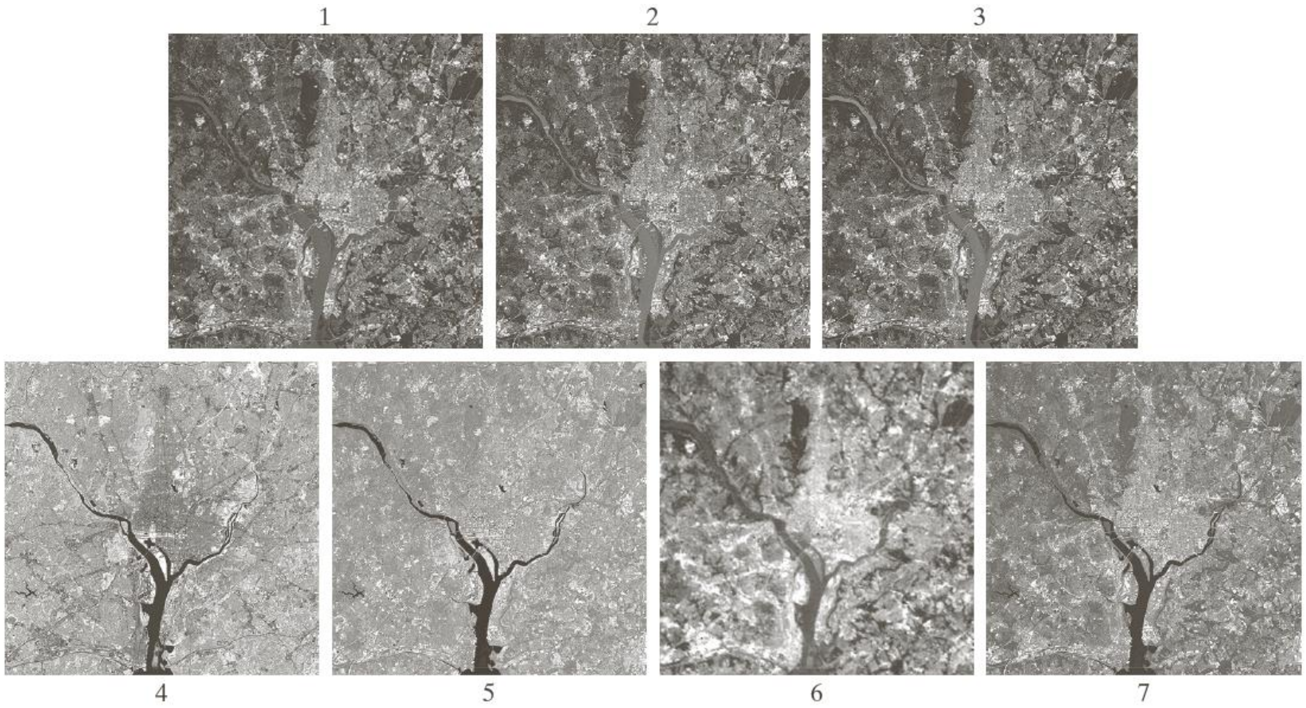
Head CT

[Gonzalez/Woods]

Image Sources

Satellite images (Washington D.C.):

Recorded with different wavelengths from the visible and infrared spectrum

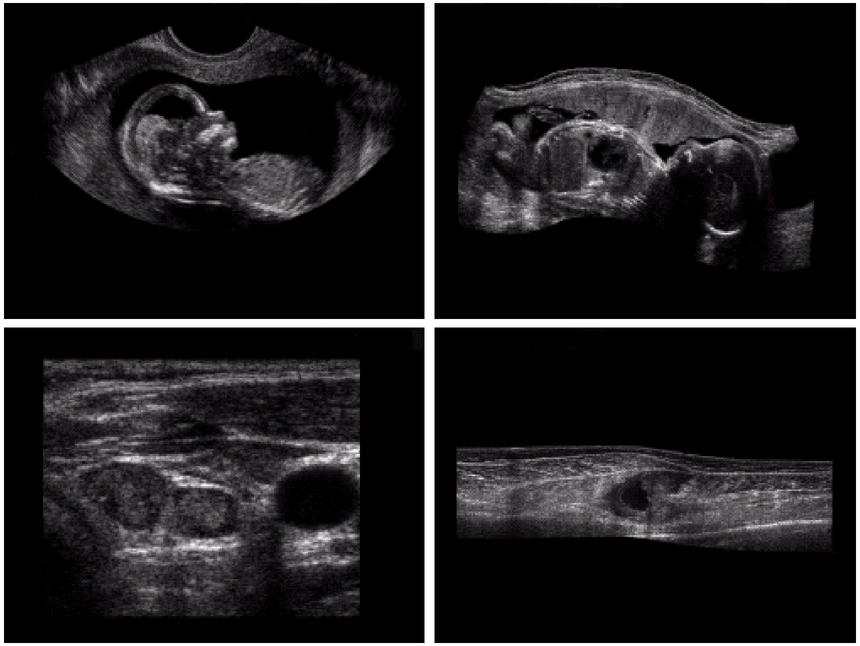


[Gonzalez/Woods]

Image Sources

There are also other imaging modalities that do not come from the electromagnetic spectrum:

Ultrasound images:



a b
c d

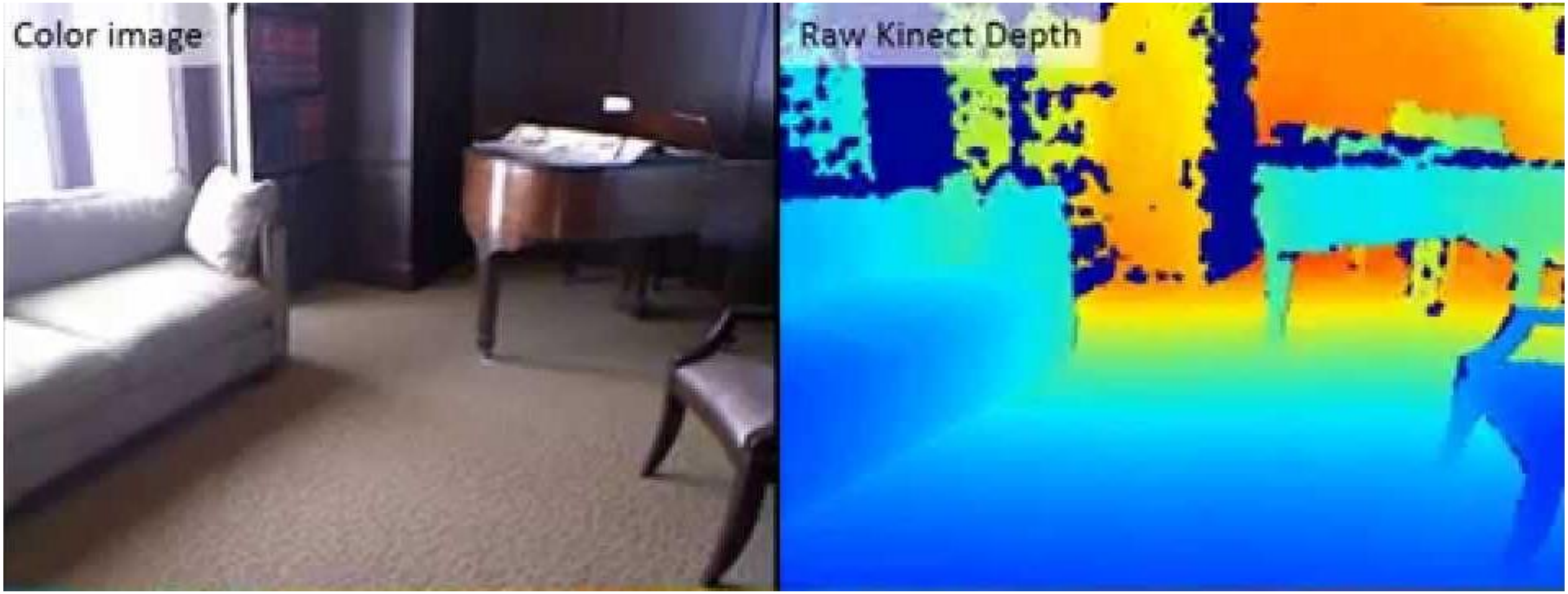
FIGURE 1.20 Examples of ultrasound imaging. (a) Baby. (2) Another view of baby. (c) Thyroids. (d) Muscle layers showing lesion. (Courtesy of Siemens Medical Systems, Inc., Ultrasound Group.)

