Image, Video, and Multidimensional Signal Processing

WS 2020/2021

Prof. Dr.-Ing. André Kaup

Chair of Multimedia Communications and Signal Processing

Friedrich-Alexander University Erlangen-Nürnberg

Remarks

Credits: L2 + E2, 4 SWS, 5 ECTS

Lecture: Monday, 12:15 – 13:45, H9 video recording
Prof. Dr.-Ing. André Kaup
Cauerstr. 7, Room 06.031, andre.kaup@fau.de

Exercises: Wednesday, 16:15 – 17:45, video conference & X2Go access
Andreas Spruck, M.Sc.
Cauerstr. 7, Room 06.020, andreas.spruck@fau.de

Graduate level course for master students in engineering

- *Restricted elective subject* for IuK, CME, CE, and ASC
- *Elective subject* for EEI, WING-IKS, MT, and others

Lecture notes: can be downloaded as PDF from StudOn printed copies available

Grading: Written exam, 90 min, date to be set by examination office, two-sided DIN A4 handwritten formulary and calculator allowed

Time Table WS 2020/2021

Monday, 12:15 – 13:45, H6 vrec		Wednesday, 16:15 – 17:45, vconf / X2Go	
02.11.20	Lecture 1	04.11.20	Exercise 1
09.11.20	Lecture 2	11.11.20	Exercise 2
16.11.20	Lecture 3	18.11.20	Exercise 3
23.11.20	Lecture 4	25.11.20	Exercise 4
30.11.20	Lecture 5	02.12.20	Exercise 5
07.12.20	Lecture 6	09.12.20	Exercise 6
14.12.20	Lecture 7	16.12.20	Exercise 7
21.12.20	Lecture 8	23.12.20	Exercise 8
28.12.20	–	30.12.20	–
04.01.21	–	06.01.21	–
11.01.21	Lecture 9	13.01.21	Exercise 9
18.01.21	Lecture 10	20.01.21	Exercise 10
25.01.21	Lecture 11	27.01.21	Exercise 11
01.02.21	Lecture 12	03.02.21	Exercise 12
08.02.21	Lecture 13	10.02.21	Exercise 13

References

J.-R. Ohm, "Multimedia Content Analysis", Berlin: Springer-Verlag, 2016.
Covers image, video, and audio signals, focus on features, signal decomposition, consistent mathematical treatment, very comprehensive.

J. W. Woods, "Multidimensional Signal, Image, and Video Processing and Coding", Amsterdam: Academic Press, 2nd edition, 2012. *Covers image and video processing and transmission, consistent mathematics, very concise.*

R. C. Gonzalez, R. E. Woods, "Digital Image Processing", Upper Saddle River: Pearson Education, 4th edition, 2018. *Textbook on basics of image processing with focus on filtering, restoration, and segmentation.*

A. K. Jain, "Fundamentals of Digital Image Processing", Englewood Cliffs: Prentice Hall, 1989. *Rigorous mathematical treatment of basic topics, signal processing point of view, my personal favorite.*

Goal of Lecture

Basic understanding in processing of image and video data

- Point and binary operations
- (Color spaces) not this WS 2020/21
- Multidimensional signals and systems theory
- Data interpolation
- Image feature detection and matching
- Image and video segmentation
- (Transformation domain image processing) not this WS 2020/21

Listening requires previous knowledge in

- Discrete Signals and Systems

Helpful are the following lectures (but no course prerequisite)

- Digital Signal Processing
- Pattern Recognition

Contents

*) Greyed chapters contain supplementary content that will not be part of lecture in WS 2020/2021

1 Point Operations

1.1 Digital Images

1.2 Image Histogram

1.3 γ-adjustment

1.4 Image Averaging

1.5. High Dynamic Range (HDR) Imaging *)

1.6 Image Subtraction *)

*) Greyed chapters contain supplementary content that will not be part of lecture in WS 2020/2021

1.1 Digital Images

An **image** is a continuous 2D light intensity signal $s(x,y)$, where x and y are spatial coordinates and s at (x,y) is related to the brightness or color of the image at that point

A **digital image** is the representation of a continuous image by a 2D array of discrete samples $s[m,n]$ called **pixels** (from "picture element")

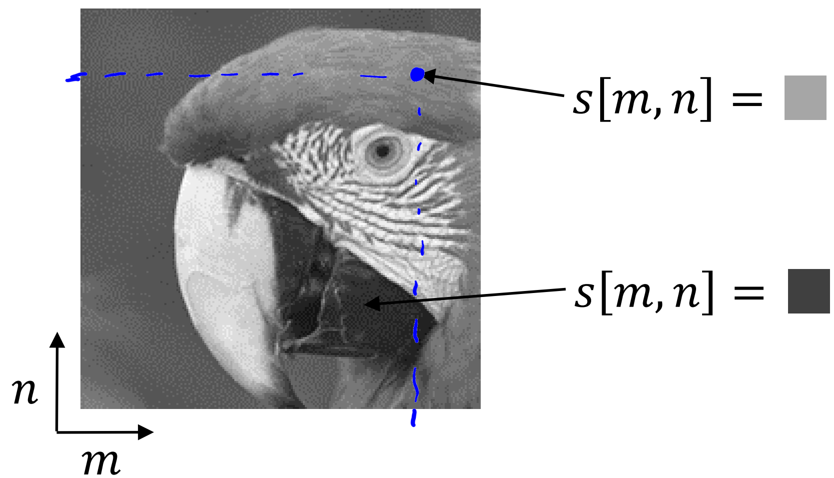
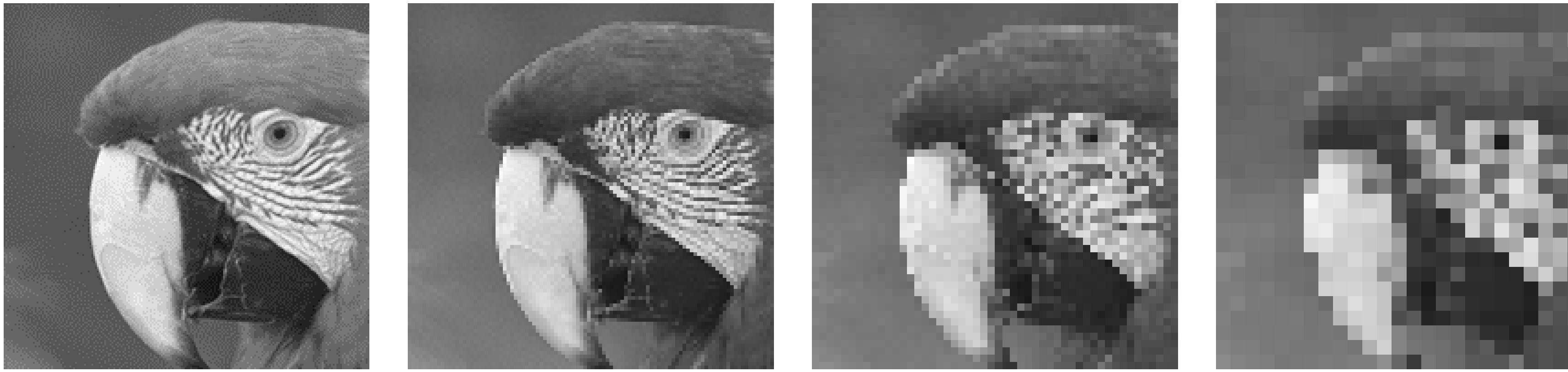


Image matrix notation: $\boldsymbol{S}_{M\times N} = (s[m,n] | m = 0, \ldots M-1, n = 0, \ldots, N-1)$

(also 1...M , 1...N)

Image Resolution



200x200 | 100x100 | 50x50 | 25x25

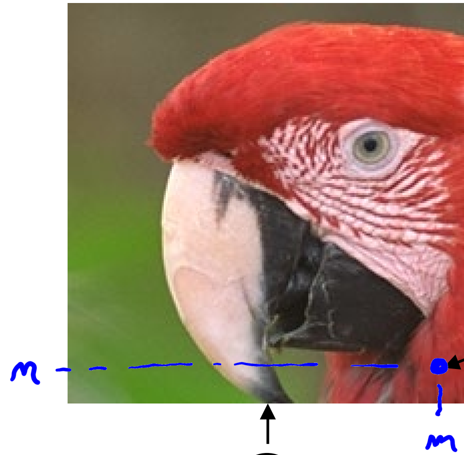
These images have been created by pixel repetition

- Nearest neighbor approach , zero order interpolation

We can do better

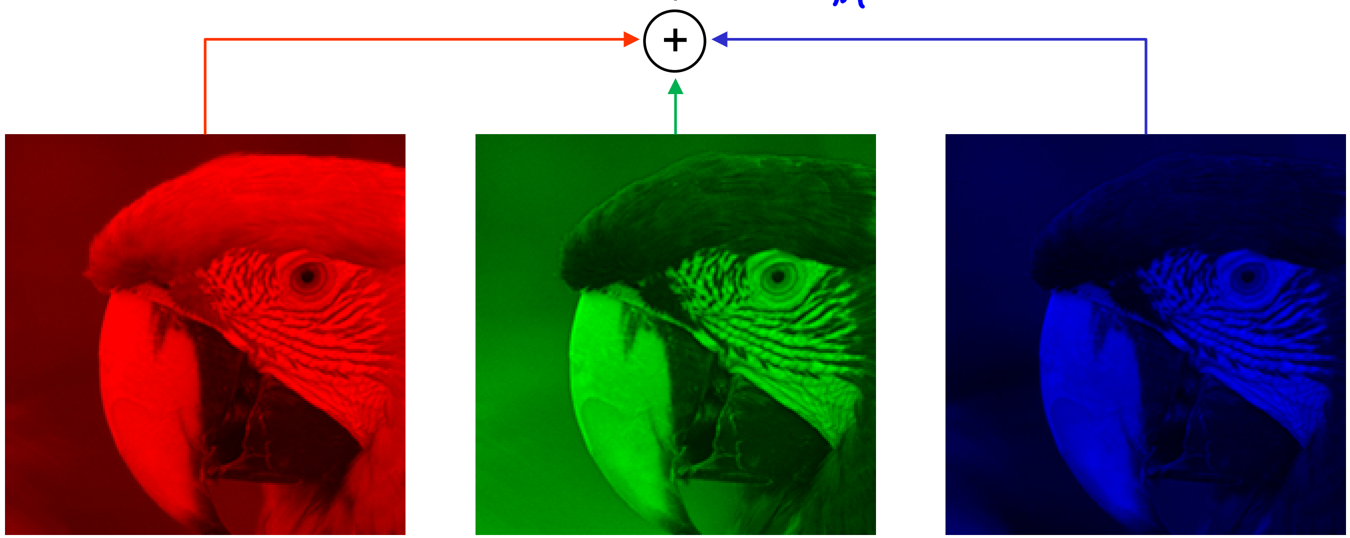
- Pre-filtering to avoid aliasing
- Smooth interpolation

