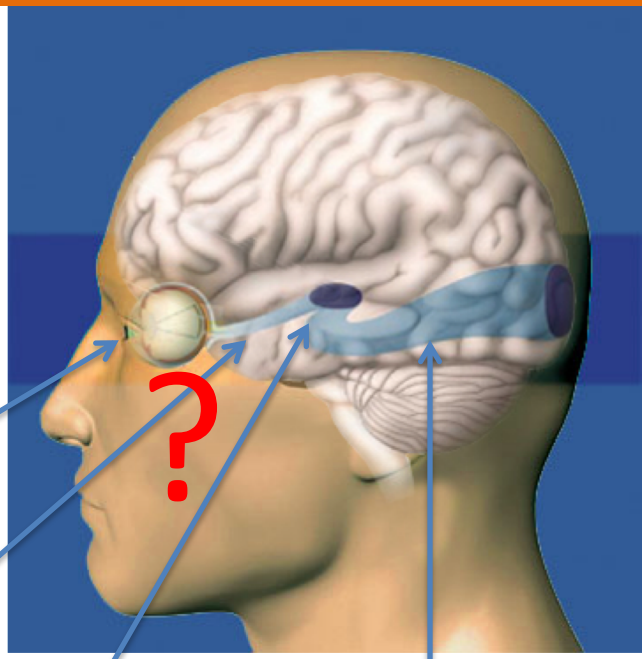
Image Data Compression

Biological visual systems

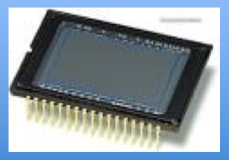
© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Hmm… how is biology related to image compression?

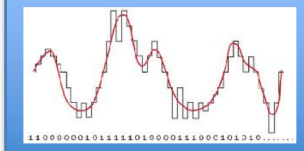
- **Human perception justifies definitions of “irrelevant” data**
- Biological solutions inspire (optimal?) processing techniques
- Human vision is adapted to natural image statistics (?)
- Evolutionary adjustments of different organisms demonstrate adaptations to specific visual environments
- Biological vision: massively parallel image processing
- Maybe gain insight into computer vision problems?
- Neurophysiology in experiencing a huge progress
- A funny topic to study!



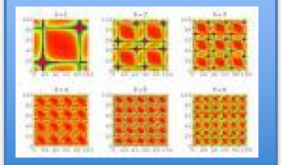
Sensor



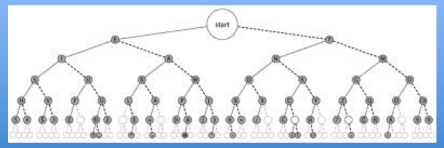
Quantization



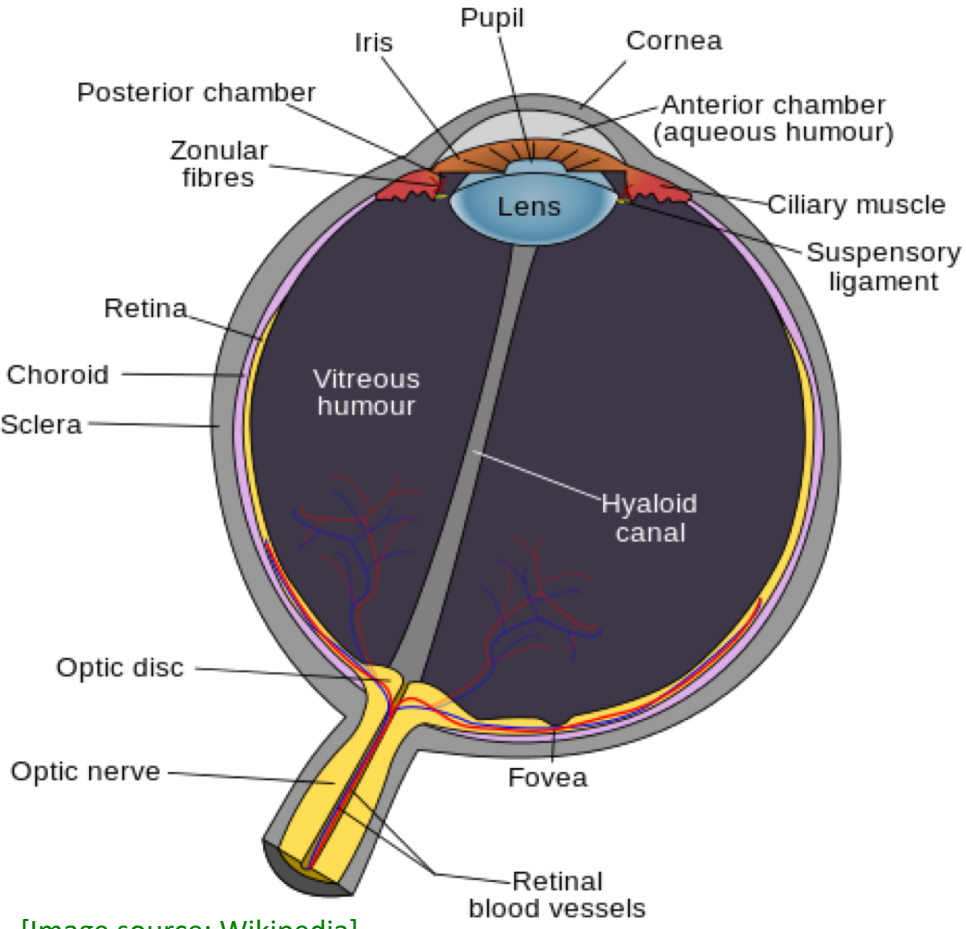
Transform



Coding



Human eye as a camera: technical specifications



[Image source: Wikipedia]

- **Image formation:**
  Adaptive lens, aperture, discretized sensor
- **Focal length:**
  About 22 mm
- **Static contrast ratio:**
  ~ 6.5 f-stops (1 : 100)
- **Adaptation to darkness:**
  ~4 sec (initial), ~ 30 min (complete)
- **Dynamic contrast ratio (with adaptation):**
  ~ 20 f-stops (1 000 000 : 1)
- **Field of view:**
  ~ 180° (h) x 120° (v)
- **Sensitivity of an adapted eye in darkness:**
  ~ ISO 800 (~ 1-2 photons per pixel)
- **Color resolution:**
  Trichromatic sensor, ~ 10 000 000 colors
- **Angular resolution:**
  ~ 6 arc sec (cf: receptive cells are ~30 arc sec in size, pitch of ~30-60 arc sec!)

All specs close to the physical limitations, a similar image processing chain is still challenging to reproduce with the current technology!

If extended to 120°x120° field of view, equivalent to a very decent 600 megapixel camera!

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

(Human) eye retina as a sensor

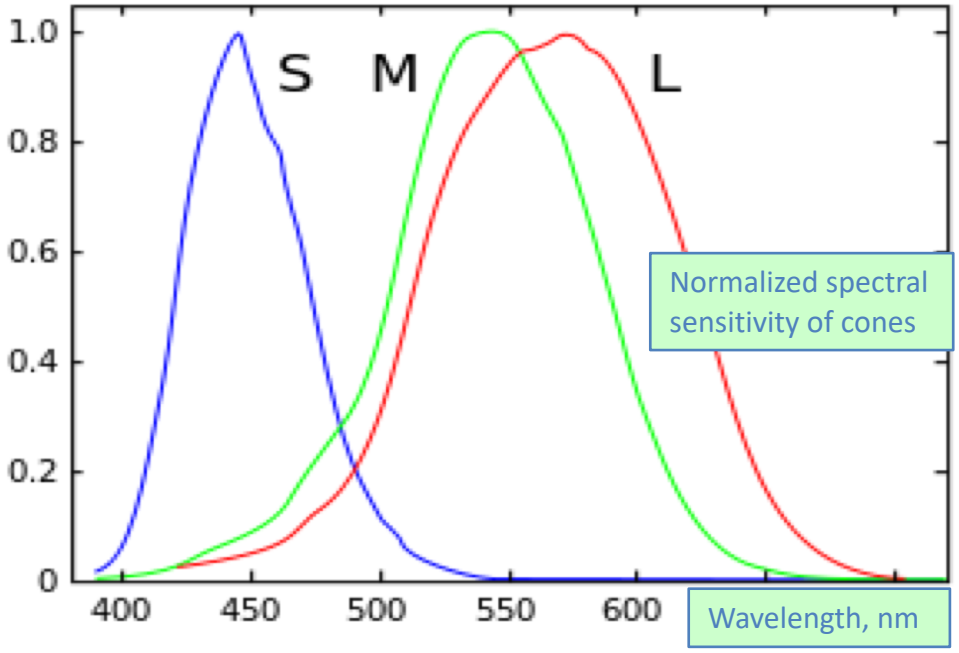
Light receptor cells:

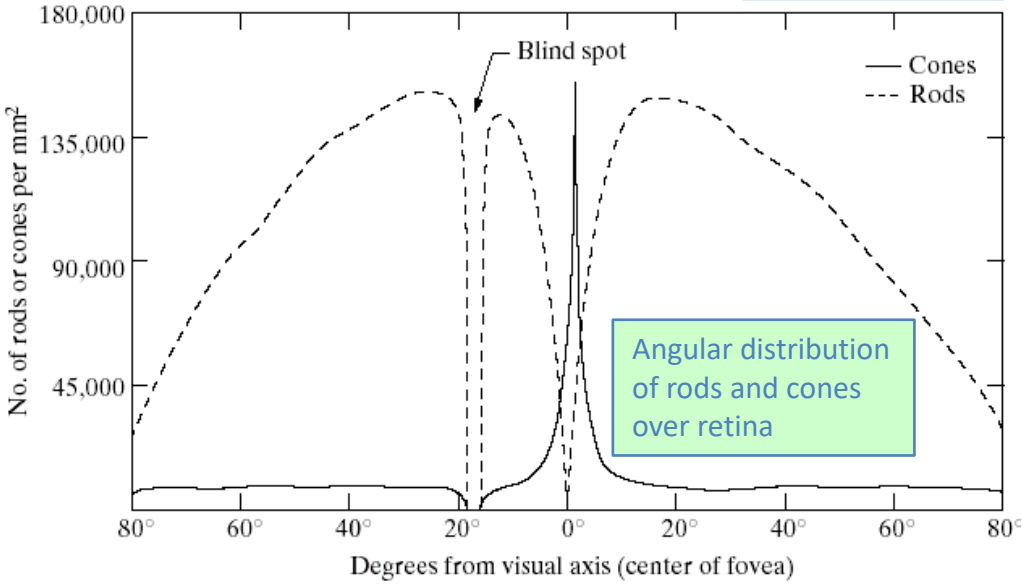
Rods

- Sensitive to brightness only
- Active at low illumination
- High sensitivity
- ~ 120 000 000 in total

Cones

- Type M – max sensitivity at ~535 nm (green)
- Type L – max at ~570 nm (yellow-red)
- Type S – max at ~445 nm (violet)
- ~ 6 000 000 in total (mostly in **fovea**)





These results justify the use of RGB, gray-scale images, and "chroma subsampling"

Chroma sub-sampling (e.g. JPEG, NTSC):

- Convert RGB signal to YCrCb representation
- Encode Y (luminance) w. highest spatial resolution
- Encode Cr, Cb (chroma) channels w. lower resolution

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Foveated imaging



General idea:
Since only a small region near the optical axis is color-sensitive and has the highest resolution, we may encode unimportant image parts with lower color and spatial resolution.



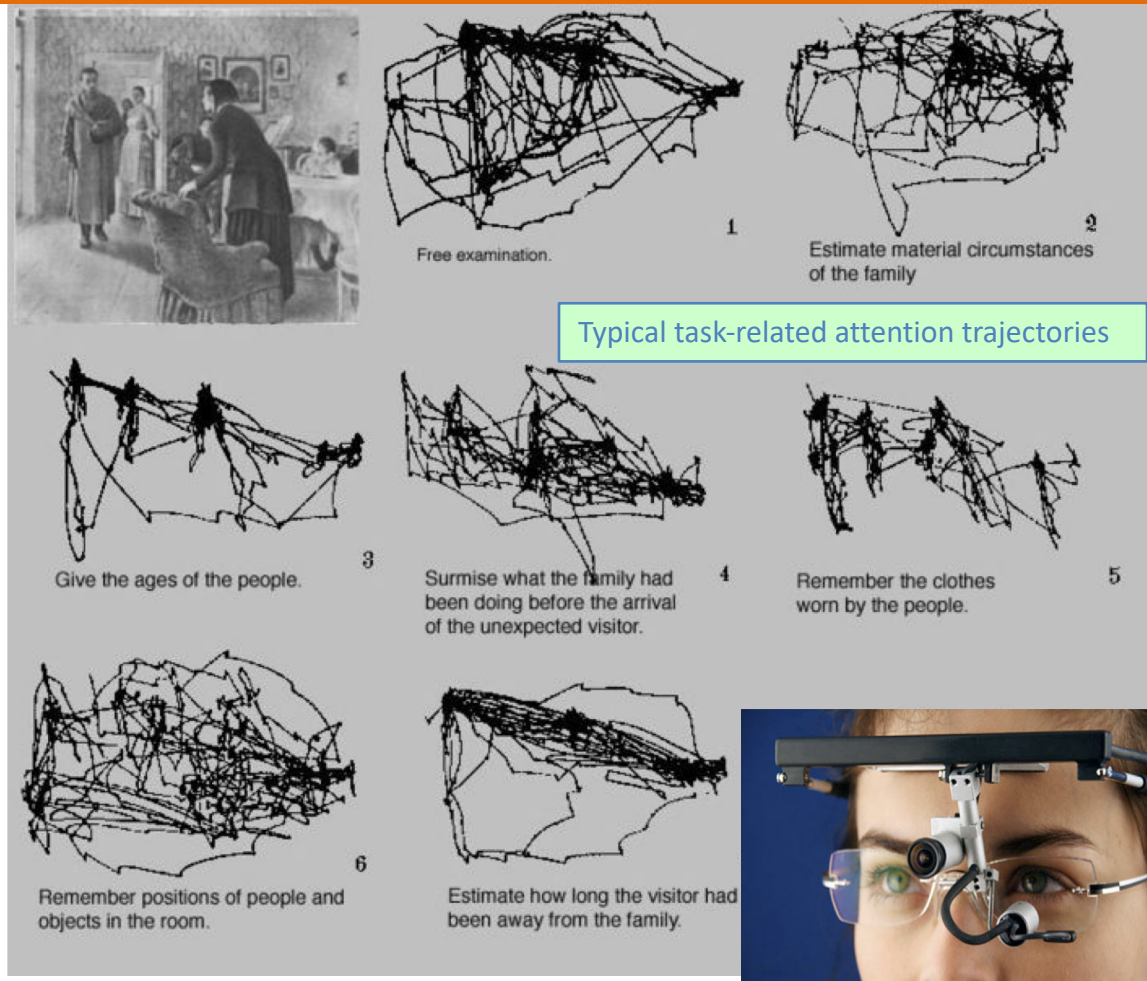
Goals:

- Save bandwidth in transmission
- **Foveated progressive loading:** transmit higher-resolution information in different areas in the order as they come to the focus of attention
- Need to know the **attention focus trajectory**
- The trajectory is person- and task-dependent

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Eye motion and tracking studies

- Motion is known to be essential to image formation
- Has to do with pigment depletion in cells, Stiles-Crawford effect, …
- Perfectly stabilized image (to ~ arc sec) disappears in a few seconds
- Controlled by dedicated high-performance nerve centers
- Gaze analysis, fixation prediction: active field of research in Fraunhofer IOSB in Karlsruhe





Types of eye motion

- Saccades: the fastest bodily motions (up to 900°/s, up to 90°)
- Fixations: periods between saccades, with smaller motions
- Smooth pursuit: specialized “prey hunting” mode
- Further modes: micro-saccades, ocular tremor, ocular drift, …

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Do humans have the best vision? (no)

Humans are trichromats – have three types of cones

- Animals can be tetrachromats, pentachromas, …
- (humans can also be tetrachromats: some women carrying OPN1MW gene on one X chromosome and OPNMW2 on the other)
- Of placentary mammals, only primates are believed to be trichromats
- Color vision evolved to match the environment
- (RGB could help apes find ripe fruit between the green tree leaves)

Most advanced color vision + most complex eyes in animal kingdom: **mantis shrimps (dodecachromat)**







- 16 sorts of photoreceptor pigments (12 for color sensitivity, 4 for color filtering)
- Sensitivity to polarized light
- Each individual eye: trinocular vision (each object is observed by three eye areas), direct depth perception
- Species *Gonodactylus smithii*: 4 linear and 2 circular polarization components (i.e. all Stockes parameters!)

… options limited by the transparency of atmosphere and solar spectrum

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

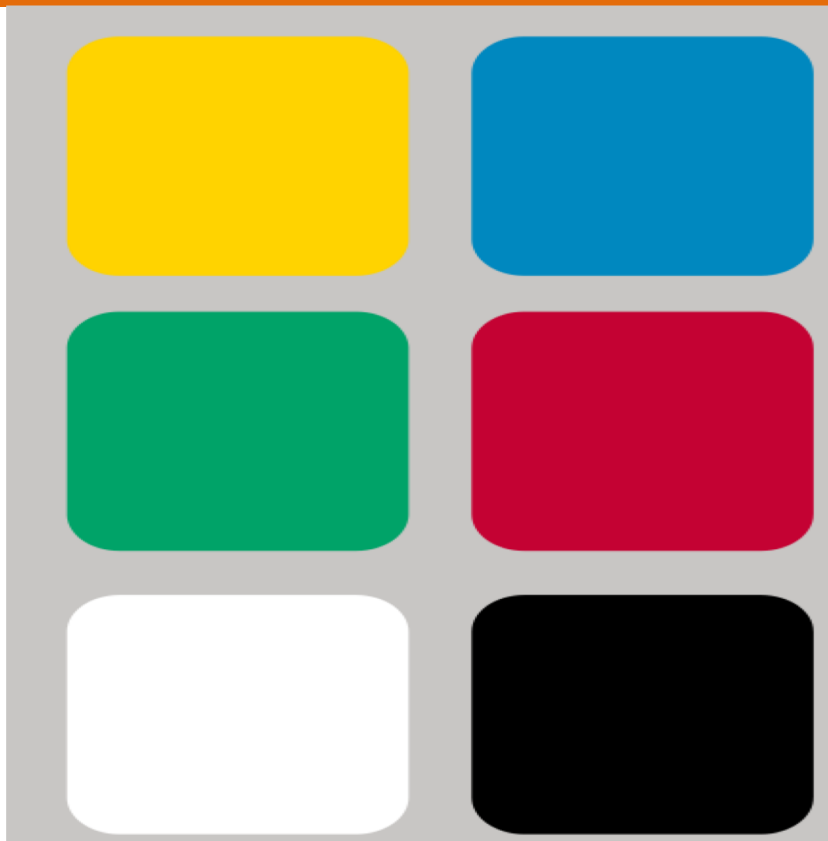
Peculiarities of human color perception

Brain does not use raw RGB signals, but rather some differences: white/black, red/green, yellow/blue.