[Gonzalez/Woods]

Image Sources

Depth images (e.g. from Kinect sensor)

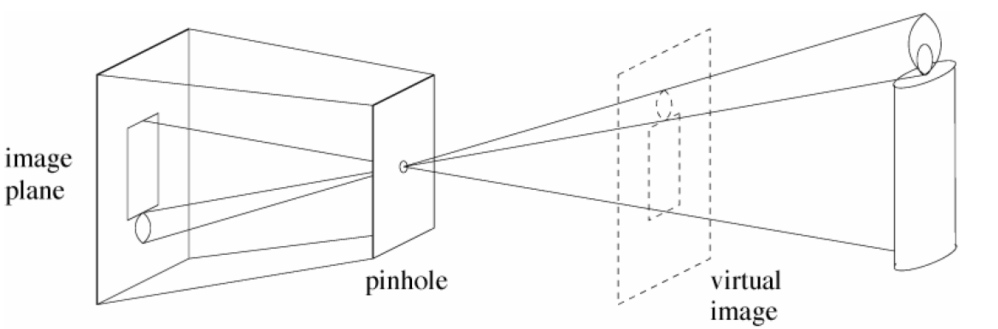
(an infrared laser projector combined with a monochrome CMOS sensor, which captures video data)

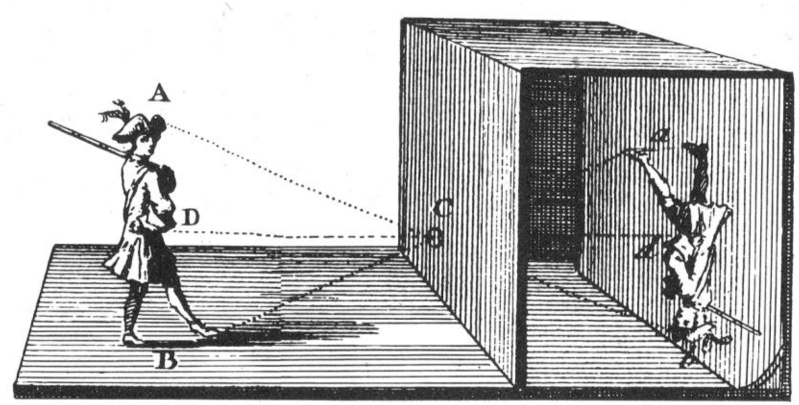


Pinhole Camera

Camera Obscura (Pinhole camera):
A small hole in a box leads to a reversed and inverted (left-right and upside-down) image in the box

First occurances probably already 30,000 BCE
as indicated by cave paintings





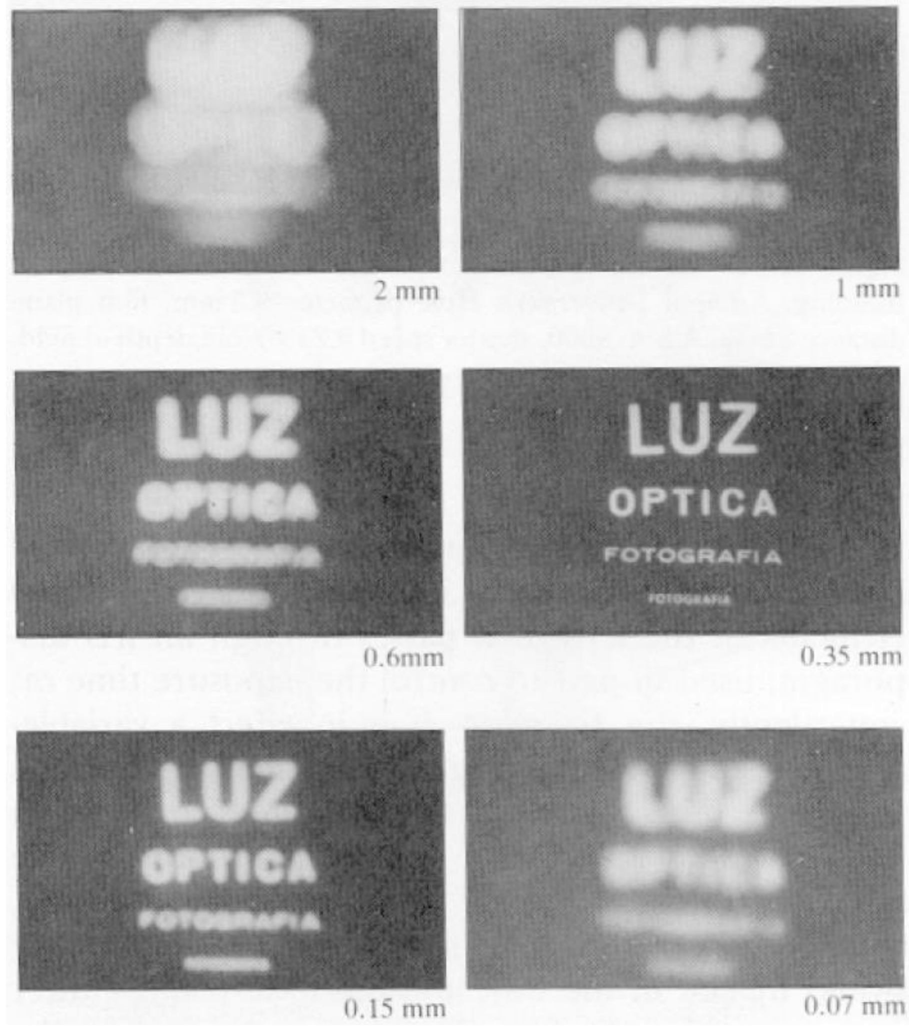
[Image: Wikipedia]

Pinhole Size / Aperture

- Pinhole too big – many directions are averaged, blurring the image

- Pinhole too small – diffraction effects blur the image

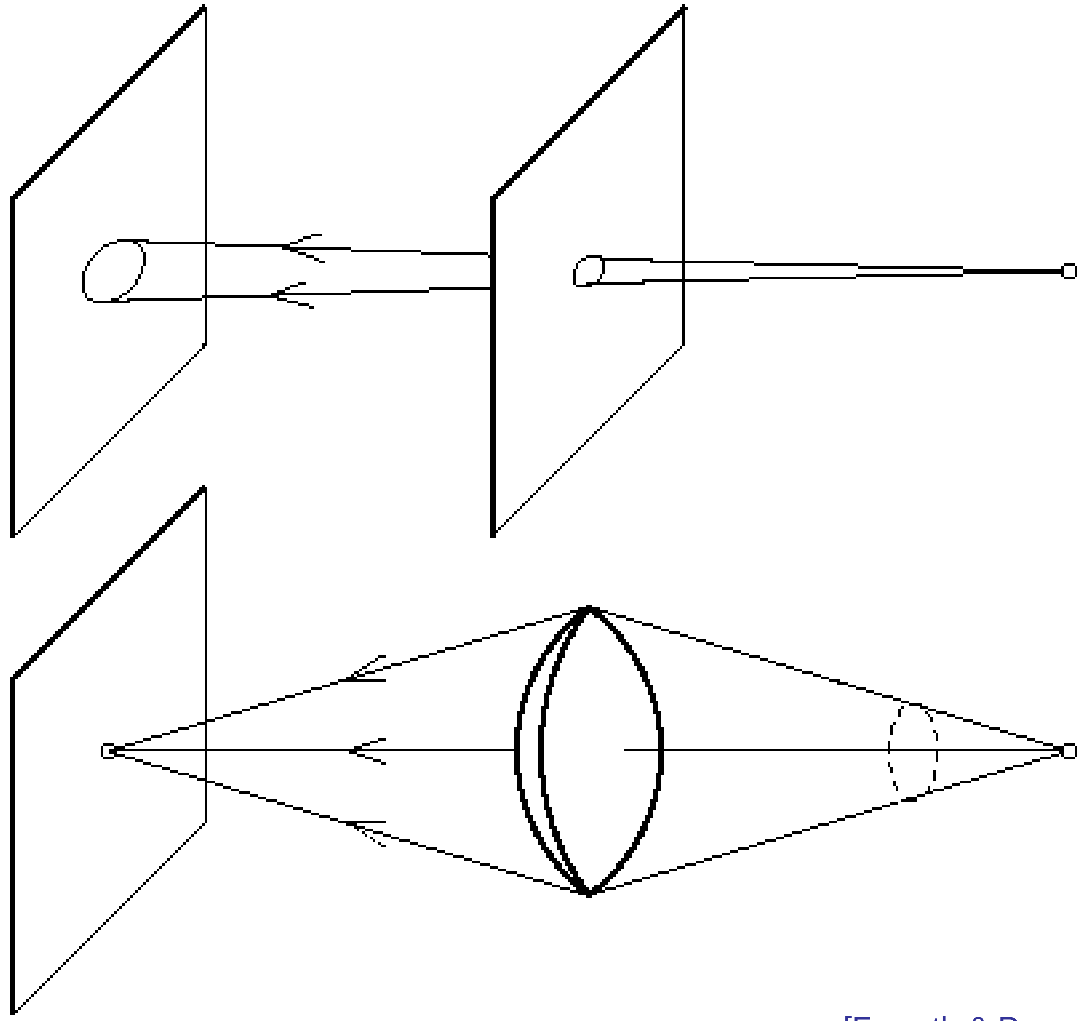
Generally, pinhole cameras are *dark*, because a very small set of rays from a particular point hits the screen.