Color Images



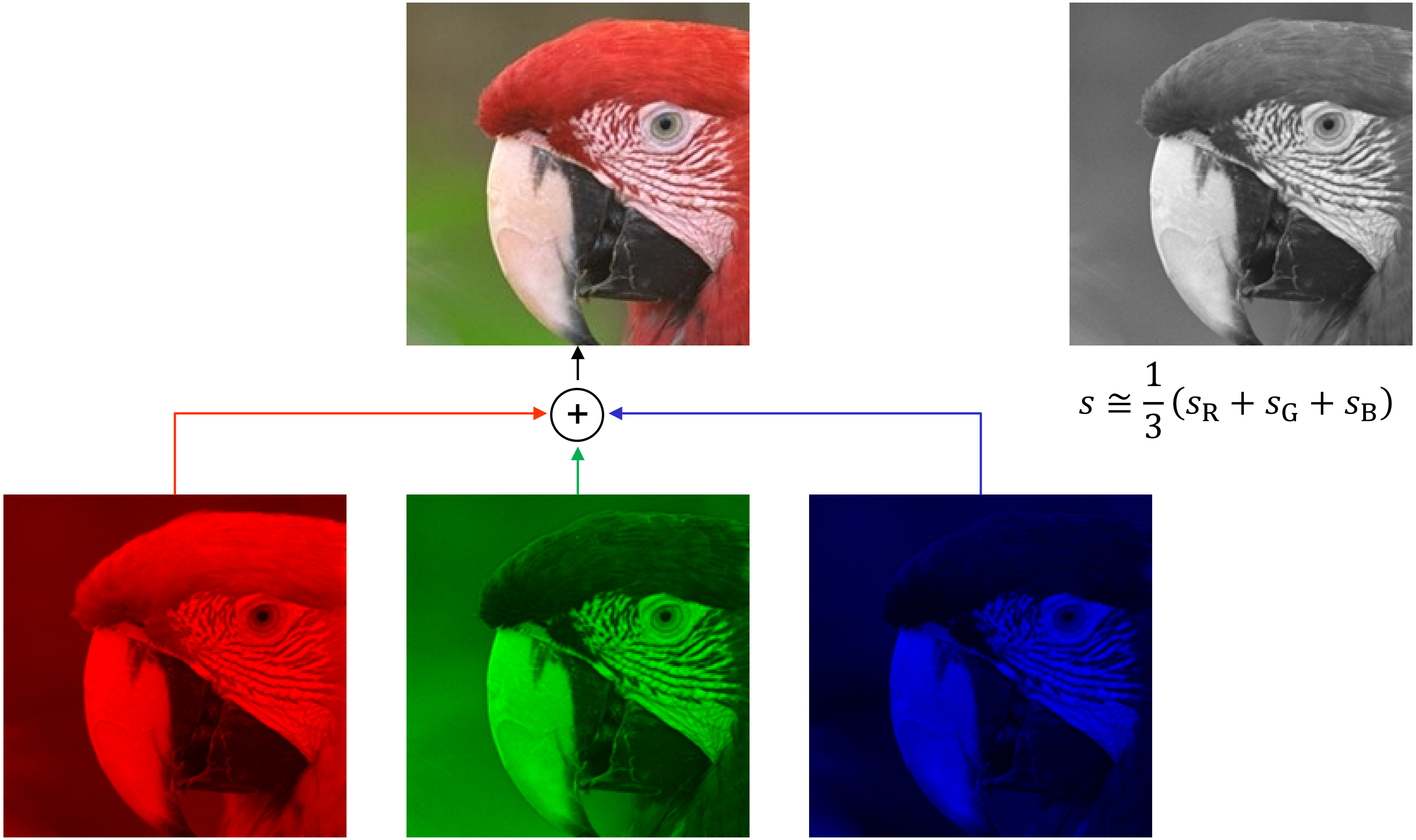
Each pixel is a vector of three color components:

$$\boldsymbol{s}[m,n] = \begin{bmatrix} s_{\mathrm{R}}[m,n] \\ s_{\mathrm{G}}[m,n] \\ s_{\mathrm{B}}[m,n] \end{bmatrix} =$$



$s_{\mathrm{R}}[m,n]$ $s_{\mathrm{G}}[m,n]$ $s_{\mathrm{B}}[m,n]$

Monochromatic Images



$$s \cong \frac{1}{3}(s_R + s_G + s_B)$$

$s_R[m,n]$ $s_G[m,n]$ $s_B[m,n]$

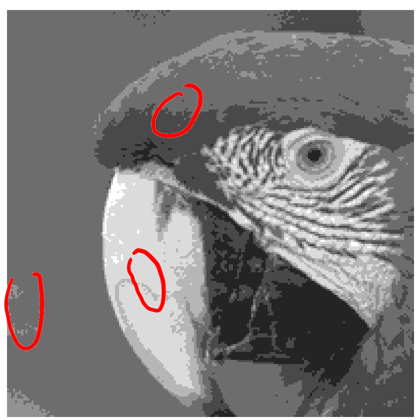
Different Numbers of Gray Levels

256 = 8 bit

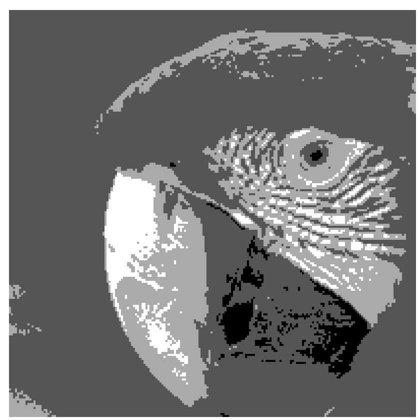
32

16

8



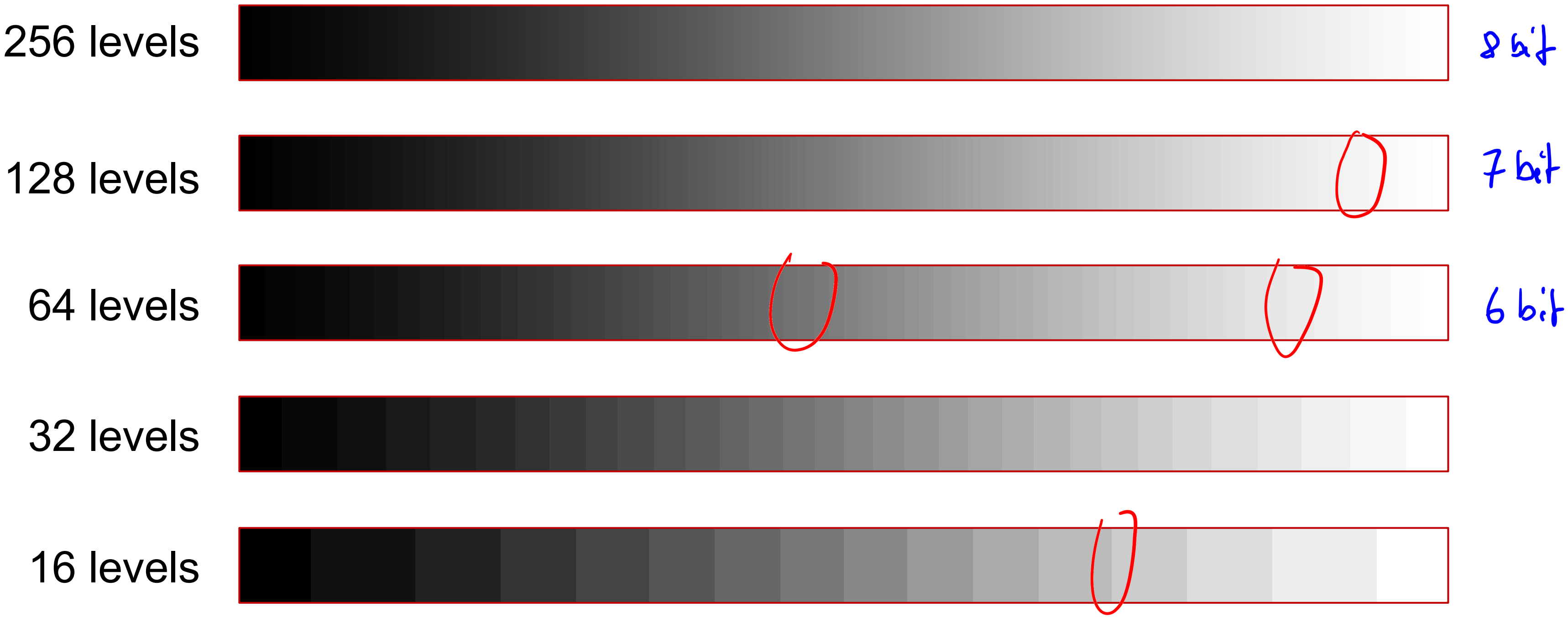
4



2 = 1 bit

Image Quantization

Contouring is most visible for a grey level ramp ("gradient")



Typically 8 bits/pixel, i.e., $2^8 = 256$ gray levels

1.2 Image Histogram

Histogram represents gray level distribution

- for B-bit images we have 2^B bins
- $\text{bin}(i) = m$ # count
- Meaning: there are m pixels with gray level value equal to i

Can be interpreted as an estimate of the **probability mass function** (PMF) of the underlying random process

$$p_X(i) = \text{Prob}(X = i) \cong \frac{1}{\underbrace{MN}_{\text{total \# samples, normalization}}} \text{bin}(i)$$

- Not differentiable
- Bins can be empty

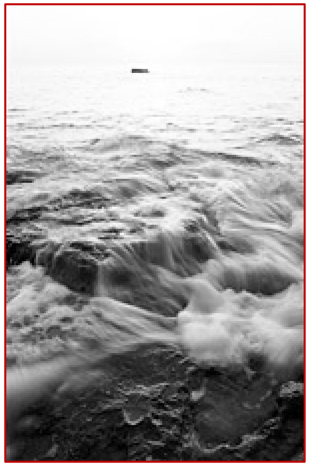
Fewer bins can be also used

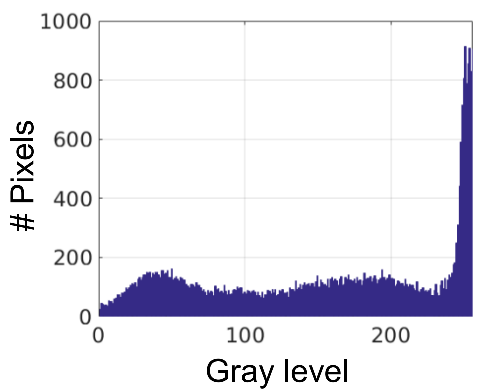
- Coarser resolution
- Smaller sample size

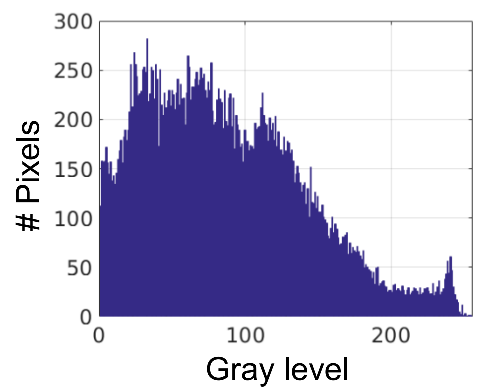
Histogram Examples

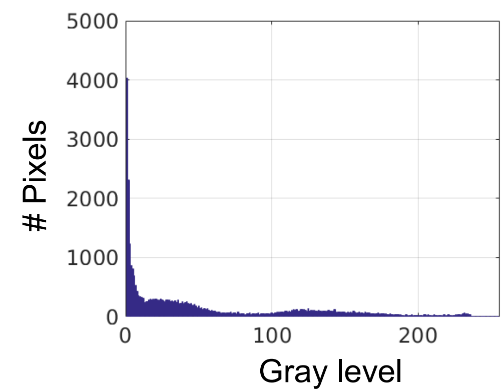
overexposure

underexposure









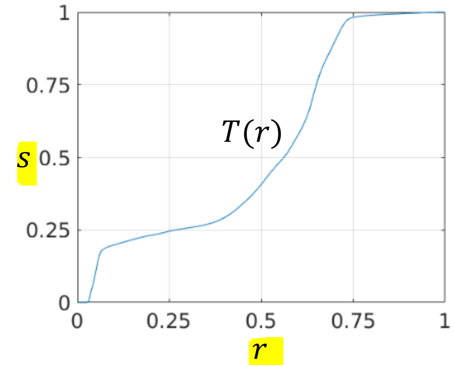
http://www.cambridgeincolour.com/tutorials/high-dynamic-range.htm

Histogram Equalization

In general, uniform histograms yield better contrast

Idea: find a non-linear transformation

$$\boxed{s = T(r)}$$



Properties

- Preserves dynamic range ($0 \leq T(r) \leq 1$)
- Preserves black-white order (monotonically increasing)

Goal: constant pdf

$$\begin{array}{ccc} r & & s = T(r) \\ & \longrightarrow & \\ p_R(r) & & p_S(s) \end{array}$$

From probability theory:

link between p_R and p_S for any given $T(\cdot)$

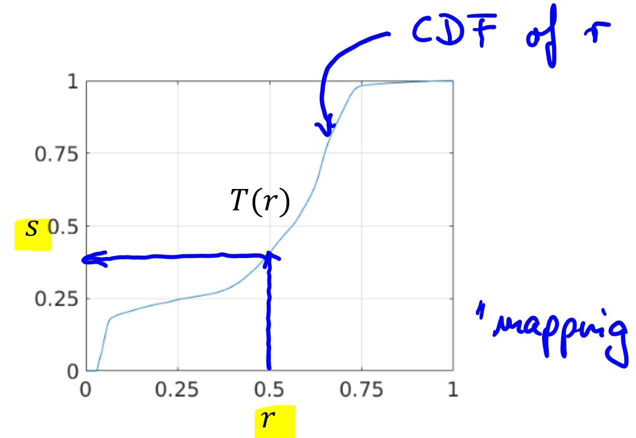
$$p_S(s) = \left[p_R(r) \frac{\mathrm{d}r}{\mathrm{d}s} \right]_{r=T^{-1}(s)}$$

Histogram Equalization

How to get a constant pdf for s? Consider **transformation function**

$$\boxed{s = T(r) = \int_0^r p_R(w)\mathrm{d}w} \; (*) \qquad 0 \le r \le 1$$

(CDF: Cumulative Distribution Function)



⇨ This implies

$$p_S(s) = 1 \quad \text{and} \quad 0 \le s \le 1$$

(Uniform distribution) (full range)

i.e. const.