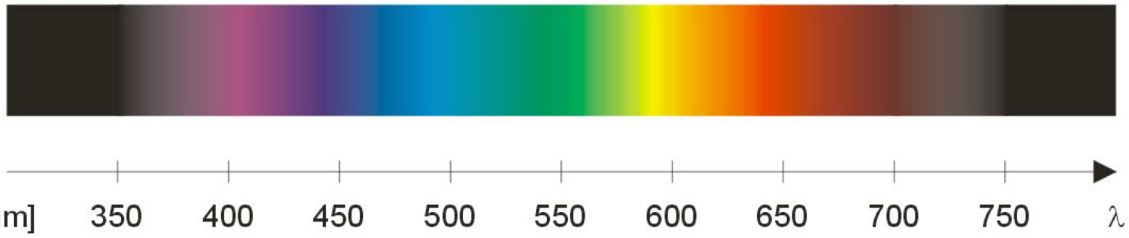
- There is no “pure red” sensor!
- Leads to complementary-color afterimages
- Impossible colors: “reen” = red + green
- Color wheel puzzle: physically impossible but …
- In RGB: resolve up to 500 levels of brightness
- (cf: 8 bits / channel enough for most situations)
- Relaxation times are different for different cells



Partially explains many puzzles in the color perception



Combinations of these color pairs are perceived as separate colors, yet they do not correspond to any wavelength!