[Forsyth & Ponce]

The Reason for Lenses

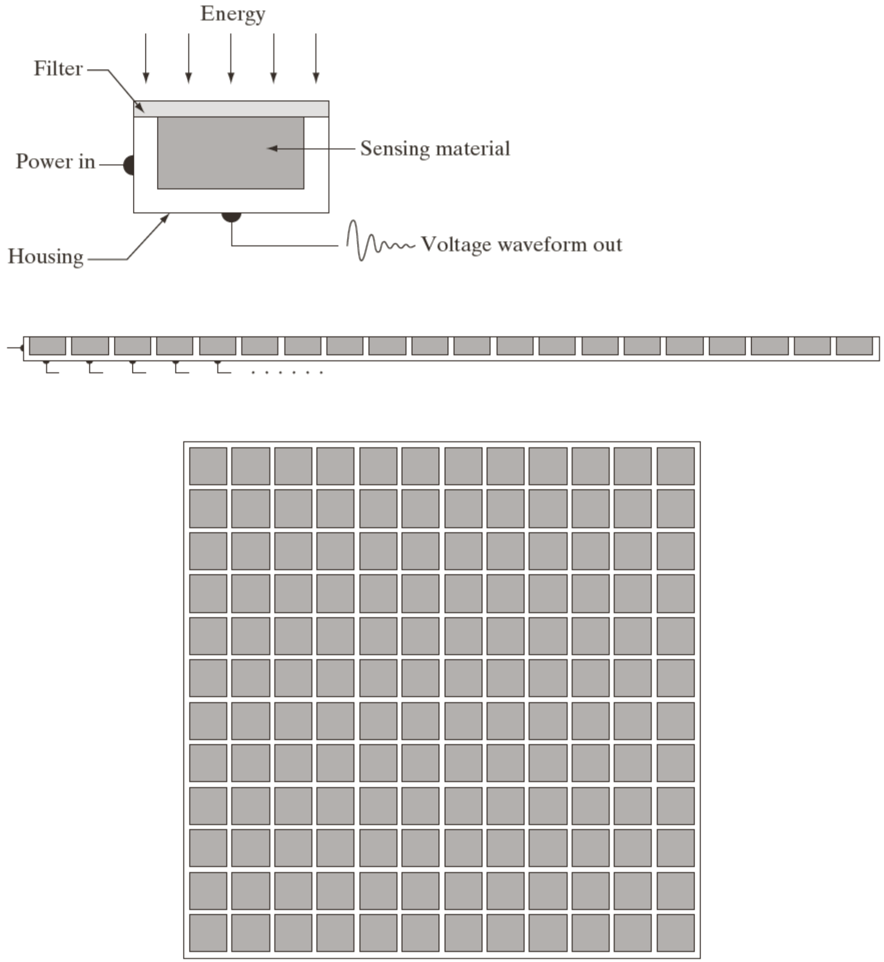
Keep the image in sharp focus while gathering light from a large area



[Forsyth & Ponce]

Imaging Sensors

Three principal sensor arrangements:



Single sensor

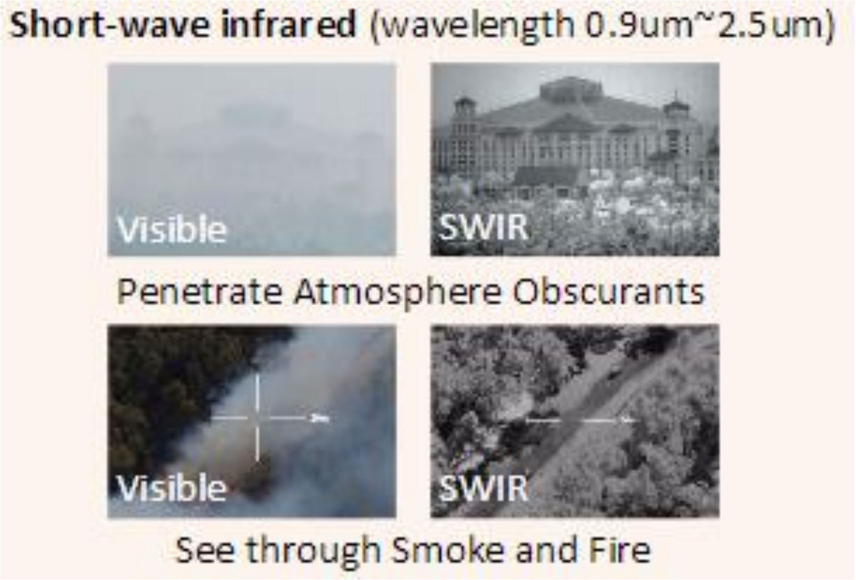
Line sensor

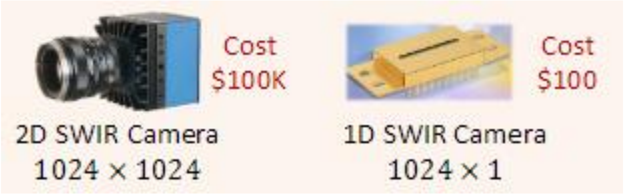
Array sensor

[Gonzalez/Woods 2017]

Imaging Sensors

While array sensors are most common, there are also line sensors, e.g. the SWIR (short-wave infrared) 1D sensor:

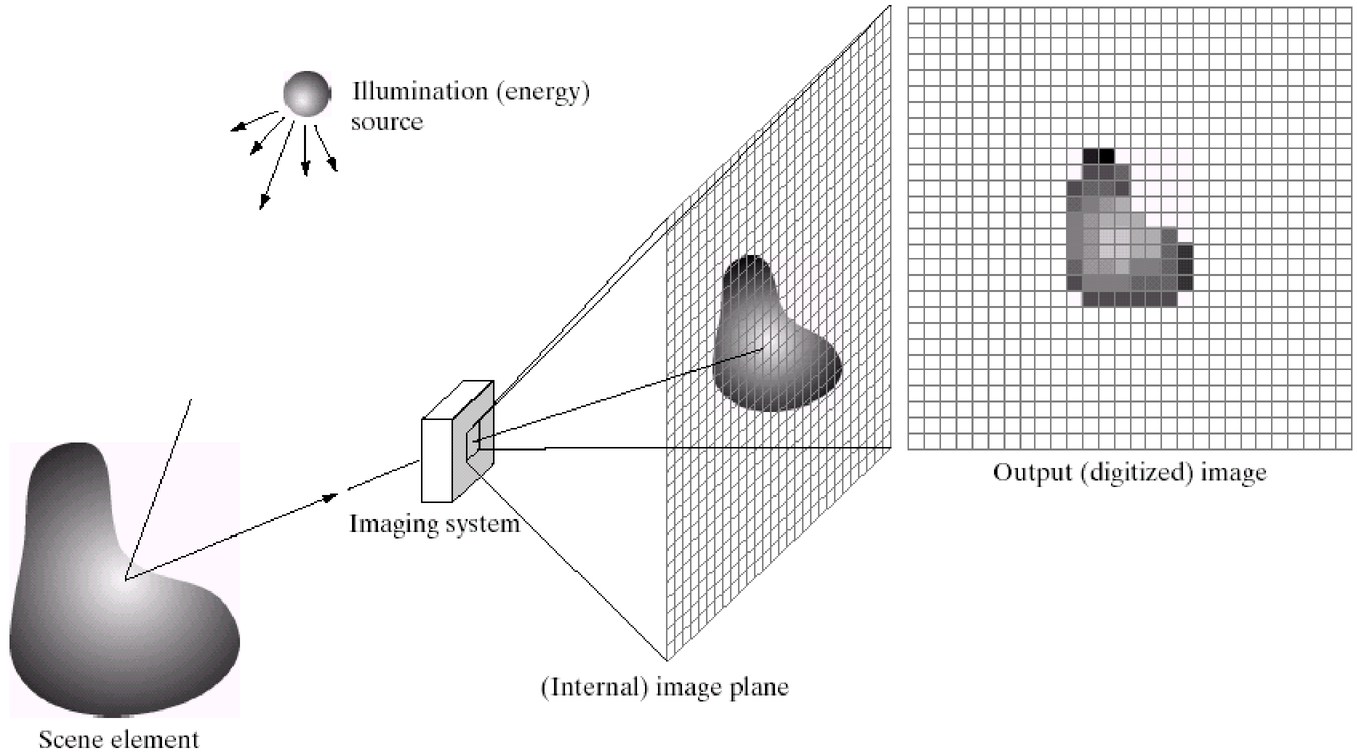




[Wang et al. 2016]

Digital image acquisition

Process to acquire digital images:



[Gonzalez/Woods 2017]