Proof:

$$p_S(s) = \left[p_R(r)\frac{\mathrm{d}r}{\mathrm{d}s}\right]_{r=T^{-1}(s)} = \left[p_R(r)\frac{1}{p_R(r)}\right]_{r=T^{-1}(s)} = 1$$

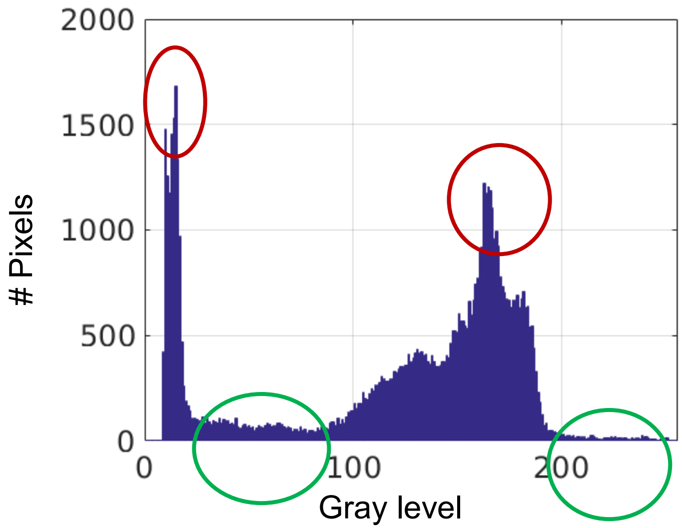
$s=T(r) \Rightarrow \frac{d}{dr}$

$$\frac{\mathrm{d}s}{\mathrm{d}r} = \frac{\mathrm{d}T(r)}{\mathrm{d}r} \overset{(*)}{=} \frac{\mathrm{d}}{\mathrm{d}r}\left[\int_0^r p_R(w)\mathrm{d}w\right] = p_R(r)$$

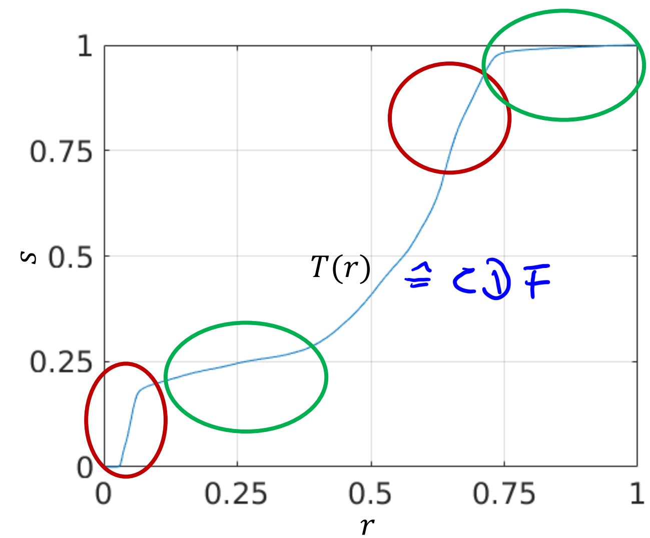
(Leibniz's rule)

Histogram Equalization

Histogram



Equalization function $T(r)$



- Large accumulations of similar colors are "**stretched out**"
- Less frequent colors are "**compacted**"

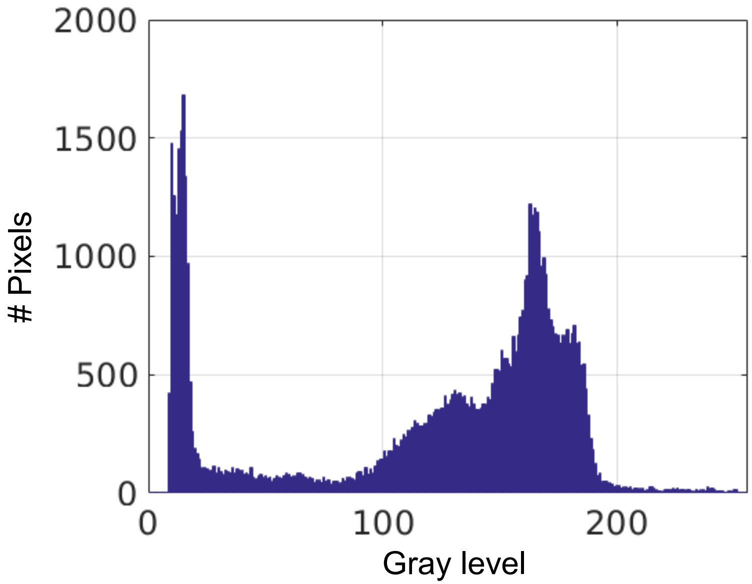
Histogram Equalization Example

Original image

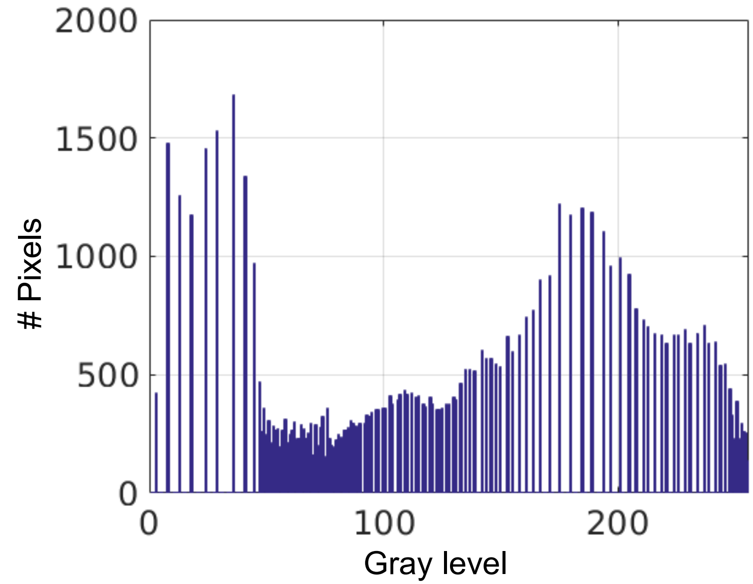
…after histogram equalization

Histogram Equalization Example

Original image



…after histogram equalization



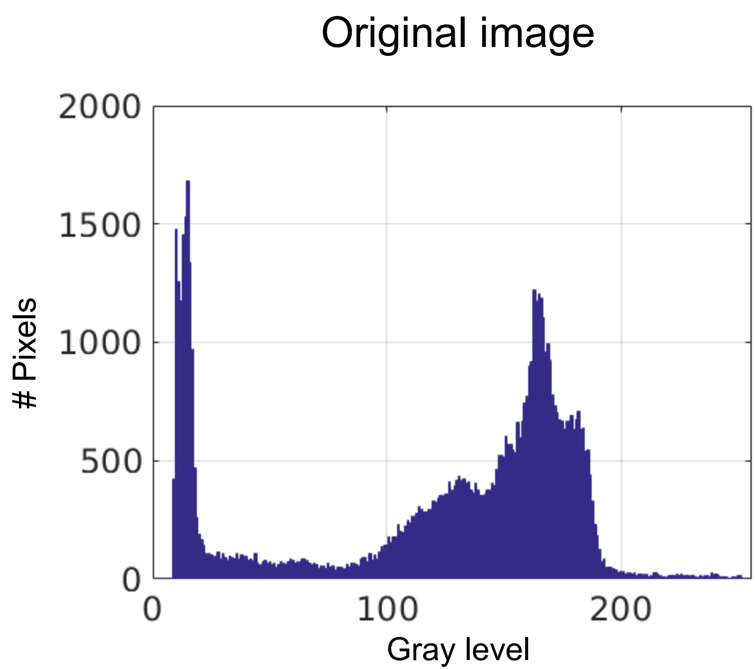
Discrete mapping function in practice:

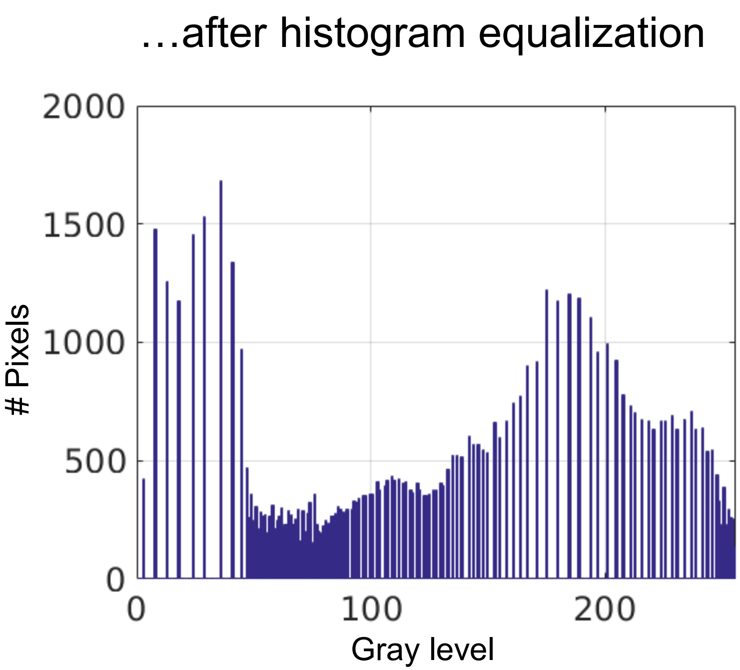
instead of (*) on 1-17

$$s_k = T(r_k) = (L-1)\sum_{j=0}^{k} p_R(r_j)$$

Scale by # grey levels

Histogram Equalization Example



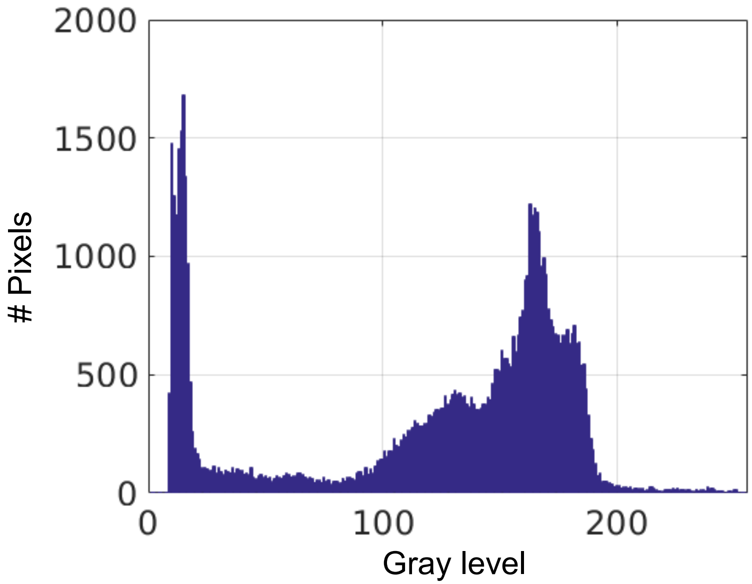


In general, after equalization, the histogram is **not** (perfectly) **flat**

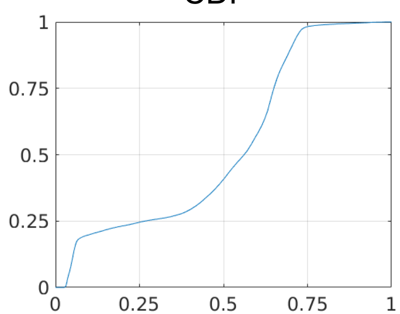
- Histogram is an approximation to pdf
- Empty bins
- **But:** "equalized" CDF is still approximately linear