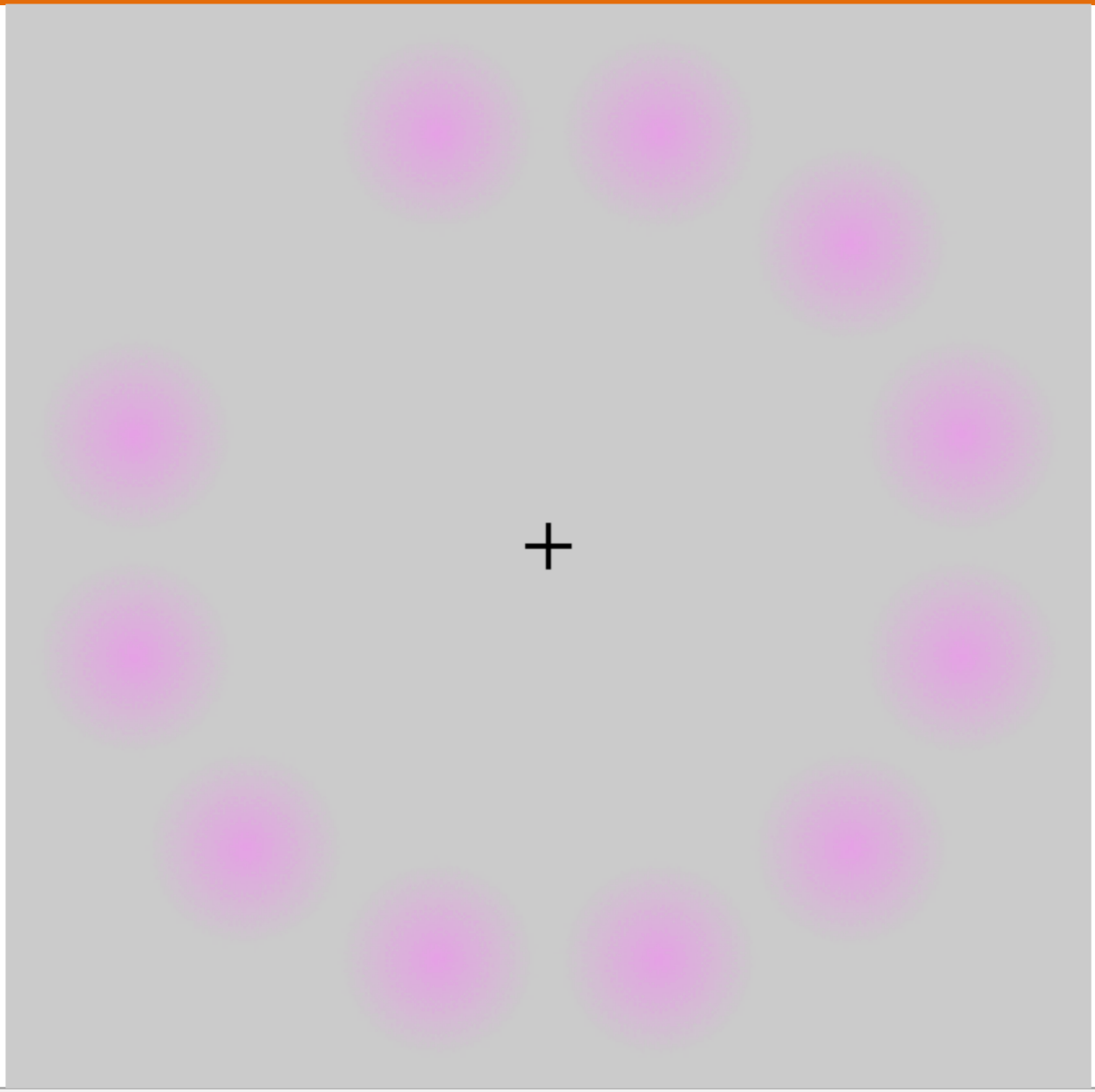


Relaxation of color perception

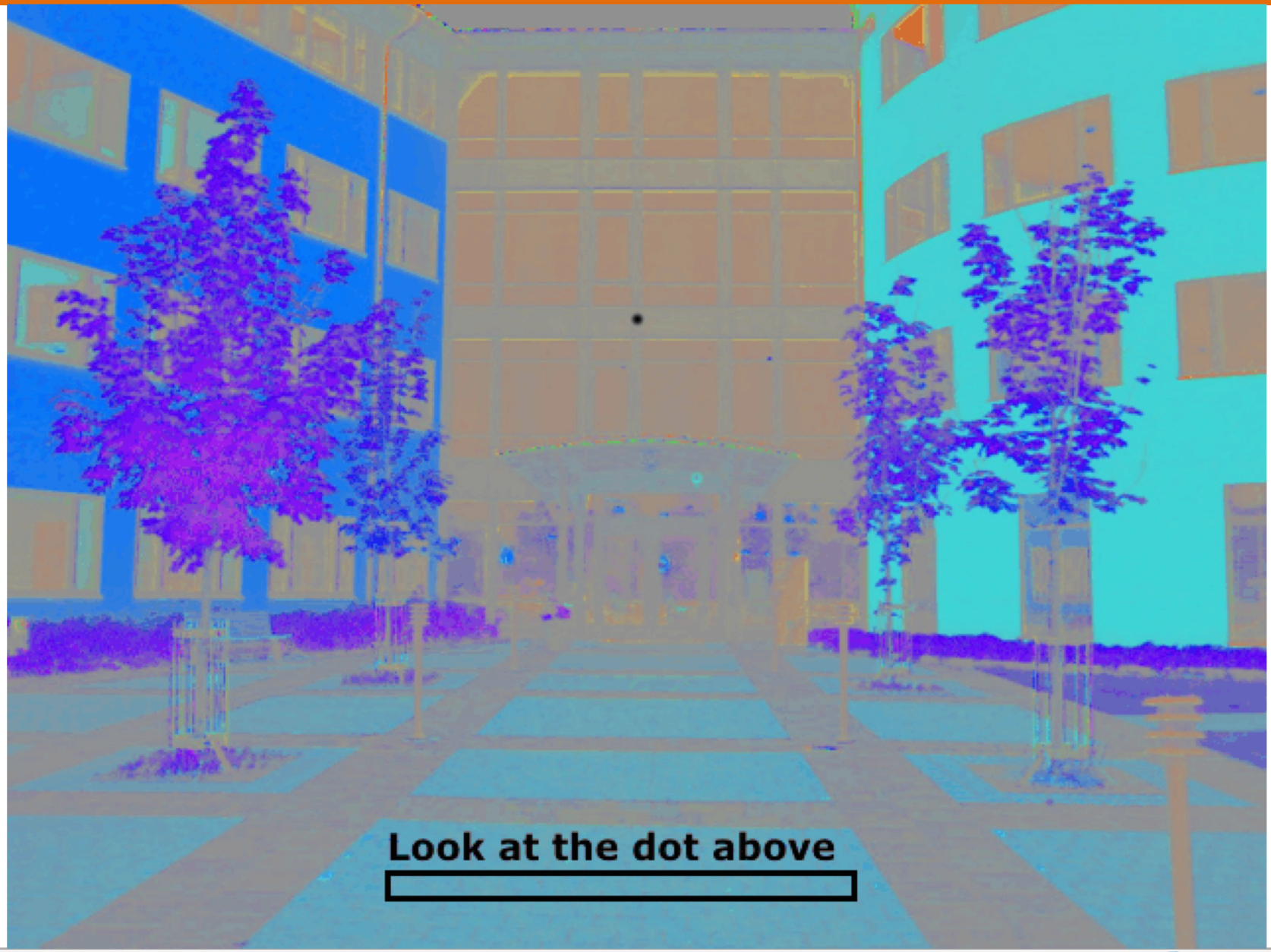
Focus on the center for 10 seconds…



Relaxation of color perception

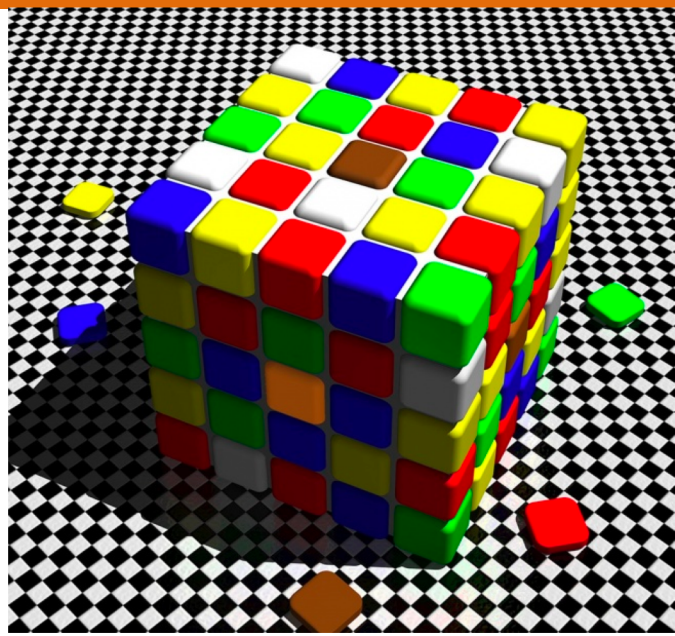


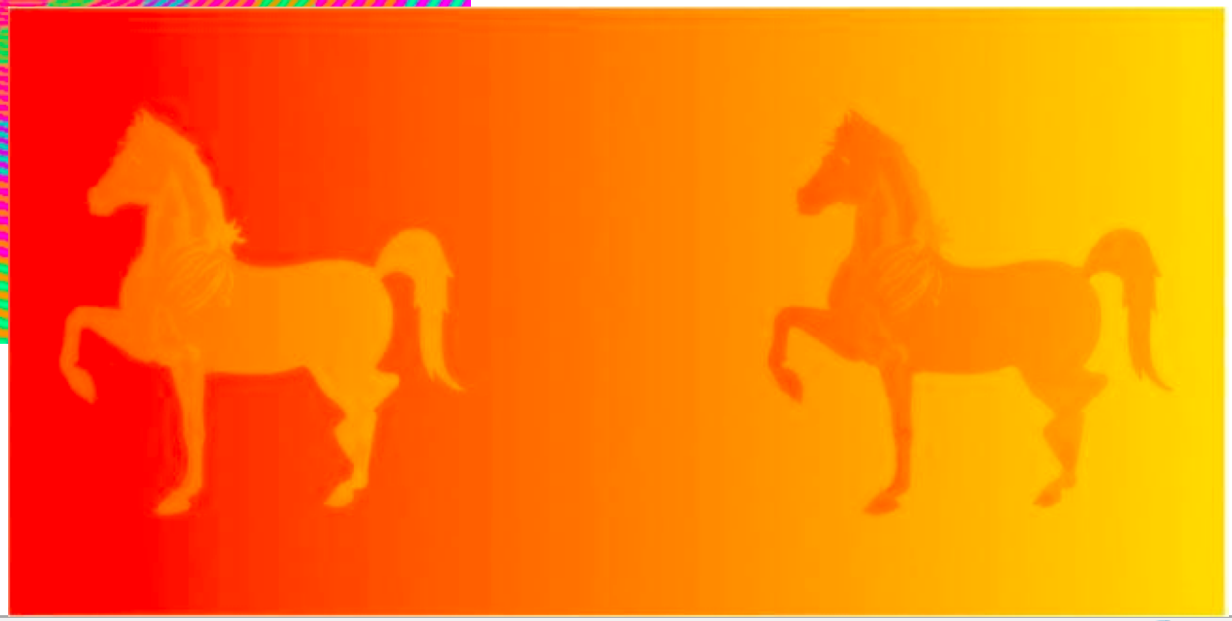
Relaxation of color perception



Context dependence of color perception







© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Context dependence of color perception

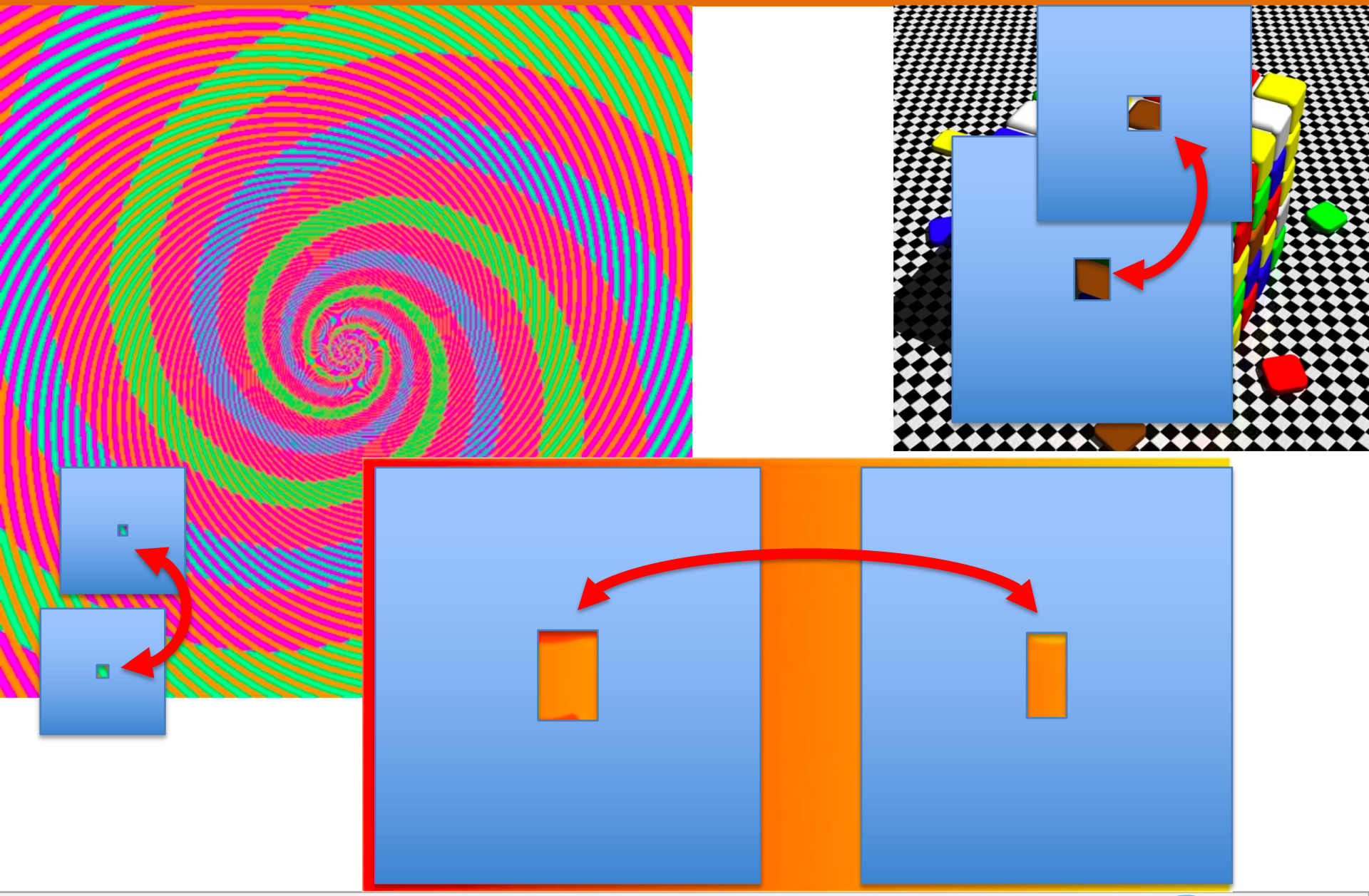
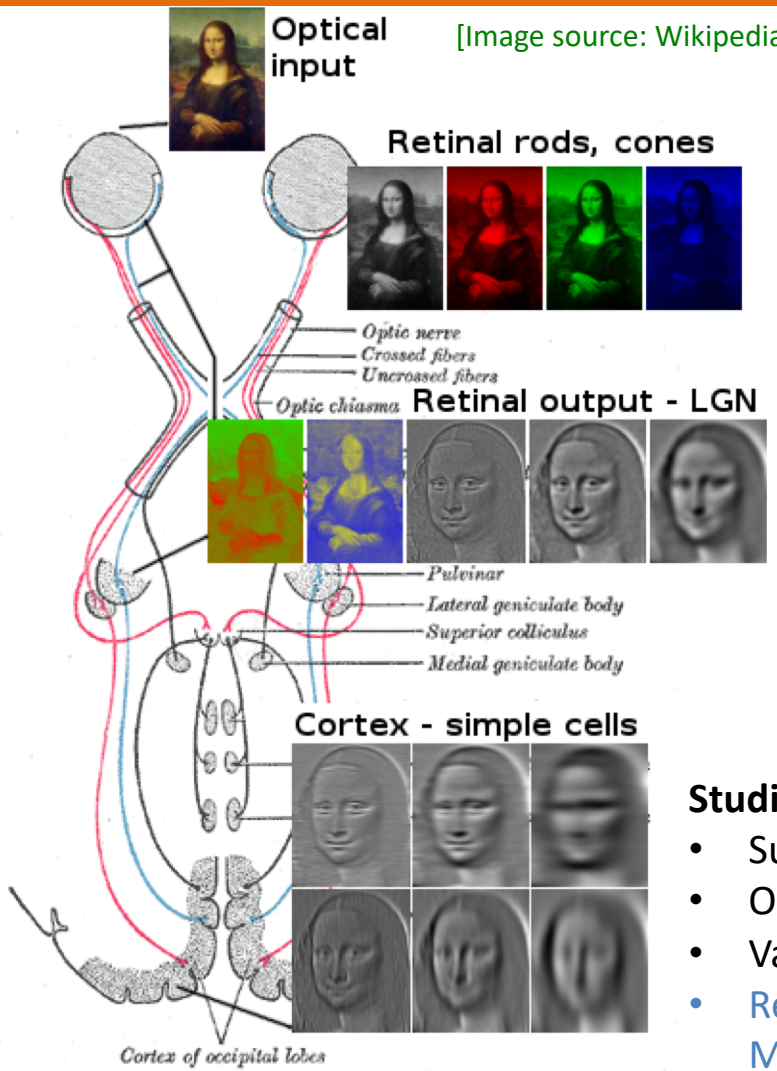


Image processing: deep neural network with feedback



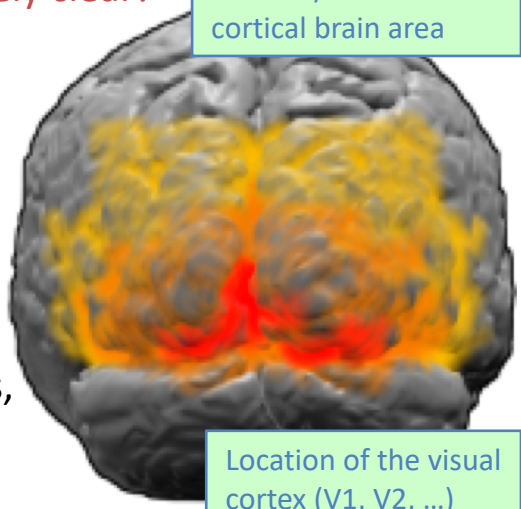
- Besides rods and cones, retina contains several layers of neurons, and a ganglion forming the output to the nerve
- Number of receptor cells / number of nerve fibers: ~ 80:1 (already a very decent rate of compression!)