Discretization of Images

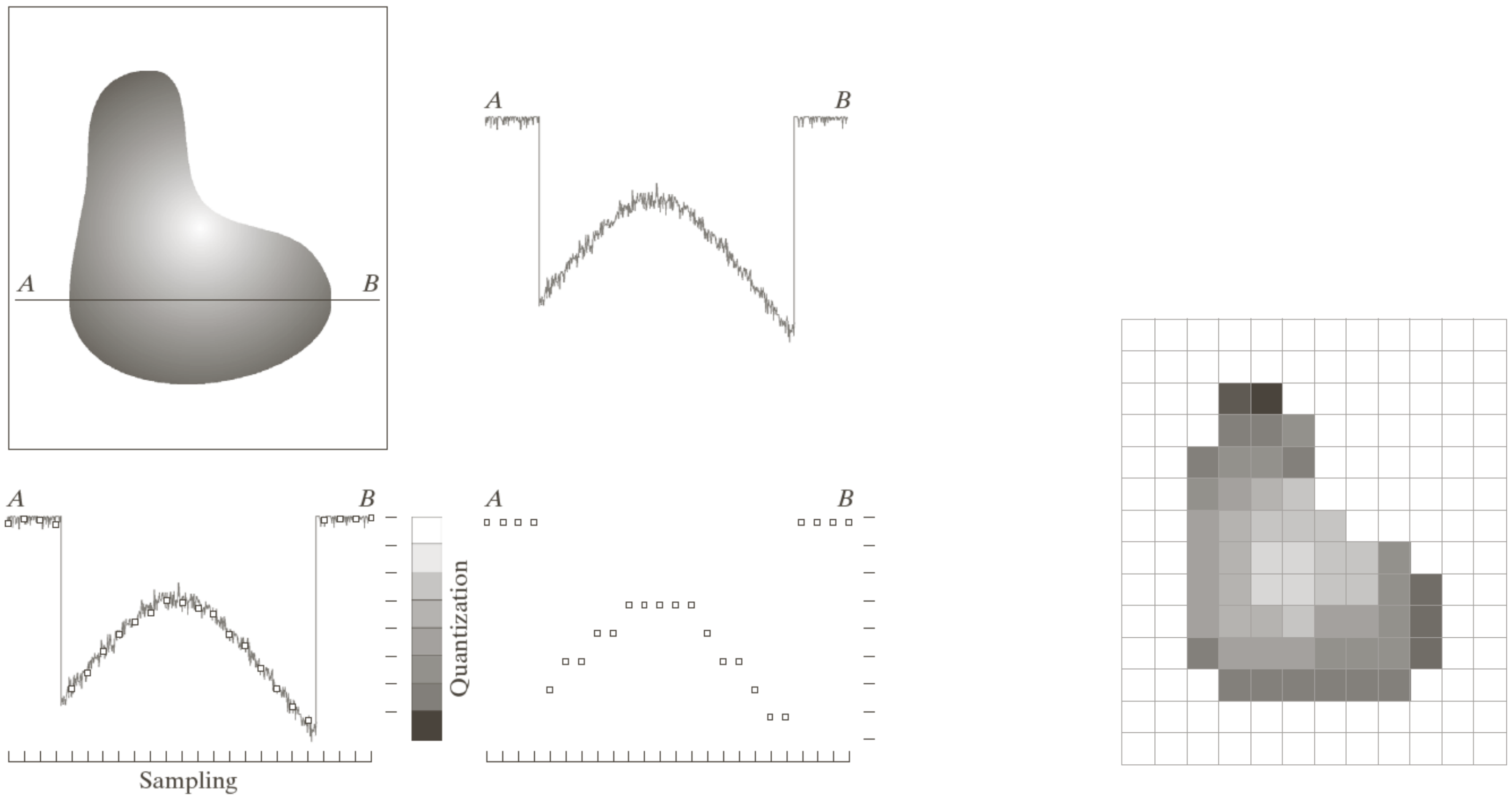
The continuous light signal must be discretized for computer processing:

- spatial quantization (sampling)
  the image plane is represented by a 2D array of picture cells
- grey value quantization
  each grey value is taken from a discrete value range
- temporal quantization
  grey values are taken at discrete time intervals

$$f(x,y,t) \Rightarrow \left\{ \begin{array}{l} f_s(x_1, y_1, t_1),\ f_s(x_2, y_2, t_1), f_s(x_3, y_3, t_1),\ \ldots \\ f_s(x_1, y_1, t_2),\ f_s(x_2, y_2, t_2), f_s(x_3, y_3, t_2),\ \ldots \\ f_s(x_1, y_1, t_3),\ f_s(x_2, y_2, t_3), f_s(x_3, y_3, t_3),\ \ldots \end{array} \right\}$$

A single value of the discretized image function is called a pixel (picture element).

Sampling and quantization

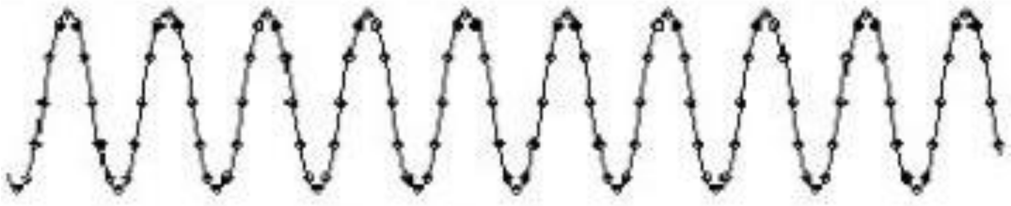


Sampling: Digitizing the coordinate values (how many grid cells)
Quantization: digitizing the amplitude values (how many values per cell)

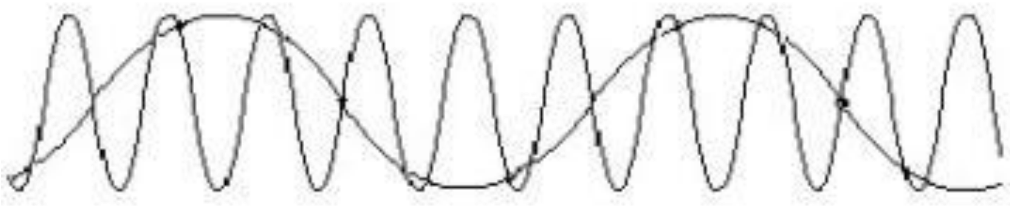
[Gonzalez/Woods 2017]

Sampling

- Reconstruction is only possible, if the “variability” of a function is captured, that means if the sampling density “fits” to the function:
- Adequately sampled signal:



- Undersampling. Effect: “aliasing”



Sampling Theorem

Sampling Theorem (Nyquist/Shannon):

tells us how high the sampling rate must be to capture all information from a signal

> A continuous, band-limited function f can be recovered completely from a set of its samples if the samples are acquired at a rate > 2µ, where µ is the highest frequency in f

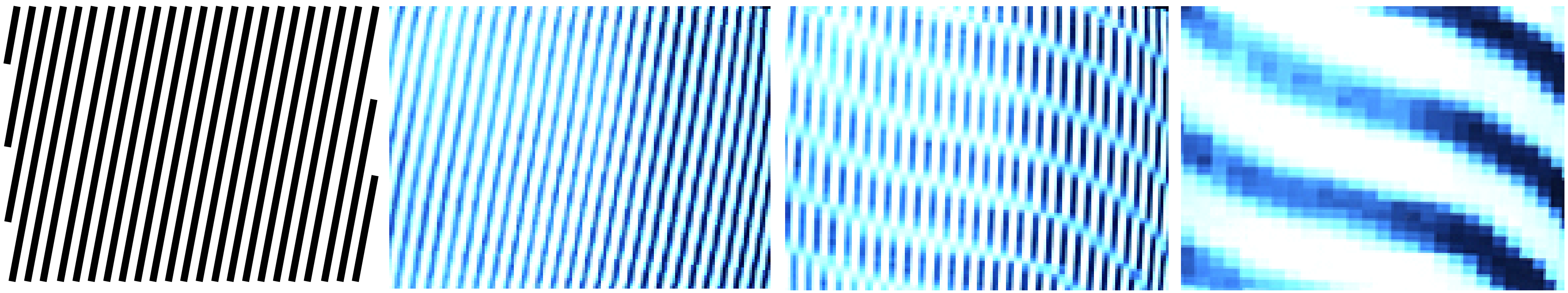
(Bandlimited function: all frequencies are in finite interval (band))

Analogous theorem holds for 2D signals with limited spatial frequencies μ_x and μ_y

Aliasing

Sampling an image with fewer samples than required by the sampling theorem may cause "aliasing" (artificial structures).