Histogram Equalization Example

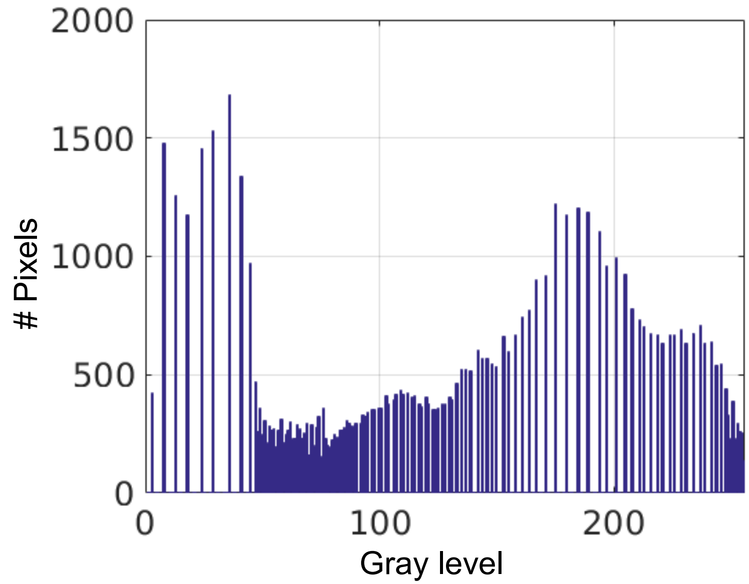
Original image



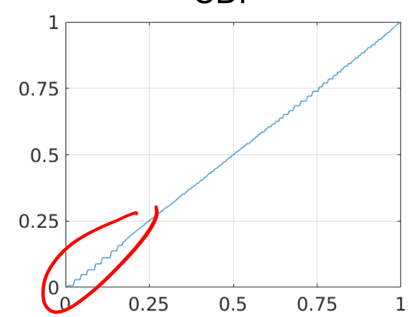
CDF



…after histogram equalization



CDF



almost linear
=> ok

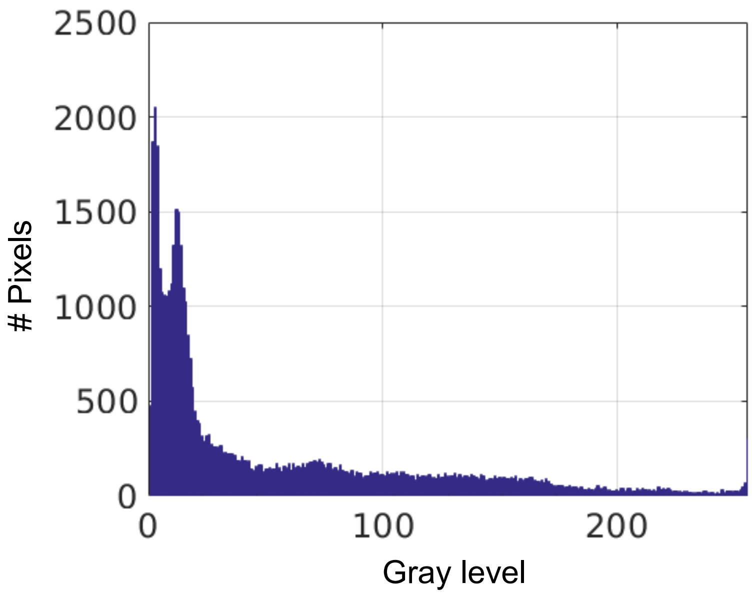
Histogram Equalization Example

Original image

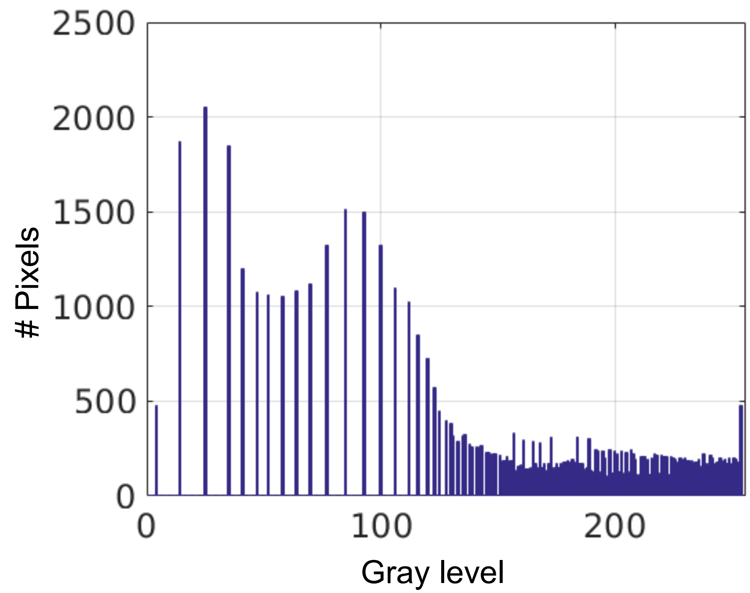
…after histogram equalization

Histogram Equalization Example

Original image



…after histogram equalization



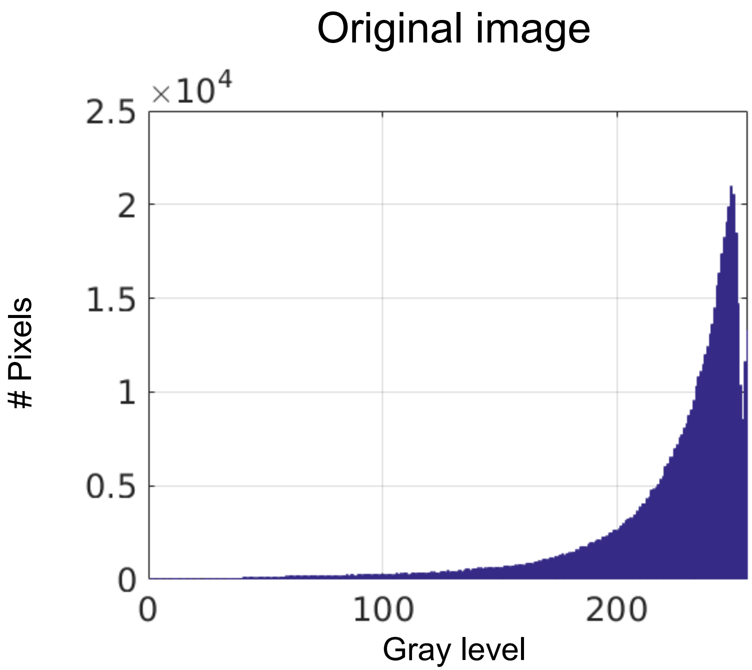
Histogram Equalization Example

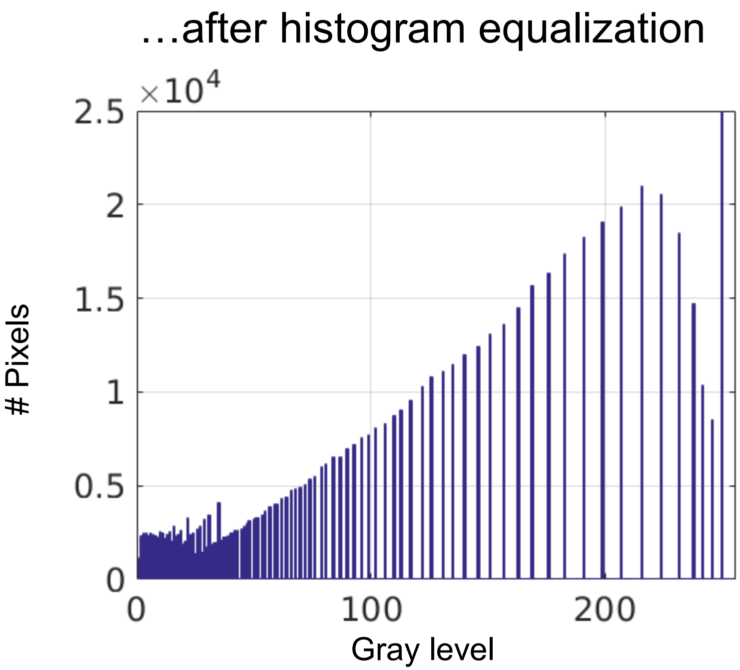
Original image

…after histogram equalization

https://photographylife.com/underexposure-and-overexposure-in-photography

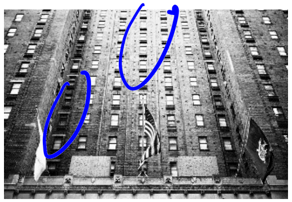
Histogram Equalization Example





Problem:

- Global mapping leads to unnatural local equalization results

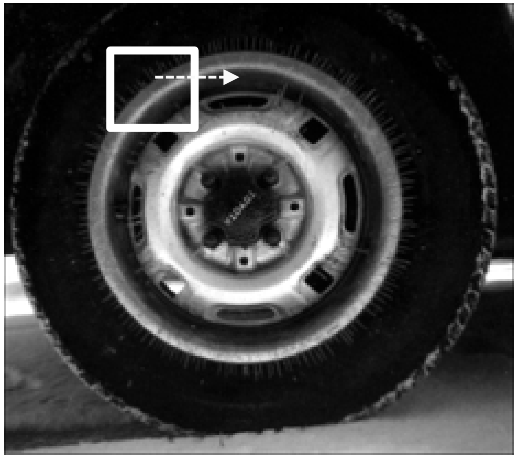


Adaptive Histogram Equalization (AHE)

How to avoid local artifacts because of global equalization?