Further notable processing centers:

- **Optic chiasma** – eye coordination for binocular vision
- **Lateral geniculate body** (LGN) – simple pre-processing, similar to wavelet transform
- **Primary visual cortex** (V1) – simple and complex processing
- Further cortex parts (V2, …) – object identification (?)
- Feedback loops run from the further parts back
- Many processing steps are non-linear
- Functions are still not completely clear!

Studied by various methods:

- Surgical methods - on animals
- Observations of traumatized humans
- Various flavors of EEG
- Recently: Functional Tomography + Machine Learning (Big Data, again!)
- Models: various statistical techniques, deep neural networks (DBN)

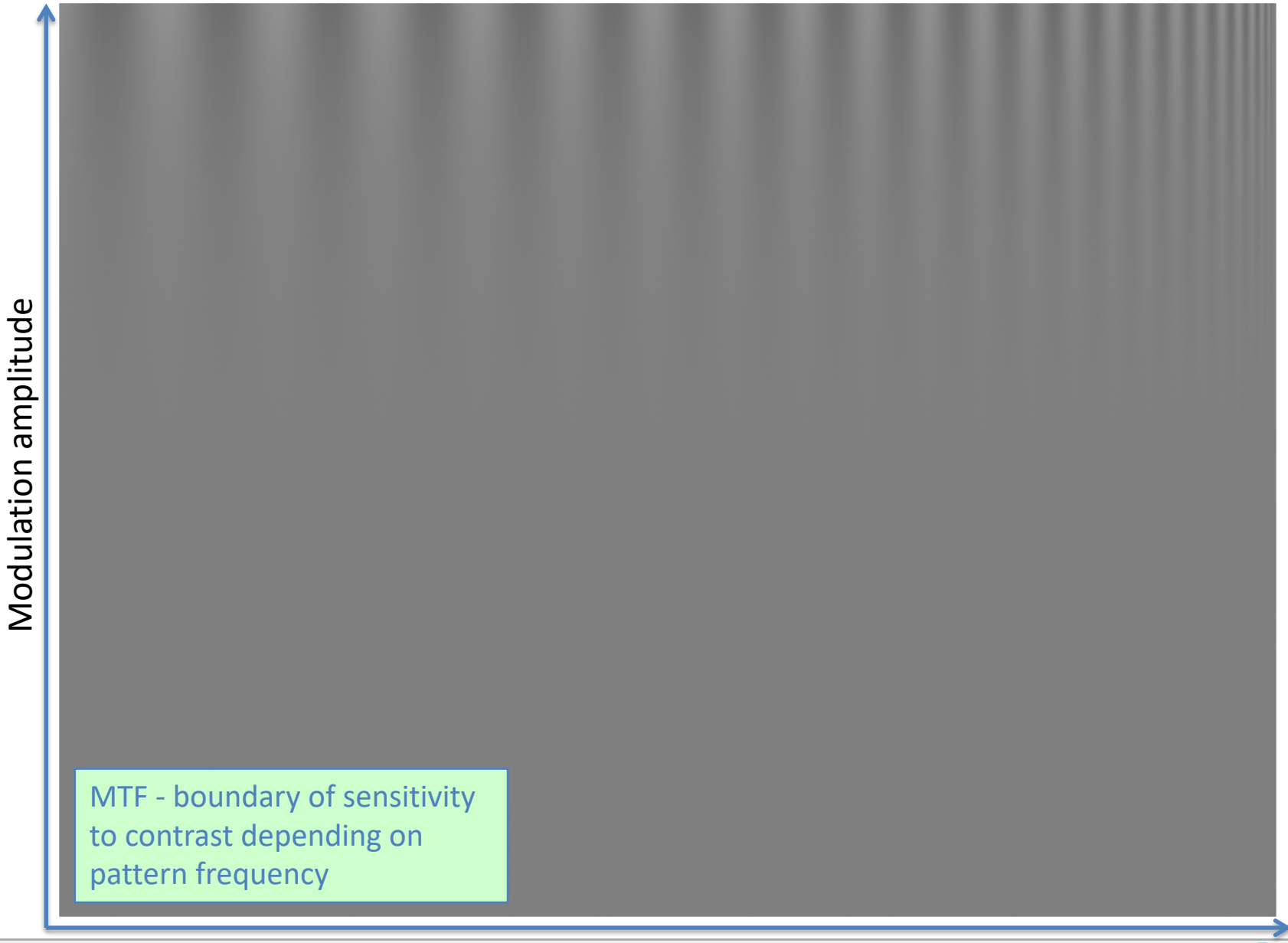


About 1/5 of the total cortical brain area

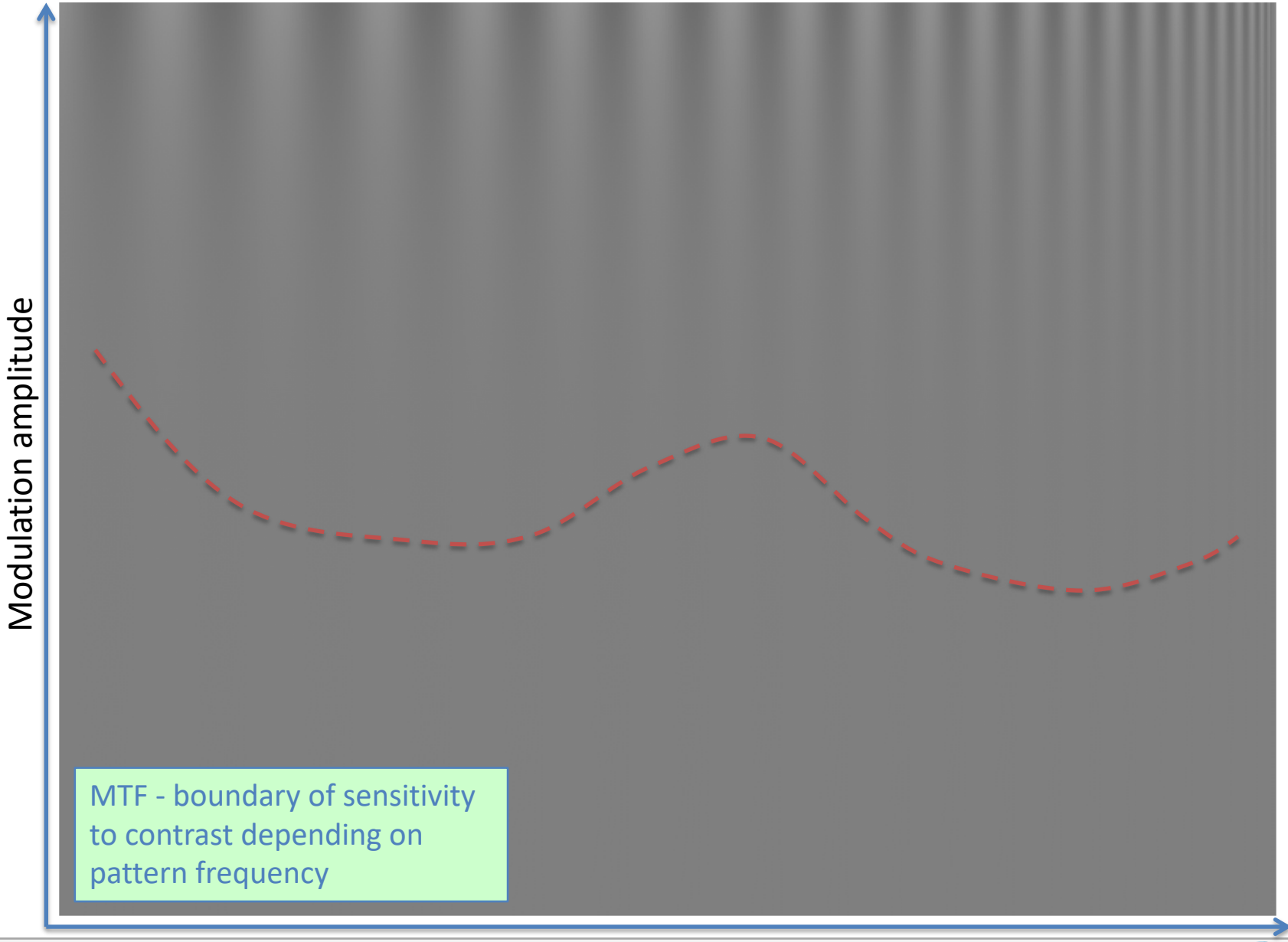
Location of the visual cortex (V1, V2, …)

A more accurate description of functions is mostly unavailable!