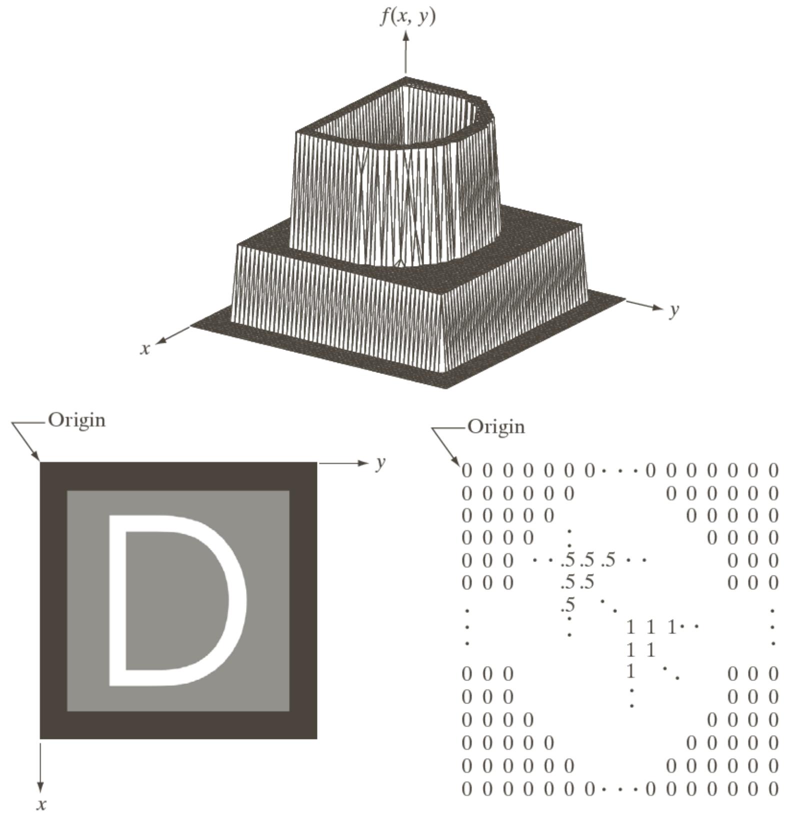
Example:



original | 143 x 128 | 71 x 64 | 35 x 32

To avoid aliasing, bandwidth of image must by reduced prior to sampling (→ low-pass filtering)

Visualizations of Images



[Gonzalez/Woods 2017]

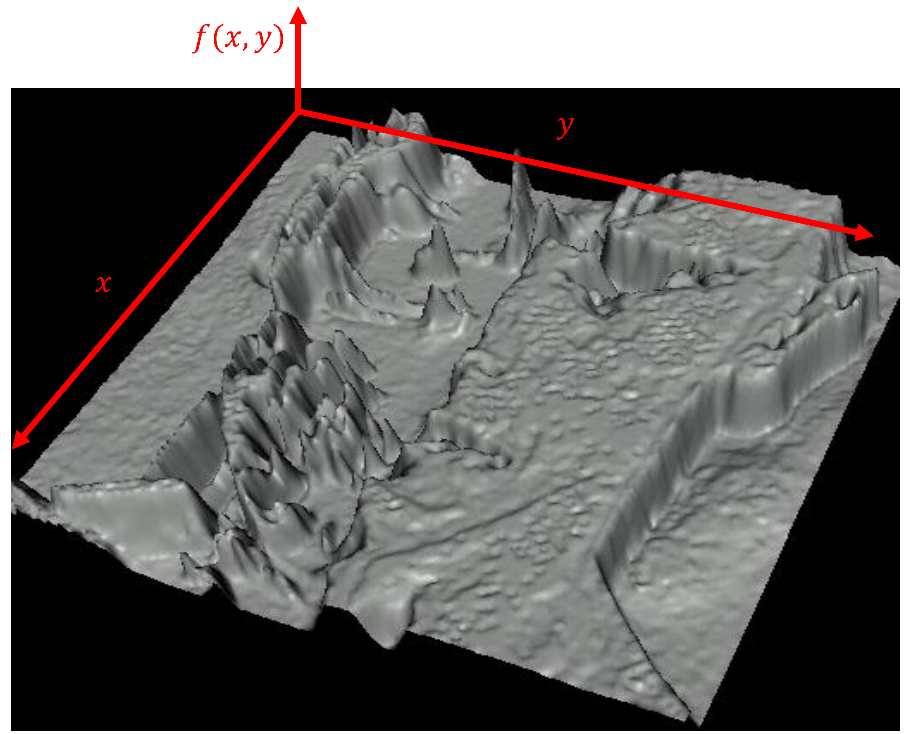
Image as Matrix

- The digital image as a matrix…
- This allows us to apply matrix and array operations

39	24	9	15	19	27	36	37	32	27	26	36	36	35	34	36	41	50	57	50	47
51	40	19	12	16	26	36	38	34	31	31	39	34	33	41	53	60	57	51	61	56
48	48	39	13	17	25	33	35	32	31	34	36	33	34	45	59	64	56	46	64	61
35	45	50	18	20	25	31	31	28	29	33	31	34	39	46	51	52	50	48	59	59
19	36	61	35	17	19	35	37	32	31	33	25	41	50	46	46	52	51	43	61	53
24	28	38	79	51	23	26	49	59	42	21	41	53	58	48	40	46	57	64	68	75
42	37	34	71	60	32	27	60	77	56	33	43	52	62	68	70	77	90	101	107	114
34	29	20	22	45	42	35	52	54	42	45	86	80	81	93	103	106	107	110	99	117
52	52	46	39	65	73	76	88	84	86	112	99	87	81	91	103	111	120	130	163	162
86	85	80	88	81	68	69	76	67	65	84	96	100	112	128	140	151	169	185	212	210
88	82	77	81	70	71	91	110	117	126	139	163	175	190	200	203	205	210	216	232	224
106	109	110	132	137	159	179	182	183	192	200	214	215	215	214	216	218	217	214	222	223
169	190	190	201	203	210	215	214	209	207	209	210	217	215	213	215	213	211	218	225	225
210	217	209	208	210	215	219	216	210	207	209	210	217	215	213	216	213	212	218	224	225
208	214	208	208	210	214	218	215	208	206	208	211	217	216	213	216	214	212	219	224	225
198	207	204	206	208	213	216	214	209	208	210	211	218	217	214	217	214	213	219	223	225
201	207	203	210	211	216	219	217	212	211	214	212	219	217	215	217	215	213	220	222	224
197	207	207	212	212	216	218	215	209	208	211	213	219	218	215	218	216	214	221	222	224
200	212	215	208	209	212	215	212	207	206	210	213	220	218	216	219	216	215	221	222	224
204	208	205	207	208	212	216	215	211	212	216	213	220	219	216	219	216	215	221	221	224
212	210	208	205	210	214	215	212	210	211	213	220	220	220	218	215	216	220	224	225	227

The Image as Signal

But you can see an image also as a (continuous) signal or function $f(x, y)$:



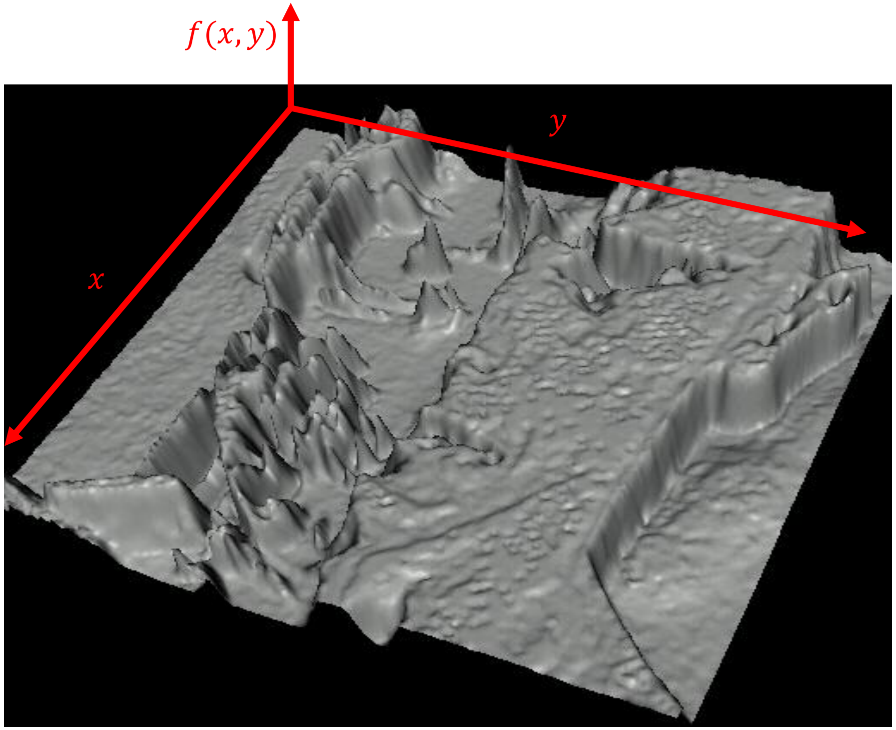
[Images: Steve Seitz]

The Image as Signal

But you can see an image also as a (continuous) signal or function $f(x, y)$:

This allows us to use terms and concepts from signal processing such as

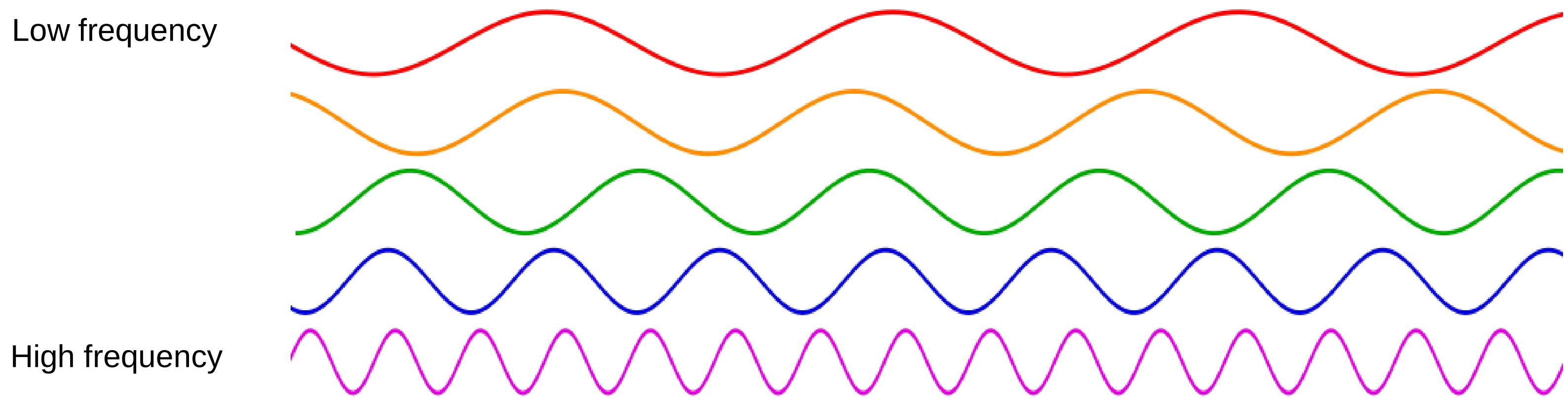
- frequencies,
- filters (low-/high-/band-pass),
- (partial) derivatives,
- noise, etc.