Apply histogram equalization **locally**

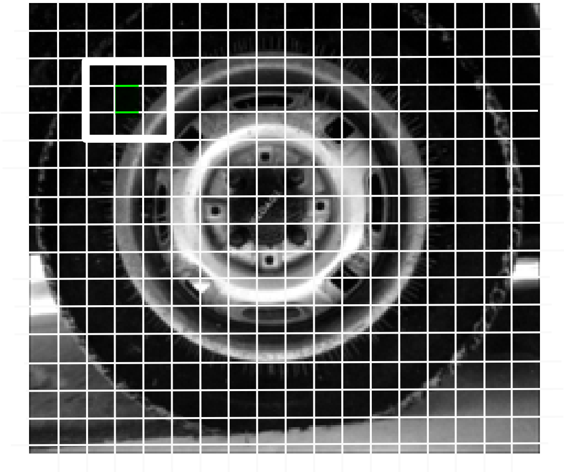
- Computes several histograms (on image sections)
- Improves local contrast



Sliding window approach

Different histogram (and mapping) for every pixel

(time consuming)



Tiling approach

Subdivide into overlapping regions, mitigate blocking effect by smooth blending between neighboring tiles

Adaptive Histogram Equalization Example

Original image

Histogram equalization (global)

Adaptive histogram equalization (8x8 tiles)

Adaptive Histogram Equalization Example

Original image

Histogram equalization (global)

Adaptive histogram equalization (8x8 tiles)

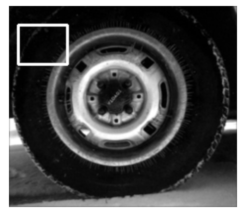
Performance of AHE:

- Good results in high contrast areas
- Noise amplification in case of low local contrast

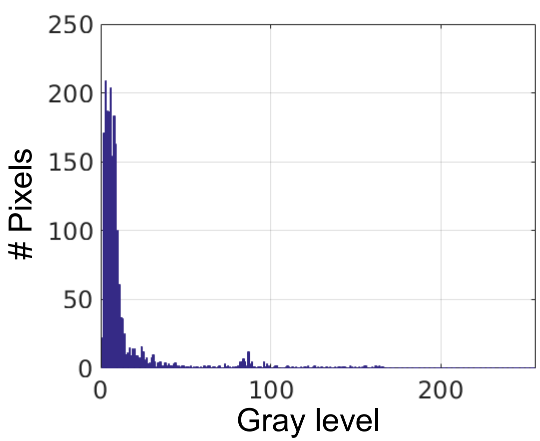
Contrast Limited AHE (CLAHE)

Adaptive histogram equalization

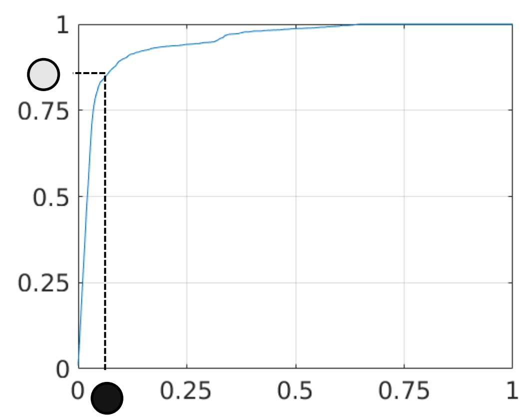
- Tries to make all gray levels equally probable
- **Over-amplifies** noise in relatively homogeneous areas



Local Histogram



CDF

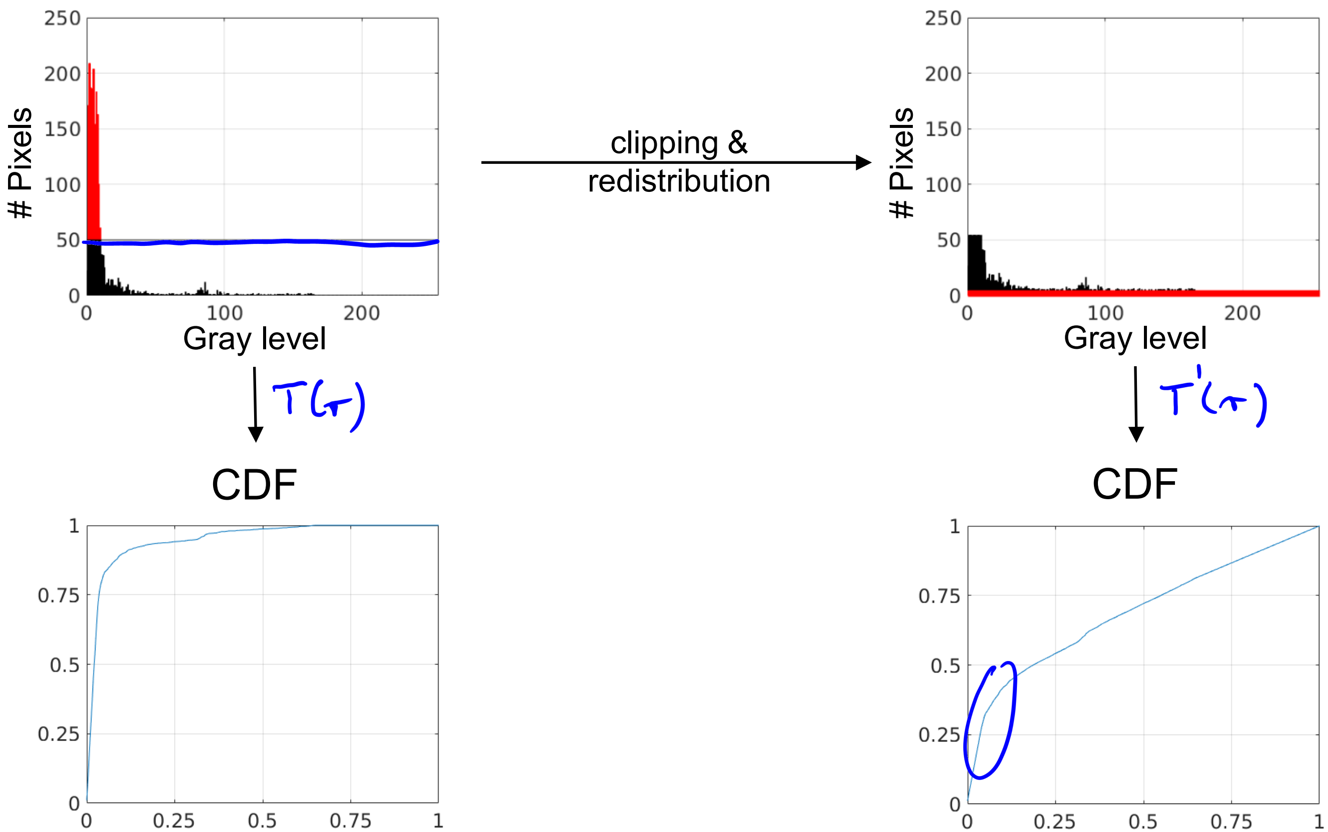


=> black "tire" mapped to white "snow"

Contrast Limited AHE (CLAHE)

Contrast limitation by histogram clipping

- This limits the slope of CDF



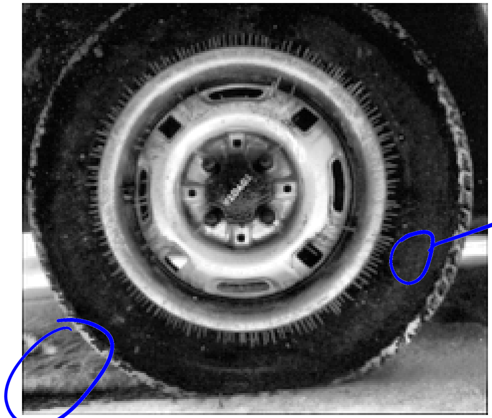
smaller slope
=> better
mapping

Contrast Limited AHE Example

Original image

Adaptive histogram equalization (8x8 tiles)

Histogram equalization (global)



good scaling

no noise amplification

Contrast limited AHE (8x8 tiles)

1.3 γ-adjustment

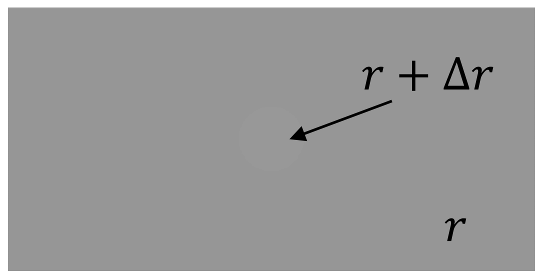
Human eye can adapt to an enormous range of light intensities

- On the order of 10^{10} ($10^{-5} \ldots 10^{6}$ cd/m^2)

candela

But it cannot adapt to all intensities simultaneously

- Brightness adaptation



Can you see the circle?

Weber's Law:

$$\boxed{\frac{\Delta r}{r} \approx \text{const}} \quad \text{(Weber fraction)}$$

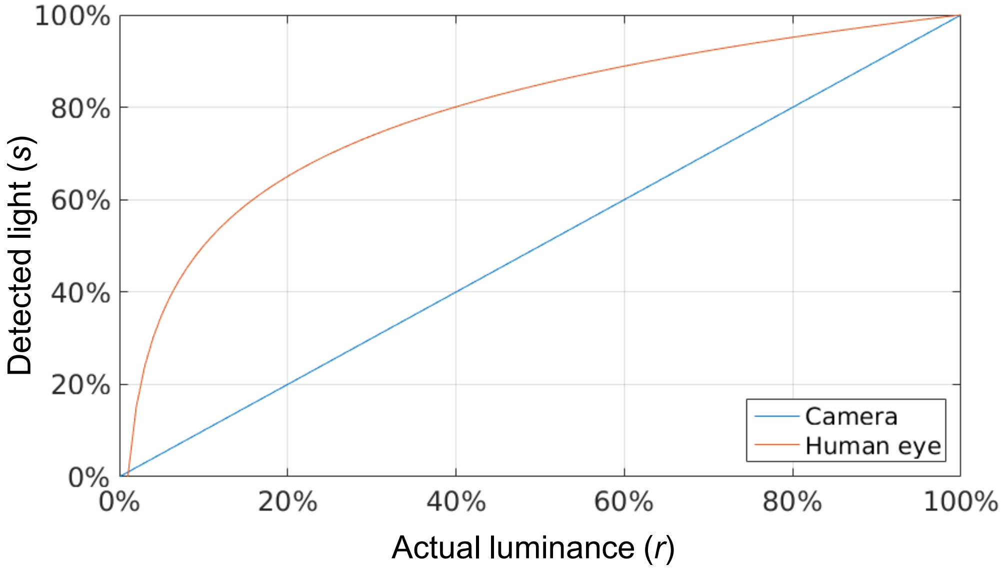
1-2%

also:
haptics: 2-3%
taste: 10-20%

Ernst Heinrich Weber (1795-1878)

γ-adjustment

Our eyes do not perceive light the way cameras do



Gustav Theodor Fechner (1801-1887)

Fechner's Law:

perceived brightness

light intensity

$$\boxed{s \sim \log(r)}$$

Digital cameras:

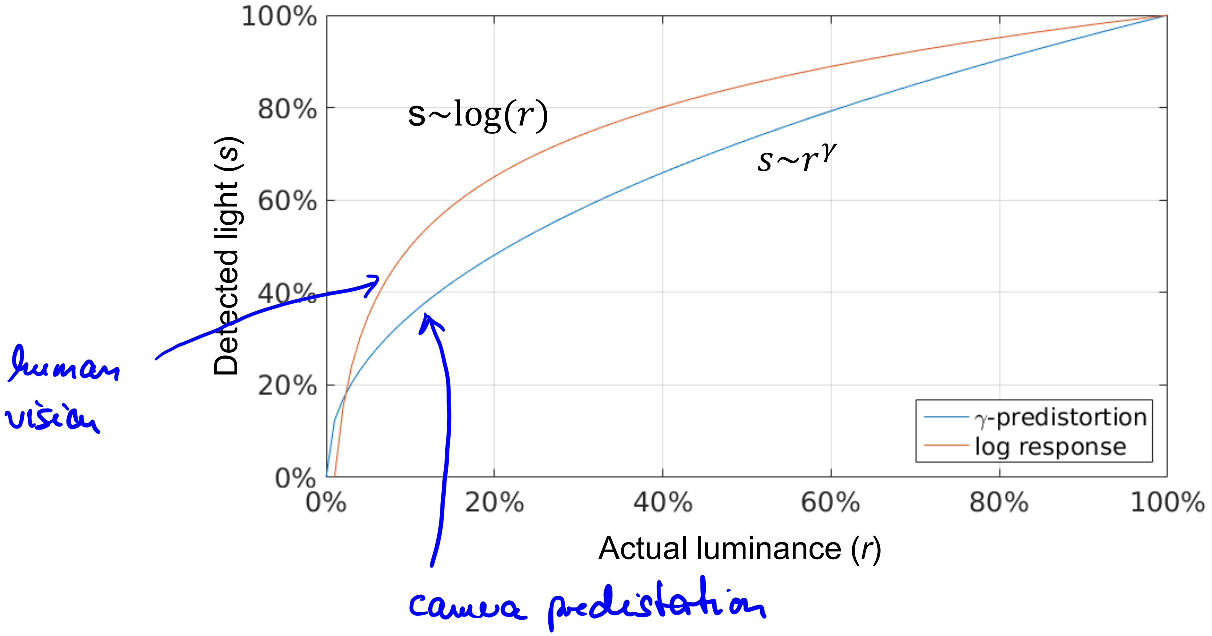
- Twice as much luminance → twice as much detected light