Simplest analysis via modulation transfer function

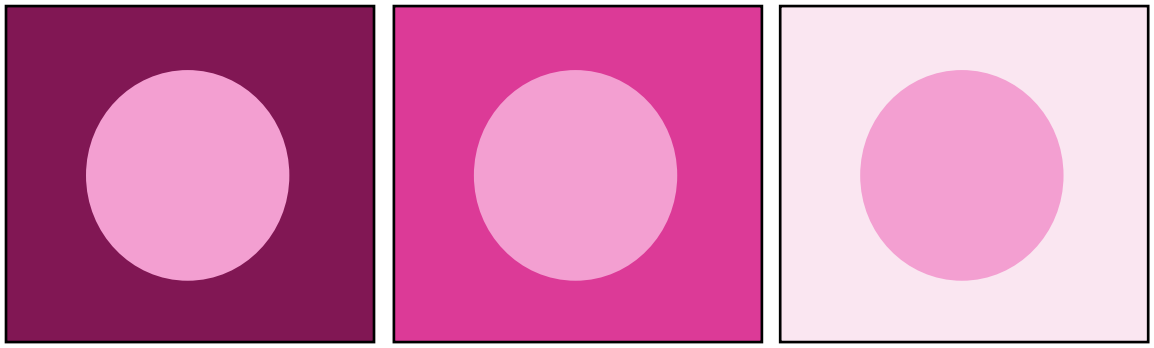


Simplest analysis via modulation transfer function



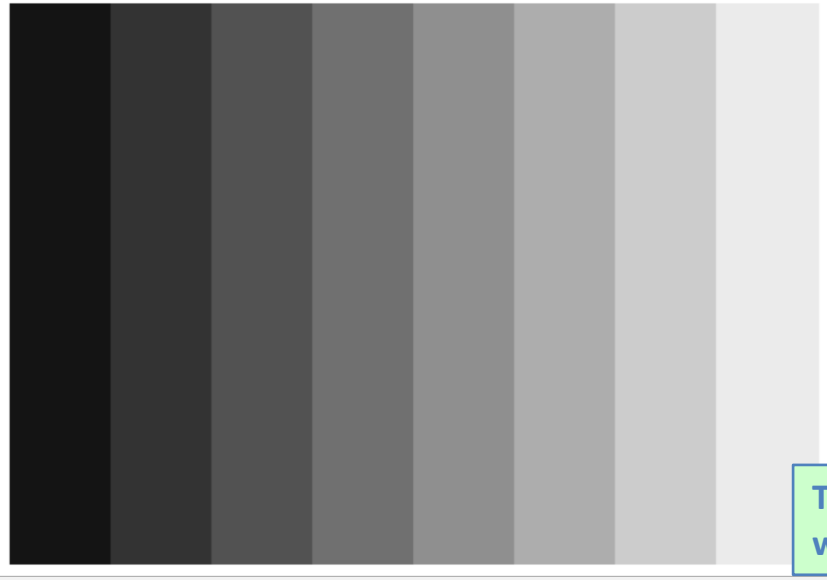
Non-linear effects in visual image processing

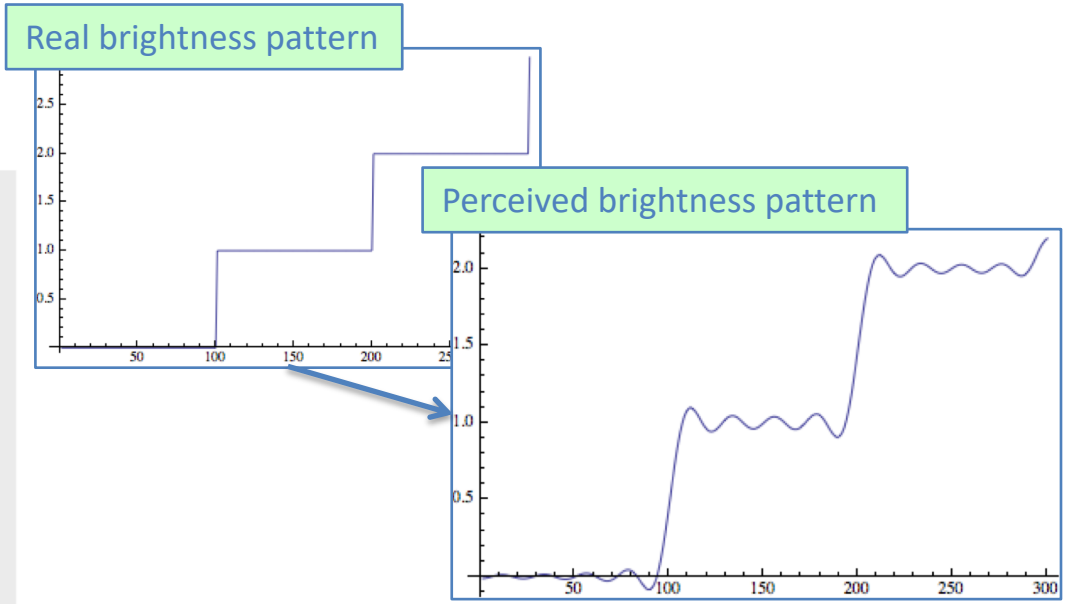
- Context-dependent sensitivity to brightness





- Perceived boundary artifacts

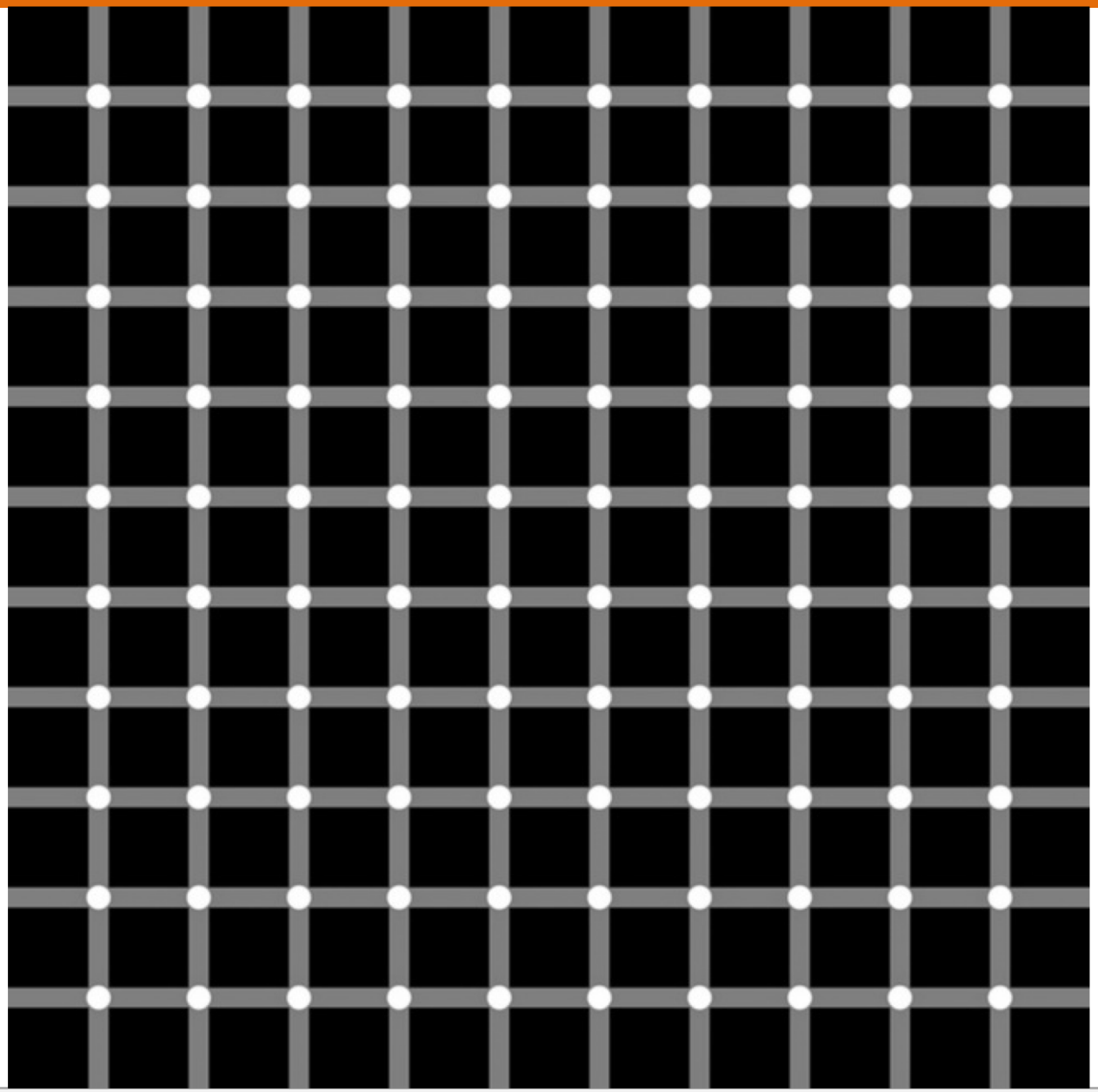




This justifies e.g. DCT-based compression methods (such as JPEG), which do not accurately reproduce hard edges and contrast steps

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Context dependence of brightness perception



© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

How nerve cells (neurons) work

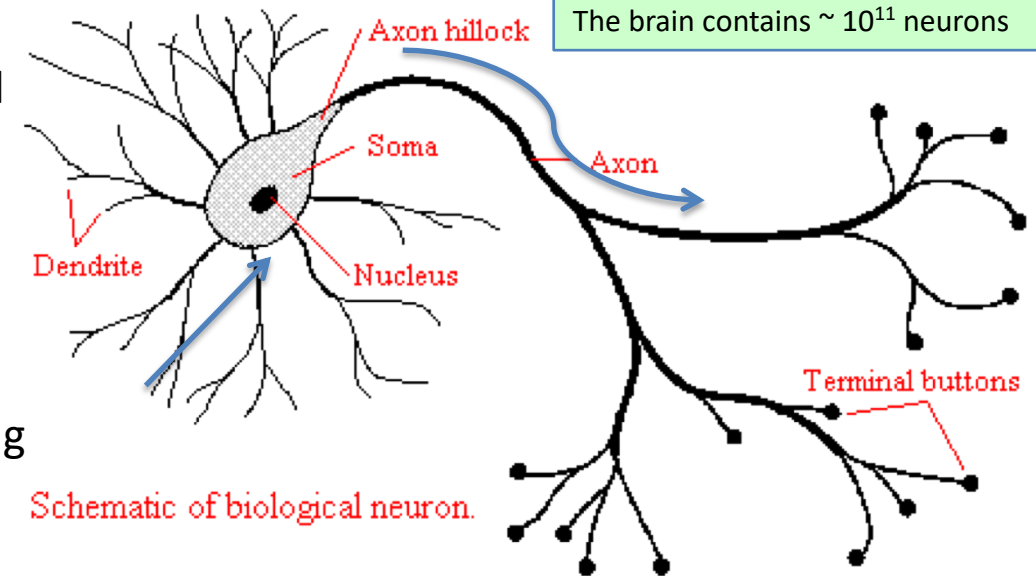
Neuron function: receive signals from other neurons, process, output the resulting signal

- **Axon:** can be many cm long
- **Action potentials ("spikes"):** short (1 ms) "electric" excitations travelling via axon
- Shape and magnitude of spikes are fixed
- Information is carried by the spike origin (i.e. sending neuron) and the time of firing

Output of a neuron consists of sequence of spikes (a **"spike train"**); a full description would record all spike timestamps

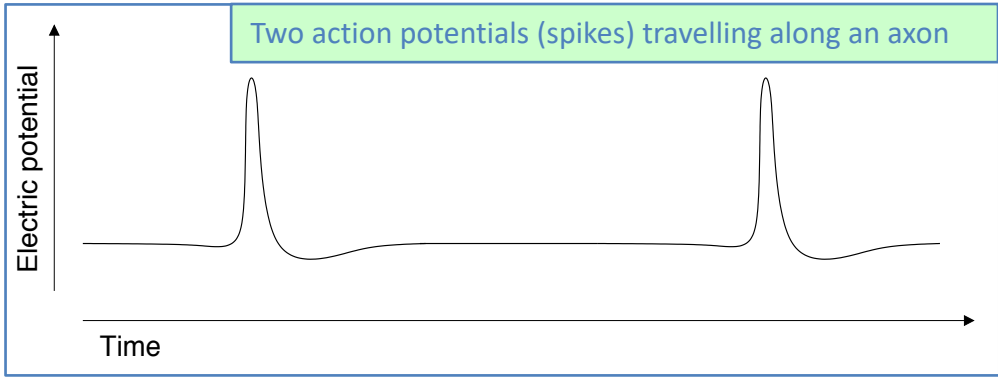
Simplification: use the **"firing rate"** (number of spikes emitted per second) as a metric

- A single scalar number, characterizes the level of activity of a cell
- **Spontaneous firing rate:** ~ 1Hz



Schematic of biological neuron.

The brain contains ~ 10^{11} neurons



Two action potentials (spikes) travelling along an axon

- Brain surface: neuron bodies ("grey matter")
- Brain volume: inter-connections ("white matter")

- There are many types of neurons, each with a specific activation pattern
- It is hard to track individual neurons, may study some volumetric data
- Studies typically produce some statistical results

Question: what are the neurons optimized for?

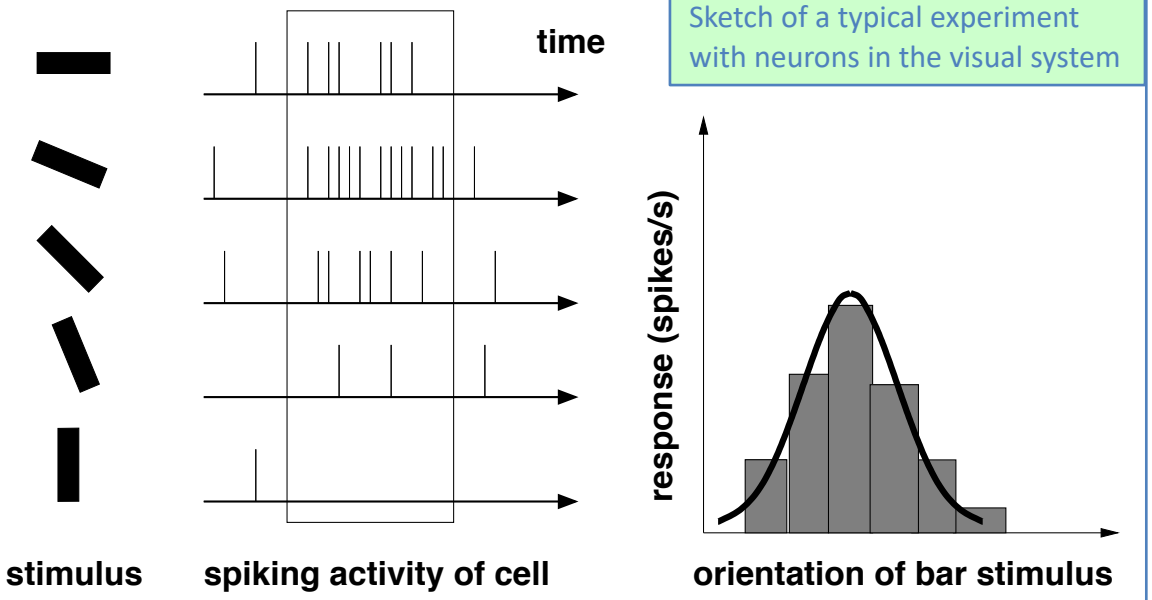
© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Linear neuron models

- Mostly describe ganglion, LGN, and V1 cells (early visual system)
- V1 contains “simple” and “complex” cells
- Method: measure firing rates of cells as a function of visual input
- Vary stimulus, record and model the neuron response
- The first experimental fact: uniform (no-contrast) areas typically yield no response
- The response of a typical neuron depends a small piece of the visual field – receptive field
- We may try and find a pattern that yields the maximum response (i.e. firing rate)…

Could also model a neuron as a function of its inputs (coming through dendrites) – a much harder task! Need to know the NN topology and the functions of neurons in advance

Sketch of a typical experiment with neurons in the visual system



Typical classical receptive fields:

- Retinal ganglion and LGN



- Simple cells in V1



Linear neuron model:

$$r_j = \sum_{x,y} W_j(x,y) I(x,y) + r_0$$

Estimation of receptive fields

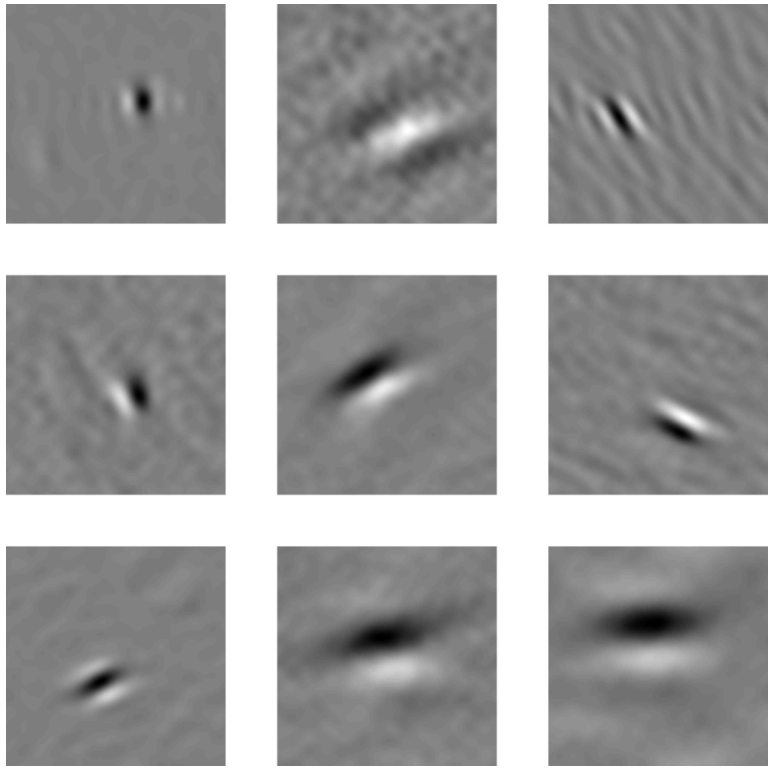
Reverse correlation study:
(rather laborious!)

$$W^* = \arg\min_W \left[d_{rms} \left(\sum_{x,y} W_j(x,y) I(x,y) + r_0, r_{\exp} \right) \right]$$

Convenient wavelet shape parameters:
Windowed Fourier transform kernels (Gabor functions)

- Orientation
- Frequency of oscillations
- Phase
- Width of envelope
- Length of envelope
- Location on retina

RFs are localized in position, orientation, and frequency – a very non-trivial fact!