[Image: Steve Seitz]

Frequencies

- What are frequencies?
- Frequency: The number of occurrences of a repeating event per unit time (temporal frequency)

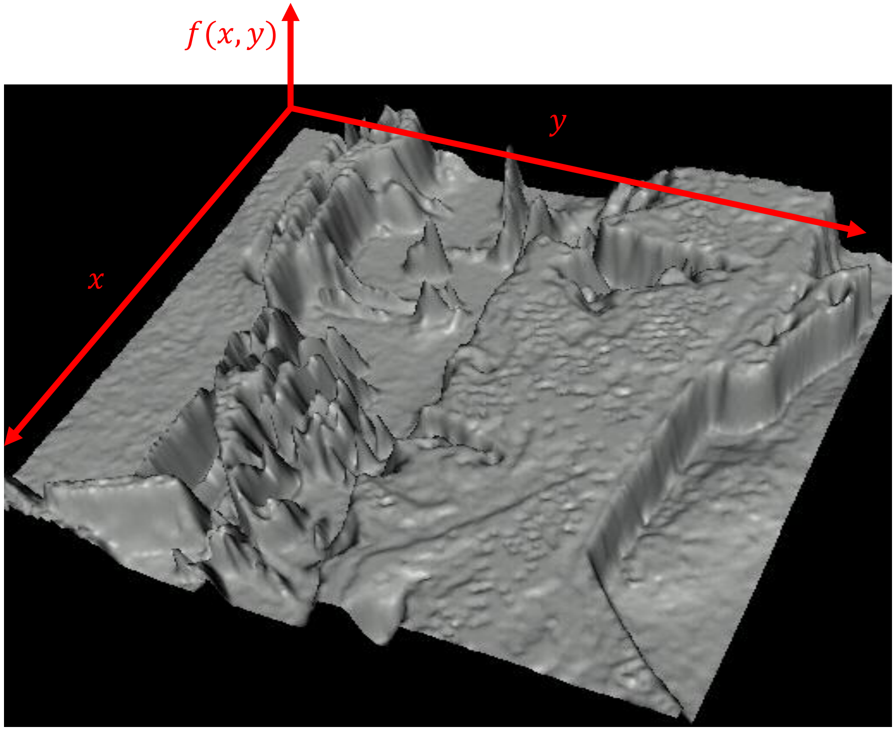


But: there is no time in images! ???

[Wikipedia: Frequency]

Frequencies

- **Spatial frequencies**: replace time axis by spatial displacement axes
- In images: the amount of (intensity) change per unit



[Image: Steve Seitz]

Noise

- What is “noise” in an image?
- “Disturbances” or pixels that vary
- Noise corresponds usually to high frequencies

Original

Salt and pepper noise

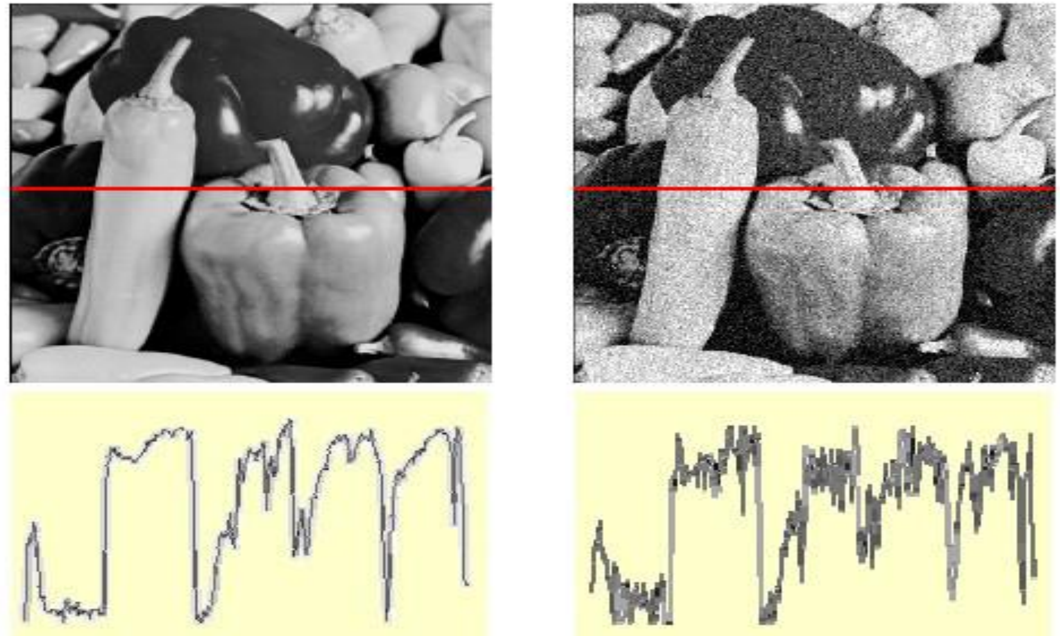
Gaussian noise

- Deviations from an ideal image can often be modelled as additive noise:

[Images: Steve Seitz]

Gaussian Noise

Image Noise



Intensity values

$$f(x,y) = \overbrace{\bar{f}(x,y)}^{\text{Ideal Image}} + \overbrace{\eta(x,y)}^{\text{Noise process}}$$

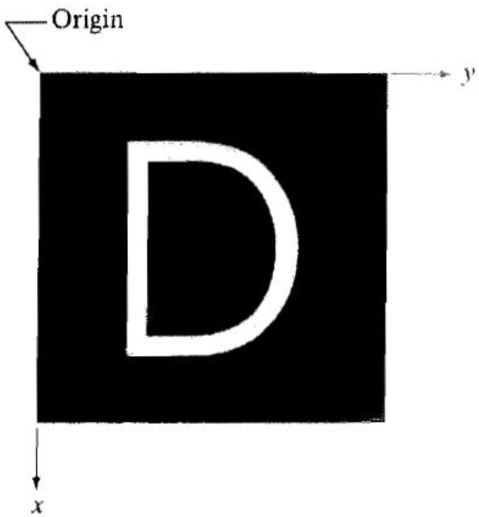
Gaussian i.i.d. ("white") noise:
$\eta(x,y) \sim \mathcal{N}(\mu,\sigma)$

[Image: Martial Hebert]

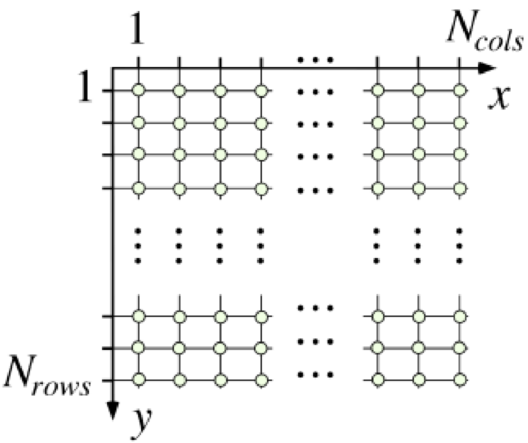
Coordinate Systems

- … a source of confusion…

Gonzalez/Woods:



Klette:

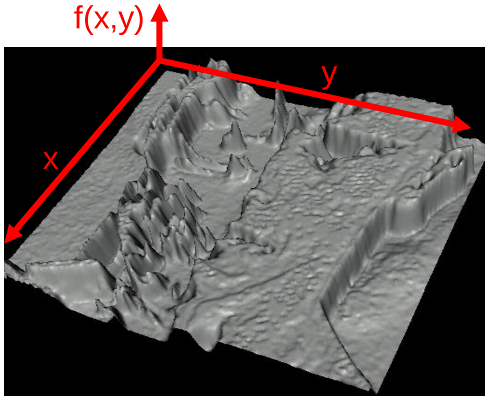


- and others in other sources…
- Solution: always look at the labels of the axes (and label the axes yourself if you draw them)