Human eye:

- Much more sensitive to changes in dark tones
- Twice as much light ≠ twice as bright

γ-adjustment

Cameras contain **γ-predistortion** circuit to mimic human eye behavior



Similar enough for most practical applications

Typically, $\gamma = {}^{1}/_{2.2}$

γ-adjustment

Without γ-predistortion

- Excess of bits for brighter tones (camera more sensitive)
- Shortage of bits for darker tones (human eye more sensitive)

Original

Linearly encoded (no γ-predistortion) with 5 bits

visible distortion in dark areas

γ-encoded with 5 bits

equal distribution over intensity

γ-encoded images store tones more **efficiently**

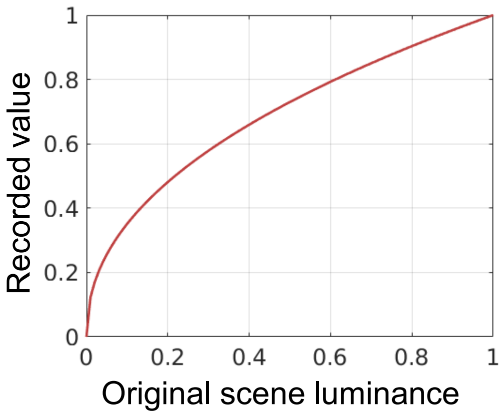
- Equal perceptual spacing

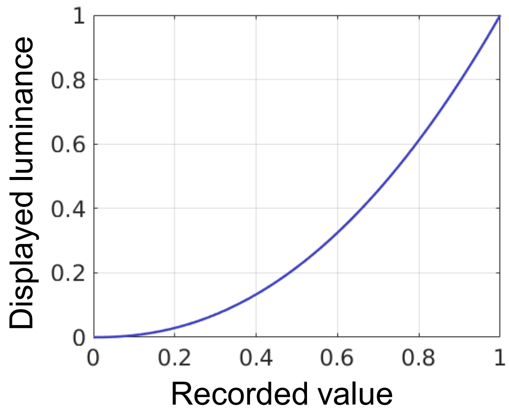
γ-adjustment

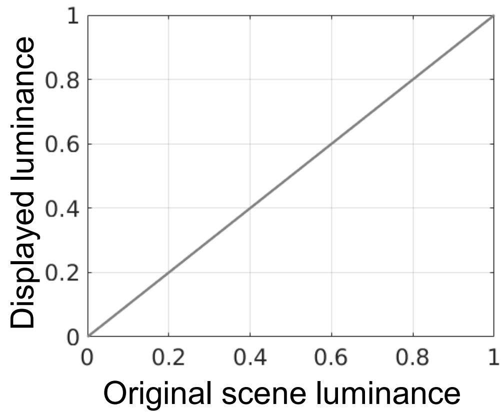
γ-predistortion must be **compensated** before viewing

- Otherwise the image would be unrealistically brightened
- Performed by video card and/or display device

Gamma workflow:



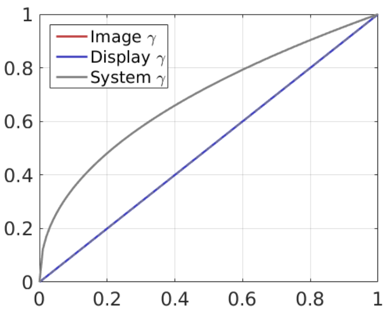




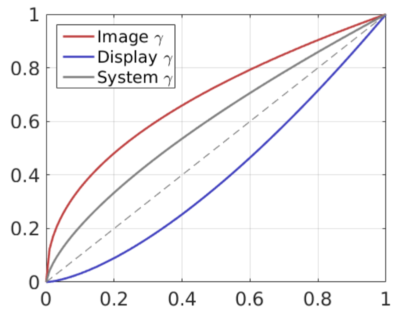
1. Image file γ (γ-predistortion) → 2. Display γ (γ-compensation) → 3. System γ (Net effect)

γ-adjustment

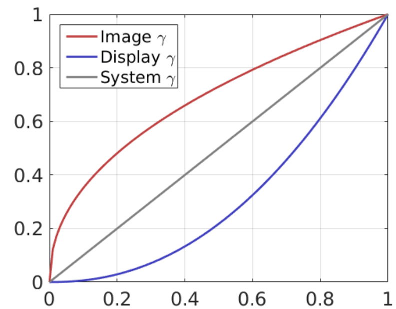
no display comp.
=> too bright



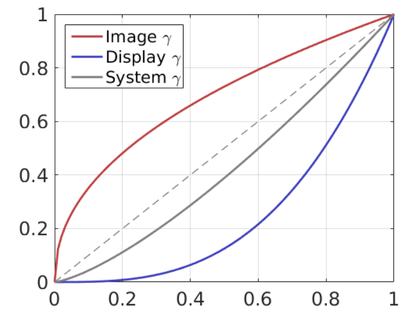
display γ-comp.
too small



display γ-comp.
inverts camera γ-pred.



display γ-comp. too big
=> too dark



Multi-frame Point Operations

Histogram and γ-correction

- Single image point operations

Multi-frame point operations

- Image averaging
- High Dynamic Range (HDR) imaging
- Image subtraction (change/anomaly detection)

Important: accurate alignment

1.4 Image Averaging

Temporal image averaging: used for noise reduction

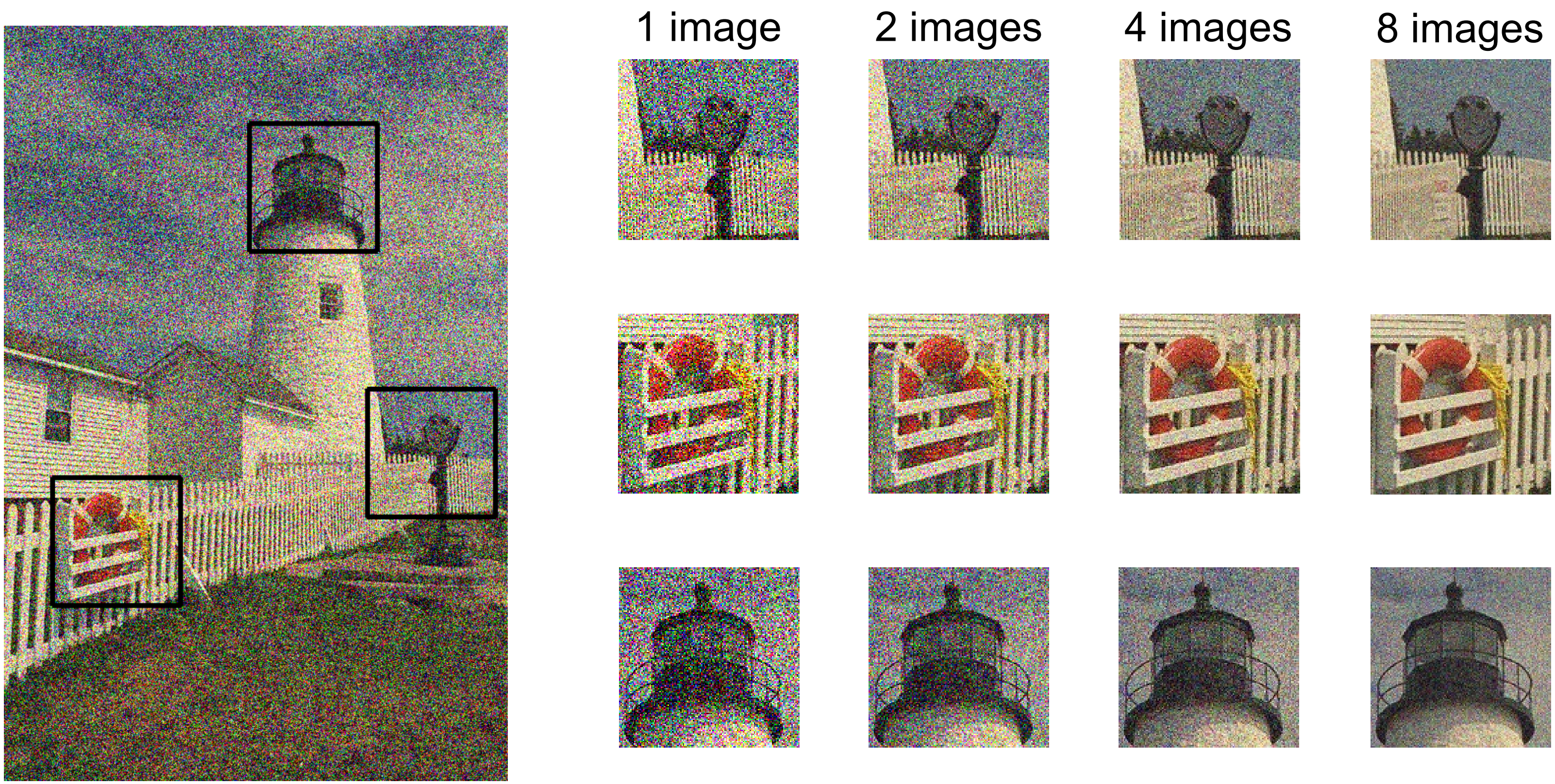


Image Averaging

Idea: take N noisy images $s_1[m,n], s_2[m,n], \dots, s_N[m,n]$ of the same scene

noisy measurement original noise

Error-free image $g[m,n]$: $s_i[m,n] = g[m,n] + v_i[m,n]$ (v_i - additive noise)

Average image: $$\boxed{\bar{s}[m,n] = \frac{1}{N}\sum_{i=1}^{N} s_i[m,n]}$$

Noise power reduction by averaging: How much noise does average image have?

$$\underbrace{E\{(\bar{s}-g)^2\}}_{\text{noise power of average image}} = E\left\{\left(\left(\frac{1}{N}\sum_i s_i\right) - g\right)^2\right\} = E\left\{\left(\left(\frac{1}{N}\sum_i (g+v_i)\right) - g\right)^2\right\} =$$

$$= E\left\{\left(\frac{1}{N}\sum_i v_i\right)^2\right\} = \frac{1}{N^2}\sum_i E\{v_i^2\} = \frac{1}{N}\overbrace{E\{v^2\}}^{\text{noise power}}$$

Provided uncorrelated noise

same noise power $\forall i$

=> noise power is reduced by 1/N

1.5 High Dynamic Range (HDR) Imaging

Broadens the tonal range

- Decreased contrast in some tones (“no free lunch”)

Cameras can actually capture a vast dynamic range

- **But:** not in a single photo (saturation problem)

Idea: create images composed of multiple exposures

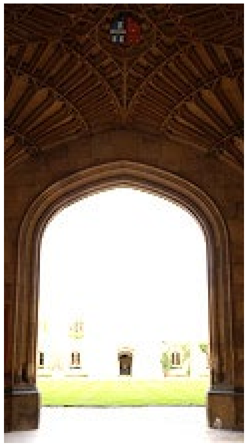
How? by varying shutter speed

- Varying aperture is not an option (it affects depth of field)

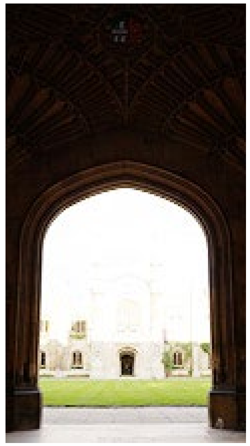
Exposure value differences (known as **stops**)