Another justification: “frequency channels” in image processing (supported by experiments):

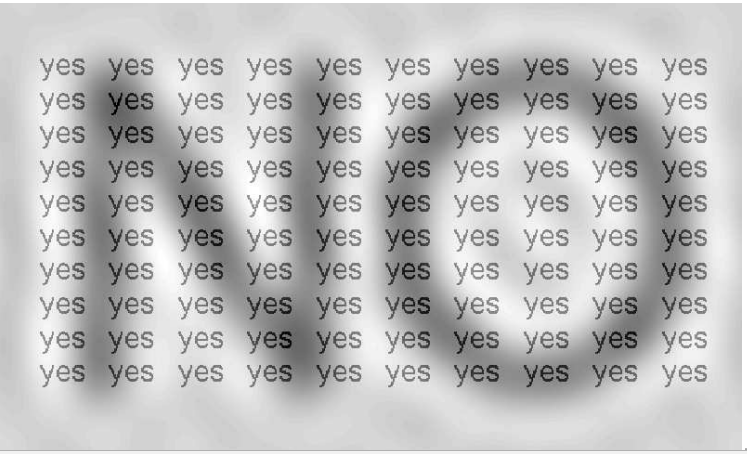
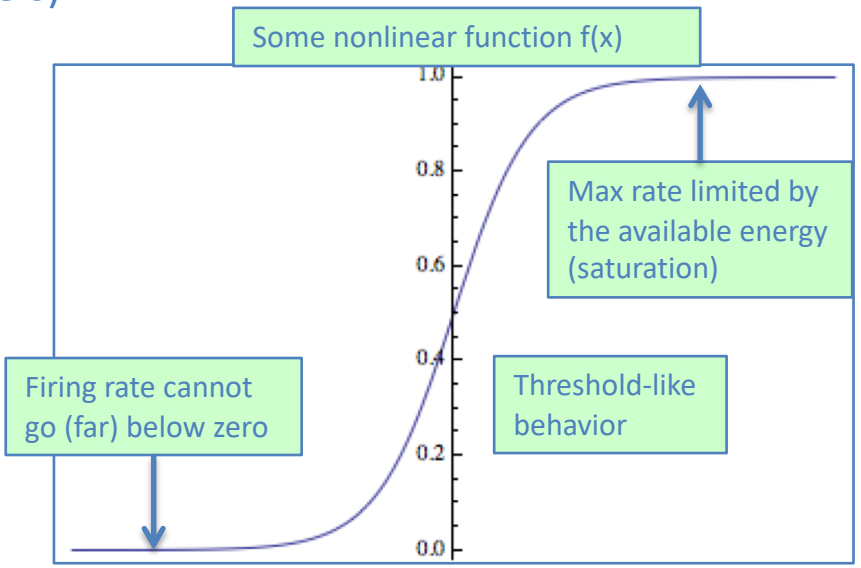


Fig. 3.7: Receptive fields of simple cells estimated by reverse correlation based on single-cell recordings in a macaque monkey. Courtesy of Dario Ringach, UCLA.

Non-linear neuron models

Simplest non-linearity (linear-nonlinear neuron models):

$$r_j - r_0 = f\left(\sum_{x,y} W_j(x,y) I(x,y)\right)$$



Complex non-linear cells and energy models:

(pooling the responses from several simple cells)

$$r_j - r_0 = \left(\sum_{x,y} W_j^{(1)}(x,y) I(x,y)\right)^2 + \left(\sum_{x,y} W_j^{(2)}(x,y) I(x,y)\right)^2$$

Phase-invariant output

Same effects of non-negativity and saturation apply

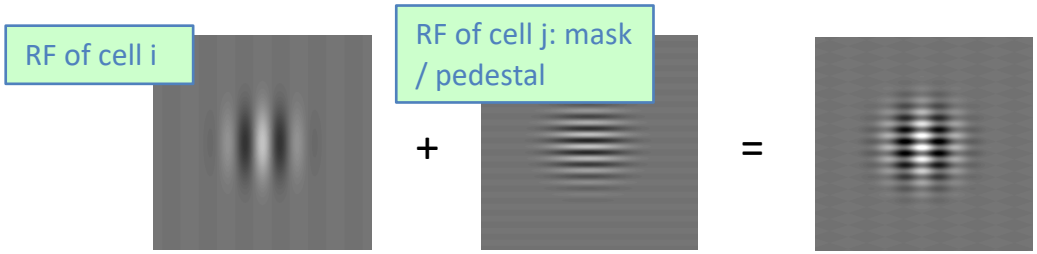
E.g. phase-shifted Gabor functions (phase-shifted by 90°)

There exist other kinds of complex cells…

Interactions between visual neurons

Principal mode of neuron interaction: inhibition

- When a cell **j** is active, output from a cell **i** decreases
- Example: two cells with orthogonal receptive fields:



$$I_1(x,y) + I_2(x,y) = I(x,y)$$

$$r_i(I(x,y)) < r_i(I_1(x,y))$$

- Known as "contrast gain control"
- Prevents saturation in high-contrast environments (e.g. by daylight)
- Effectively normalizes the variations of luminance over the scene
- In retina, similar normalization of luminance ("luminance gain control"), roughly equivalent to DC component subtraction

Divisive normalization model:
(currently considered as the most accurate model of simple cells)

$$r_j = \frac{f\left(\sum_{x,y} W_j(x,y) I(x,y)\right)}{\sum_{i=1}^{K} f\left(\sum_{x,y} W_j(x,y) I(x,y)\right) + \sigma^2}$$

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Topographic organization of neurons

- In the retina, the receptive fields of ganglion cells are necessarily linked to their physical positions
- Experimental fact: adjacent neurons in LGN and V1 tend to have receptive fields in the adjacent areas of visual field (retinotopy)
- In V1, the orientation (location, frequency, phase) of RFs shift gradually along the surface of cortex
- Similar organization is true for the other senses: tonotopic map of hearing (frequency-based), complete body map for the sense of touch, etc.

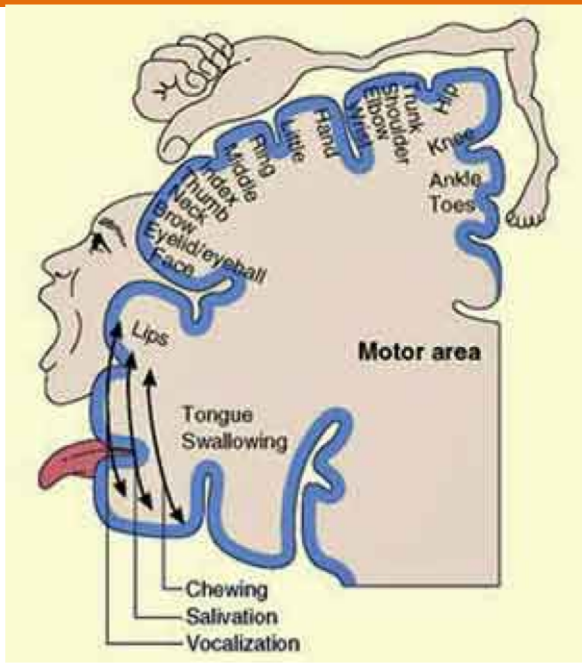
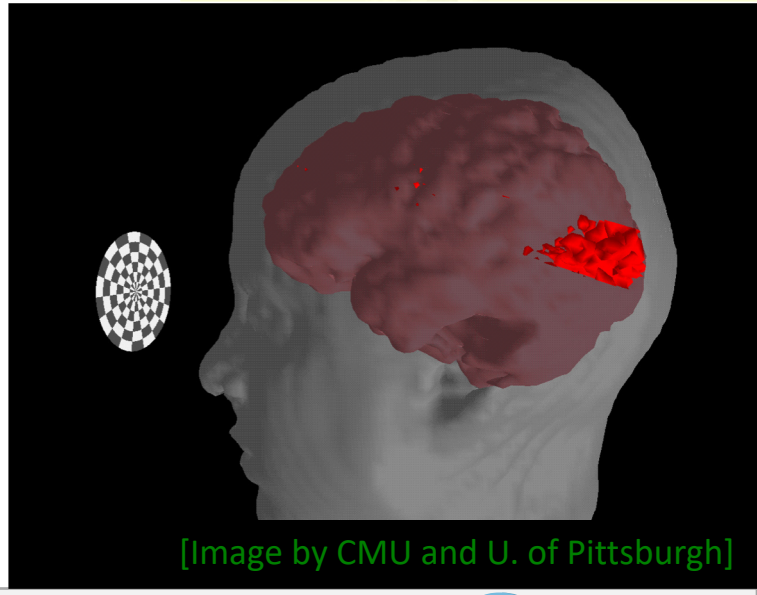


Image processing beyond the primary visual cortex

- Further areas – V2, V4, V5 (“extrastriate cortex”)
- V5 analyzes motion
- V4 extracts contours, edges, crosses, circles, non-Cartesian gratings, color, temporal changes
- The functions of V2 are not yet clear (similar to V1?)
- Human extrastriate cortex is very different from that of monkeys, harder to study and generalize



[Image by CMU and U. of Pittsburgh]

© 2016-17 Alexey Pak, Lehrstuhl für Interaktive Echtzeitsysteme, Fakultät für Informatik, KIT

Functional neuroimaging and “thought identification”

[Keim, Brandon’ 08]

- fMRI tracks in real time the consumption of oxygen in the brain (as a representative measure of neuron activity)
- Current spatial resolution: **2-3** mm
- If a subject sees an image, the respective 3D brain activity pattern may be recorded
- Inverse problem: given an activity pattern, may guess which image (of a pre-defined set) is observed!
- Identification accuracy: **72-90%** (random choice: **0.8%**)


[Image source: Wikipedia]

- One interesting extension: recovering of an unknown observed image based on a “dictionary” of learned images
- Again, a very demanding Big Data application!
- May even recover scenes that one visualizes during sleep!



- So far, relatively limited success, but the methods develop rapidly
- Further progress expected from increasing the resolution to µm-scale with e.g. PET, SPECT, MEG, …