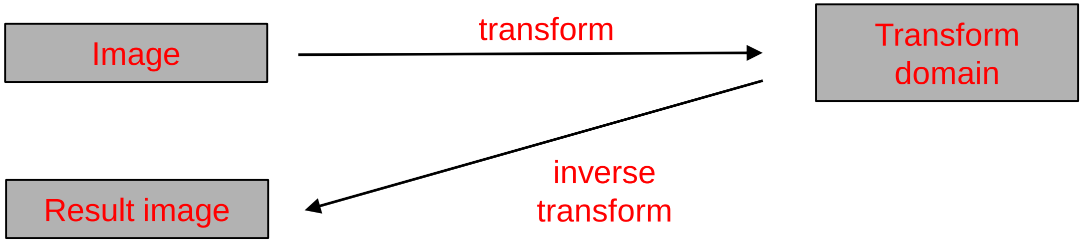
Image Processing Domains

Images can be processed in the

- Spatial domain: the image plane itself



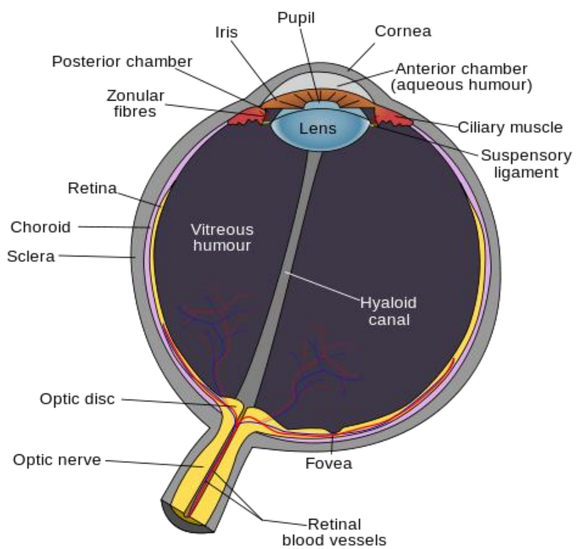
- Transform domain: e.g. the frequency domain

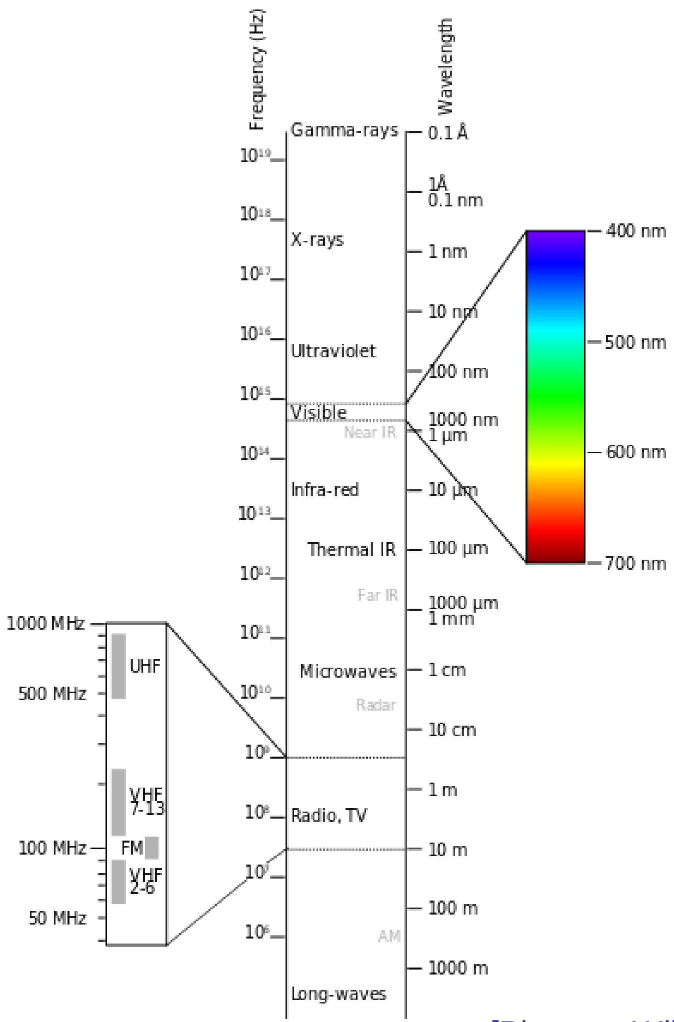


Most of computer vision takes place in spatial domain

Color

- Humans can discern thousands of color shades (but only two dozen shades of gray)





[Pictures: Wikipedia]

Color

Color often simplifies object identification

Where are the apples?

[Picture: http://steinigergarten.blogspot.de/2014/10/apfelsortenbestimmung-und-apfelernte.html]

Color Spaces

Some computer vision tasks are easier in appropriate color spaces.

Some popular color spaces:

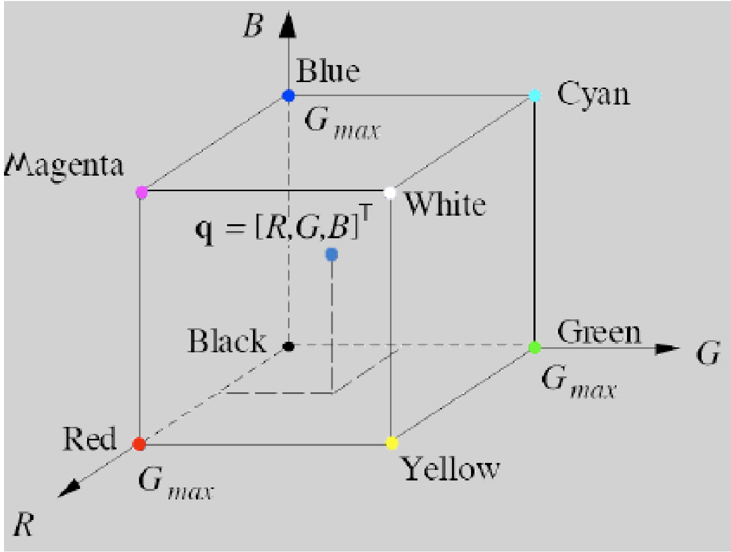
- RGB
- HSV/HSI/HSL
- Color-opponent spaces, e.g. LAB
- there are many more…

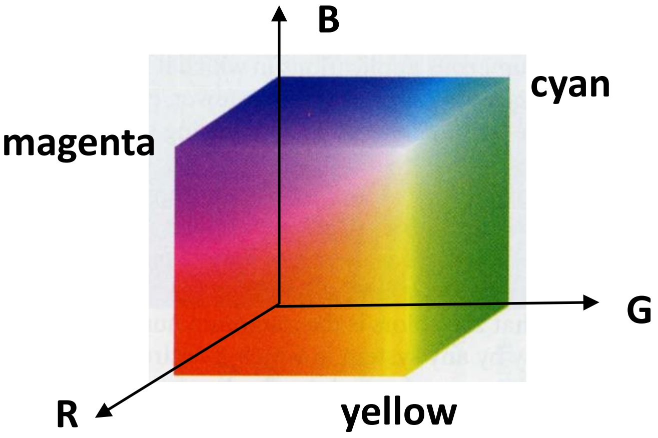
[Wikipedia]

RGB

RGB is the most common color space in computer vision.

Different colors are generated by adding different portions of red (R), green (G), and blue (B).





[Klette]

RGB Images of a Natural Scene

The RGB channels of a color image are rendered as grey value intensity images:

R+G+B

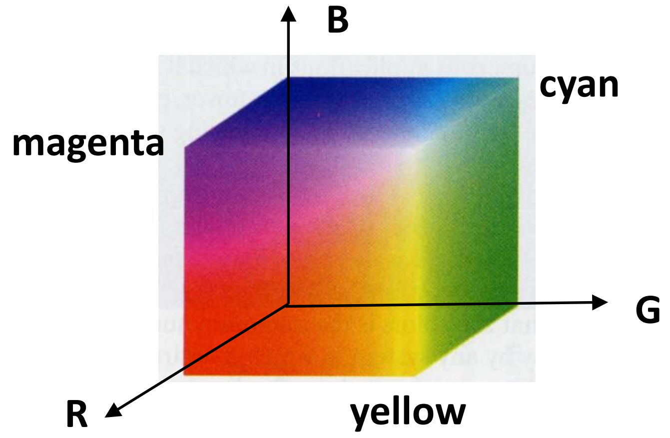
R

G

B

Color spaces

If each channel is encoded by 8 bits, we can represent $256^3 = 16{,}777{,}216$ colors

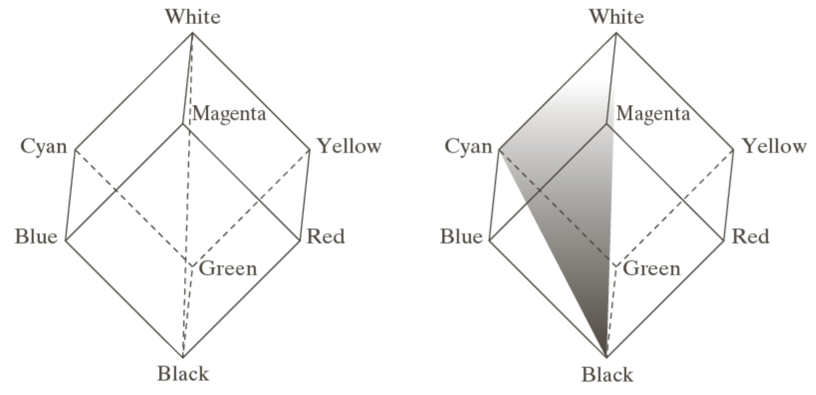


HSL/HSV/HSI

Hue-Saturation color spaces:

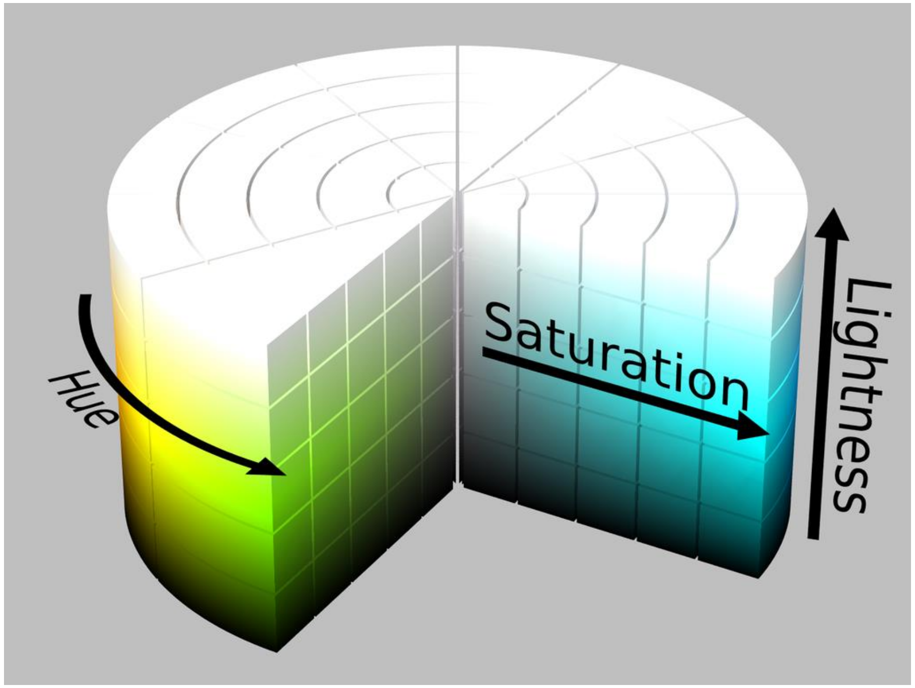
- HSL: Hue-Saturation-Lightness/Luminosity
- HSV: Hue-Saturation-Value
- HSI: Hue-Saturation-Intensity

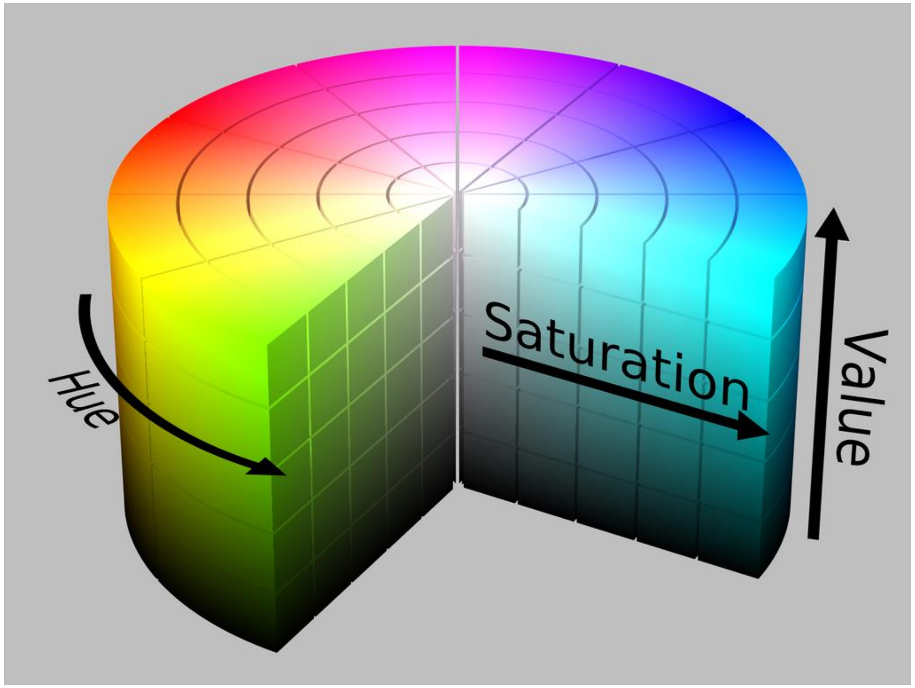
- The most common cylindrical-coordinate representations
- Re-arrangement of RGB geometry



[Image: Gonzalez/Woods 2017]

HSL/HSV





More intuitive than RGB

Better choice for selecting colors!

[Wikipedia]

HSI Color Space

HSI can be derived from the RGB model:

$$I = \frac{R+G+B}{3}$$

$$H = \begin{cases} \theta & \text{if} \quad B \le G \\ 360 - \theta & \text{if} \quad B > G \end{cases}$$

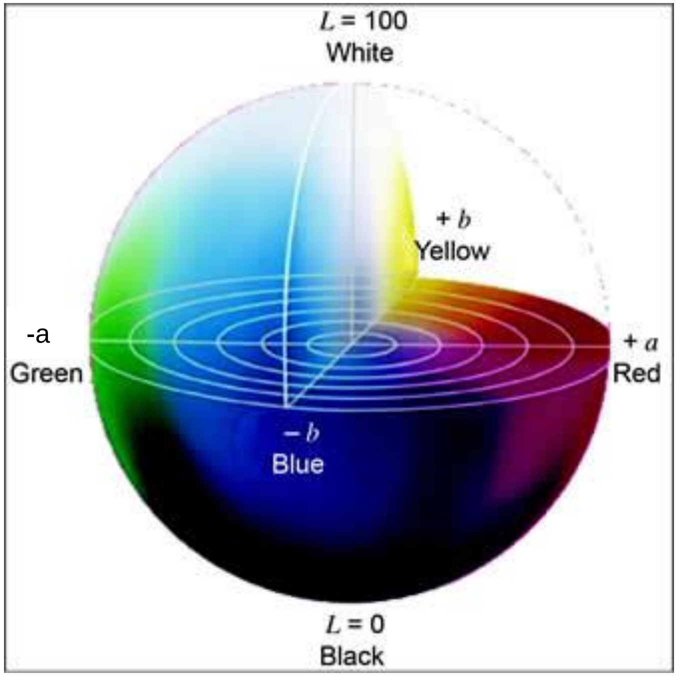
$$S = 1 - \frac{3}{R+G+B} \min(R,G,B)$$

$$\theta = \arccos\left(\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right)$$

HSV/HSL: “Value” and “lightness” are computed differently, e.g. value $V = \max(R, G, B)$. See [Wikipedia “HSL and HSV”]

Opponent Color Spaces

- Opponent Color Spaces have three axes with opponent colors (colors perceived by humans as opponents):
  - Red-green
  - Blue-yellow
  - Black-white



Example:

- The Lab color space (also CIELAB, or L*a*b* color space) is a color-opponent space with dimensions
  - L for lightness
  - a for the red-green dimension
  - b for the blue-yellow dimension
- Close to human perception (“perceptually uniform”)
- More in Lecture “Computer Vision 2”

[Image: http://www.rehab.research.va.gov/jour/05/42/5/bicchierini.html]

Summary

- The human visual system
- The electromagnetic spectrum as source of images
- Image acquisition and digitization
- Image representations
- Frequencies in images
- Noise in images
- Image processing domains (spatial vs transform domain)
- Color Models

Primary Literature

- Rafael C. Gonzalez and Richard E. Woods: Digital Image Processing, Addison-Wesley Publishing Company, 4th edition: (Standard edition) 2017.
  (parts from chapters 1 and 2)
  Note: there is also a Global Edition which contains mainly the same content but one chapter less and page numbers and figures vary

Secondary Literature

Literature on the human visual system:

- *The brain from top to bottom* – An interactive website about the human brain and behavior: http://thebrain.mcgill.ca
  (The senses → Vision → Level of explanation: advanced)

- Peter K. Kaiser: *The Joy of Visual Perception: A Web Book.* York University, *http://www.yorku.ca/eye/* (Figure)
- Quiroga, R. Quian, et al.: *Invariant visual representation by single neurons in the human brain. Nature* 435.7045 (2005): 1102-1107.
- Palmer: Vision Science: Photons to Phenomenology
- Birbaumer/Schmid: Biologische Psychologie

Other literature:

- D.A. Forsyth, J. Ponce: Computer Vision, A Modern Approach
  (2nd edition), Prentice-Hall 2012
- Wang, Jian, et al. "Dual Structured Light 3D Using a 1D Sensor." *European Conference on Computer Vision*. Springer International Publishing, 2016.