- Refers to doubling (+1 stop) or halving (-1 stop) of the captured light

High Dynamic Range (HDR) Imaging

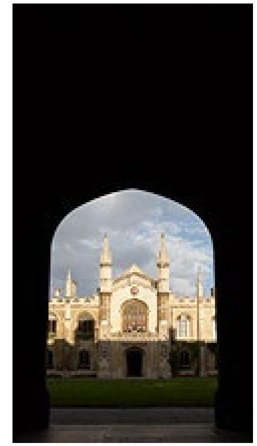


Reference (over-exposure)

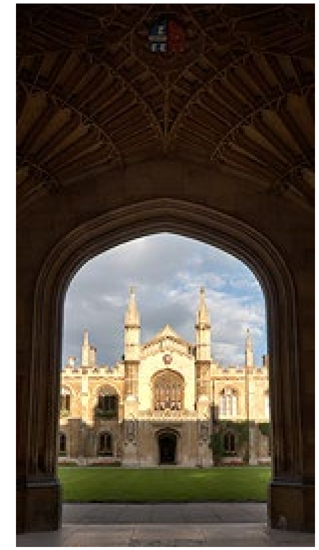


-1 stop

-2 stops



-3 stops (under-exposure)



Combined image

Creating HDR images consists of 2 steps:

1. Recovering a radiance map from a collection of images
2. Converting the radiance map into a display image

http://www.cambridgeincolour.com/tutorials/high-dynamic-range.htm

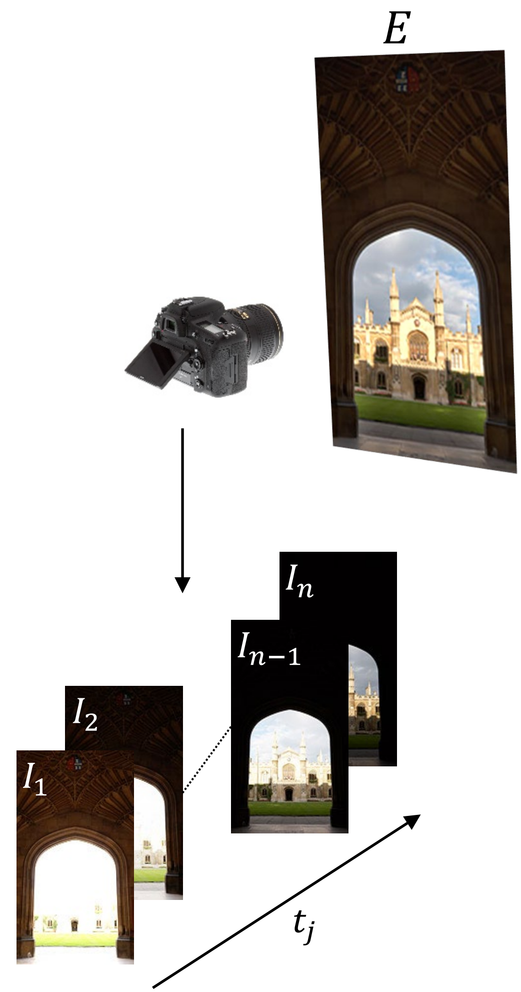
Radiance Map Recovery

Reciprocity equation

- Given exposure time t_j, it maps scene luminance (radiance) r to pixel value s_j

$$\boxed{s_j[m,n] = f\left(r[m,n]t_j\right)}$$

- Exposure value: $r[m,n]t_j$



Camera response function f

- Assumed to be monotonically increasing (therefore invertible)

How can we estimate f?

http://www.imaging-resource.com/PRODS/nikon-d750/ZPR-NIKON-D750-RR-LCD.JPG

Radiance Map Recovery

Response function estimation

- Rearrange reciprocity equation…

$$s_j[m,n] = f\left(r[m,n]t_j\right)$$

$$\downarrow$$

$$f^{-1}\left(s_j[m,n]\right) = r[m,n]t_j$$

- …and take logarithms

$$\underbrace{g\left(s_j[m,n]\right)}_{\text{Unknown function}} = \log\left(f^{-1}\left(\underbrace{s_j[m,n]}_{\text{known pixel intensities}}\right)\right) = \log\left(\underbrace{r_j[m,n]}_{\text{Unknown radiance}}\right) + \log\left(\underbrace{t_j}_{\text{known exposure time}}\right)$$

Radiance Map Recovery

Solve for $g = \log(f^{-1})$ and r in a least-squared error sense

- Quadratic function → easy to solve

$$\underset{g,\,E[m,n]}{\text{minimize}}\left\{\underbrace{\sum_{(m,n)}\sum_{j}\left(w\left(s_j[m,n]\right)\left[g\left(s_j[m,n]\right)-\log(r[m,n])-\log\left(t_j\right)\right]\right)^2}_{\text{squared error}}+\lambda\underbrace{\sum_{k=1}^{2^B-1}\left(w(k)g''(z)\right)^2}_{\text{smoothness term}}\right\}$$

Smoothness term

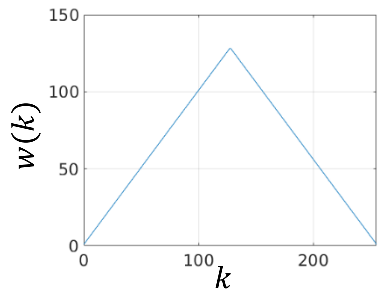
- Discrete approximation of 2nd derivative:

$$g''(k) = g(k-1) - 2g(k) + g(k+1)$$

- Balance factor λ chosen according to the amount of noise expected in s_j

Weighting function w

- suppresses the effect of pixel saturation



Radiance Map Recovery

If $(g, \log r)$ minimize the error then $(g, \log r) + \alpha$ do so as well

- Scale factor $\alpha \in \mathbb{R}$
- To establish α, we force the central value to have unit exposure, i.e.

$$g\left(\frac{2^B - 1}{2}\right) = \log(1) = 0$$

Radiance map:

$$r[m,n] = \exp\left(\frac{\sum_j w\big(s_j[m,n]\big)\big(g\big(s_j[m,n]\big) - \log(t_j)\big)}{\sum_j w\big(s_j[m,n]\big)}\right)$$

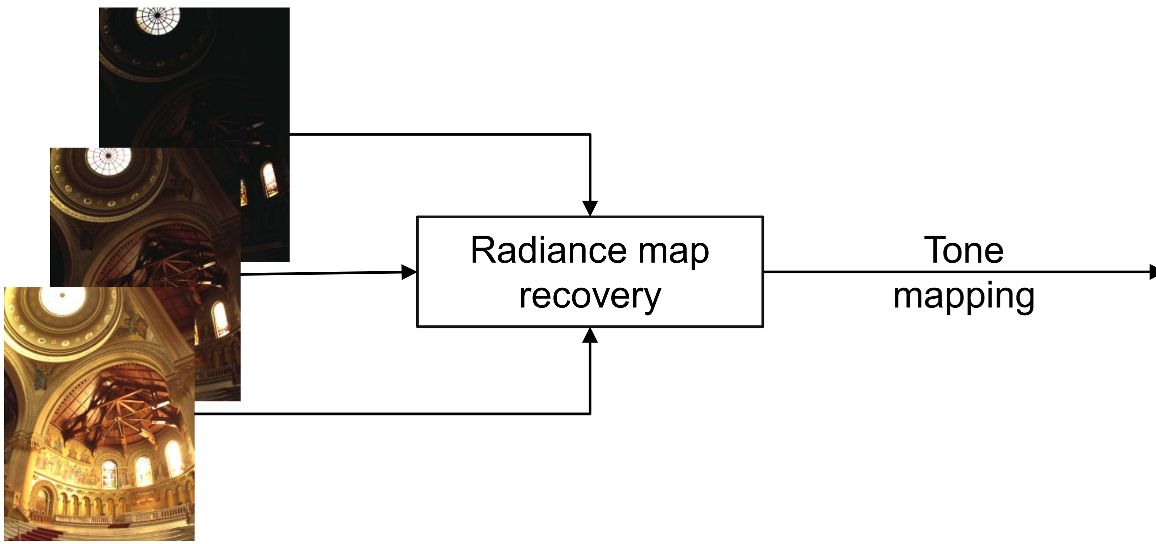
Combining multiple exposures reduces noise in the radiance map

Tone Mapping

Radiance maps correspond to HDR images

- Dynamic ranges of $1:10^5$ and more
- Cannot be displayed on normal screens ($\approx 1:10^2$)

Solution: tone mapping from HDR into [0,1]



Linear mapping

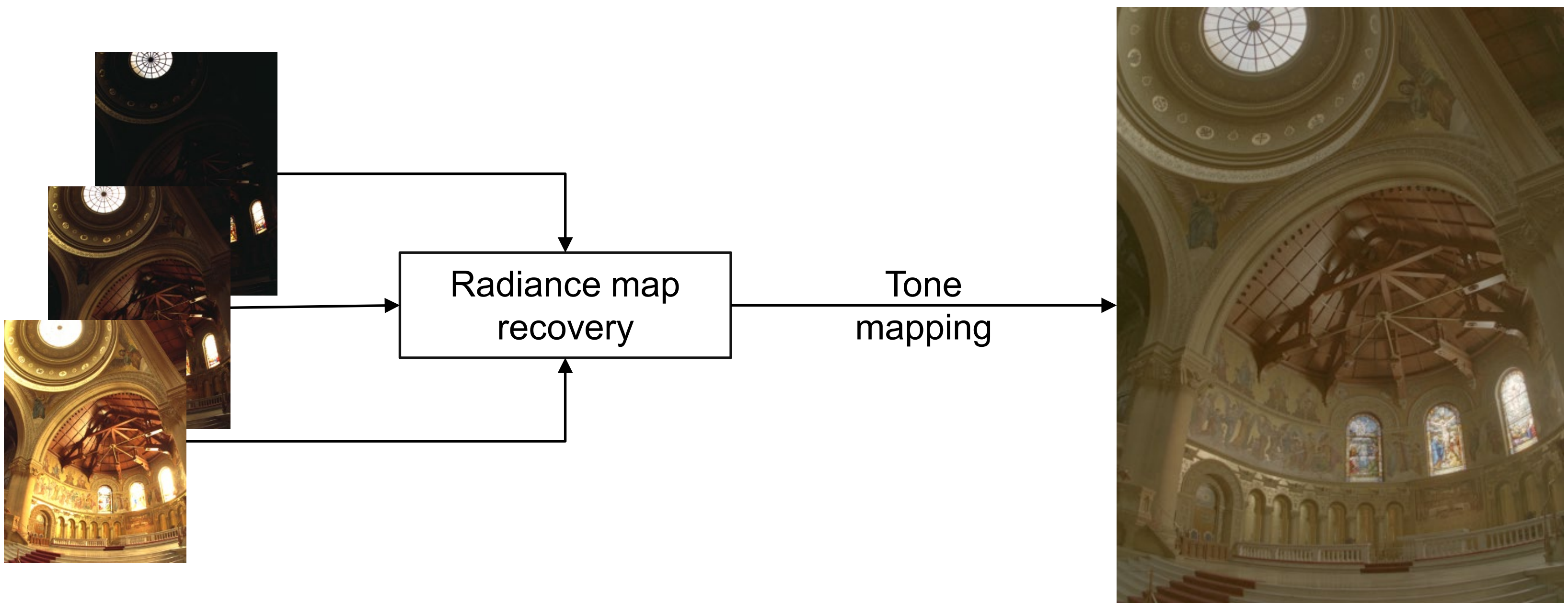
Tone Mapping

Radiance maps correspond to HDR images

- Dynamic ranges of $1:10^5$ and more
- Cannot be displayed on normal screens ($\approx 1:10^2$)

Solution: tone mapping from HDR into [0,1]

Logarithmic mapping



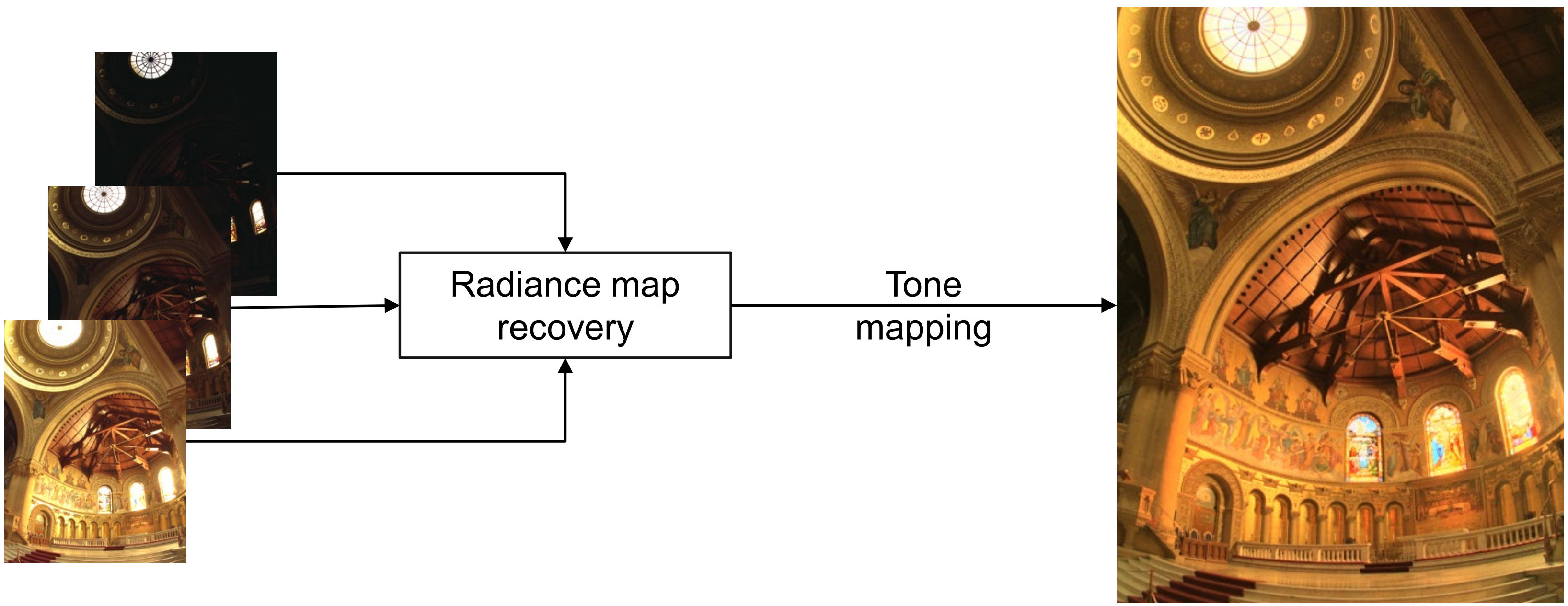
Tone Mapping

Radiance maps correspond to HDR images

- Dynamic ranges of $1:10^5$ and more
- Cannot be displayed on normal screens ($\approx 1:10^2$)

Solution: tone mapping from HDR into [0,1]

Global Reinhard's mapping



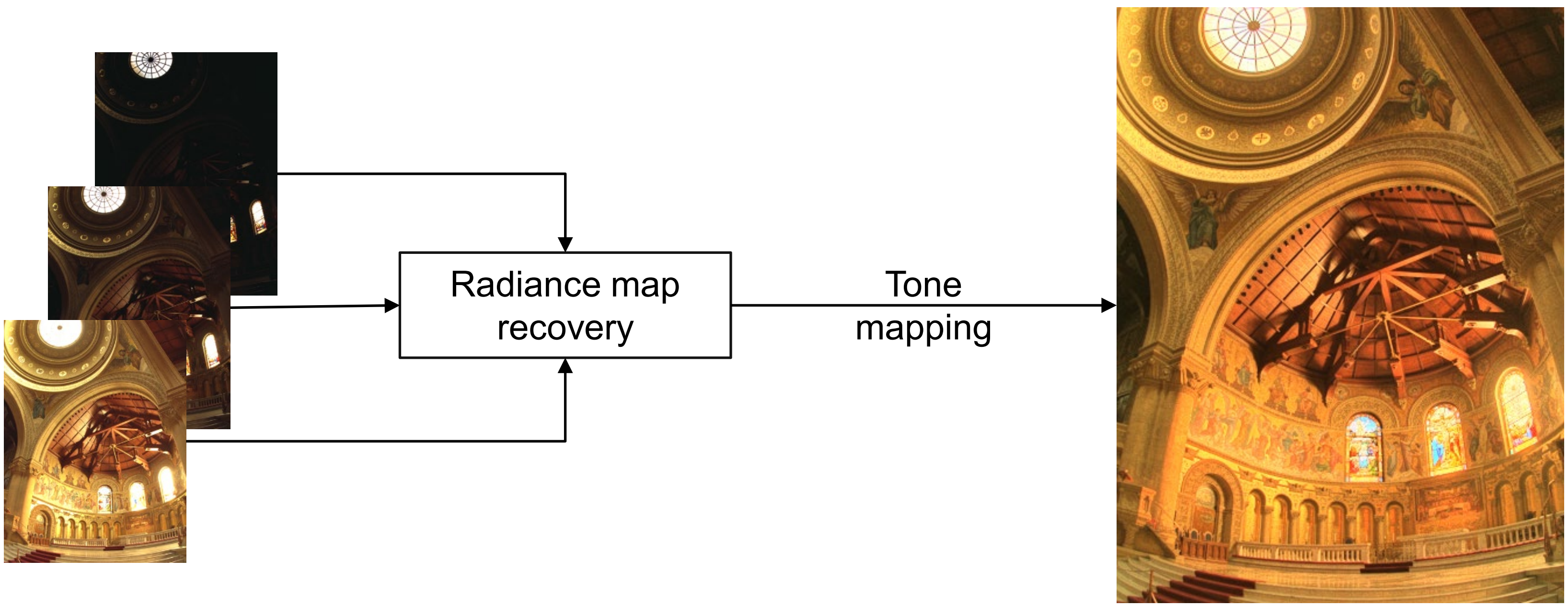
Tone Mapping

Radiance maps correspond to HDR images

- Dynamic ranges of $1:10^5$ and more
- Cannot be displayed on normal screens ($\approx 1:10^2$)

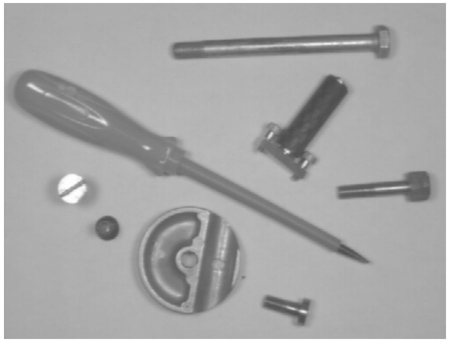
Solution: tone mapping from HDR into [0,1]

Local Reinhard's mapping

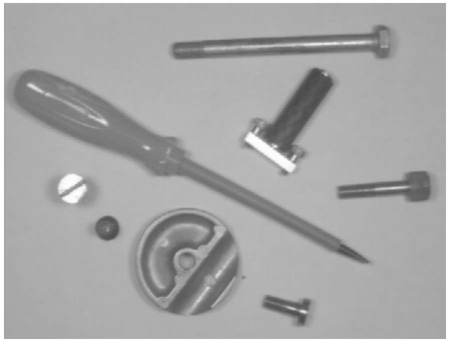


1.6 Image Subtraction

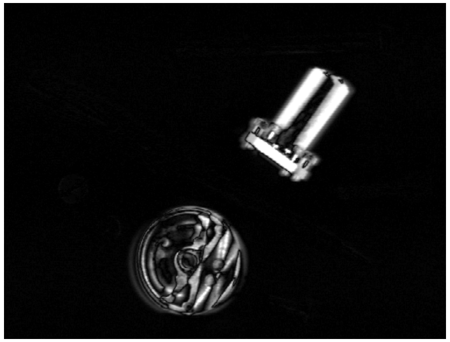
Find differences (errors/anomalies) between 2 mostly identical images



\-

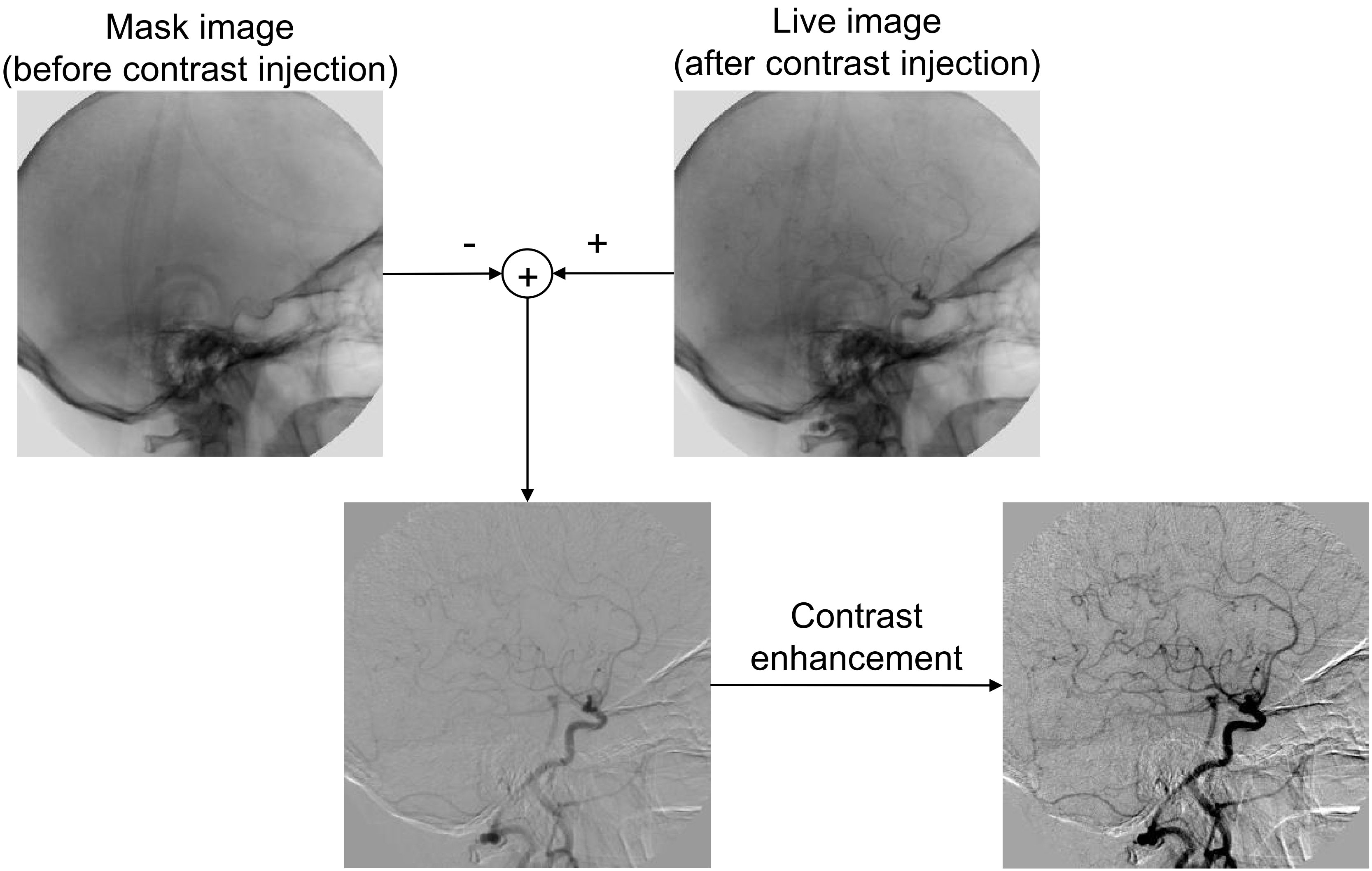


=



http://homepages.inf.ed.ac.uk/rbf/HIPR2/pixsub.htm

Digital Subtraction Angiography



http://www.isi.uu.nl/Gallery/DSA/