Computer Vision

Module M2:
Image representations

Krzysztof Krawiec
http://www.cs.put.poznan.pl/kkrawiec/

Poznan University of Technology, 2021-2023

Digital images and raster images

Digital image

A data structure reflecting (in a certain way) the content of a physical scene.

Types of representations:

- Dense: Assumes uniform sampling of the scene.
  Example: Raster image
- Sparse: Maintains information only about selected points in the scene.
  Examples:
  - vector graphics, e.g. Scalable Vector Graphics (SVG)
  - point clouds acquired with a rangefinder (laser or ultrasound-based).

Most of research and applications revolve around raster images.

- So does this course.

Raster image

A data structure storing the result of scene scanning in *n* spatial dimensions.

- n = 2: planar image
  - 2-dimensional array, *matrix*
  - Element of the domain: pixel, pel = picture element, image point
  - Width = # of columns, X, height = # of rows, Y
  - Get used to different conventions: Cartesian (X,Y) and algebraic (Y,X)
- n > 2: volumetric image (usually)
  - n-dim array, *tensor*
  - (X,Y,Z): Volumetric image.
    - Element of the domain: voxel = volume element,
  - (X,Y,t): Video sequence
  - Other, e.g. (X,Y,Z,t)
    - E.g. in functional Magnetic Resonance Imaging (fMRI)

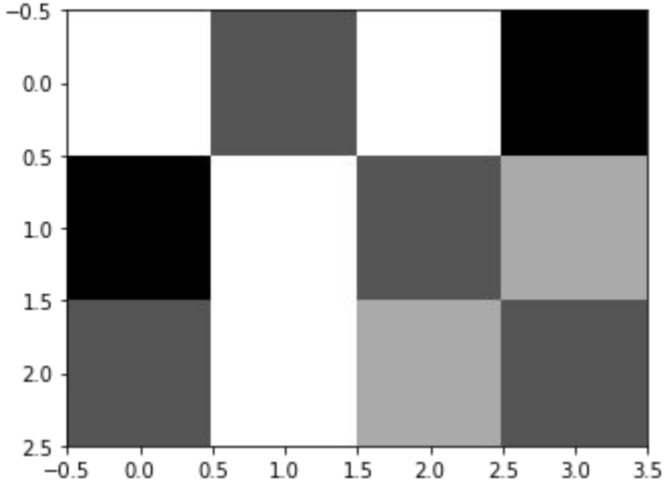
3	1	3	0
0	3	1	2
1	3	2	1

Raster image

Pixel ≠ point.

Pixels/voxels have finite dimensions; mathematical points are 0-dimensional.

- Implication: a different definition of neighbourhood.
- Misinterpretations possible.
- Q: What are the coordinates of the upper-left corner in this raster?



Raster image as a function

A mapping from space to a value set:

$$f : \mathbb{N}^n \to \mathbb{V}$$

a.k.a. *characteristic function* of an image.

Motivations:

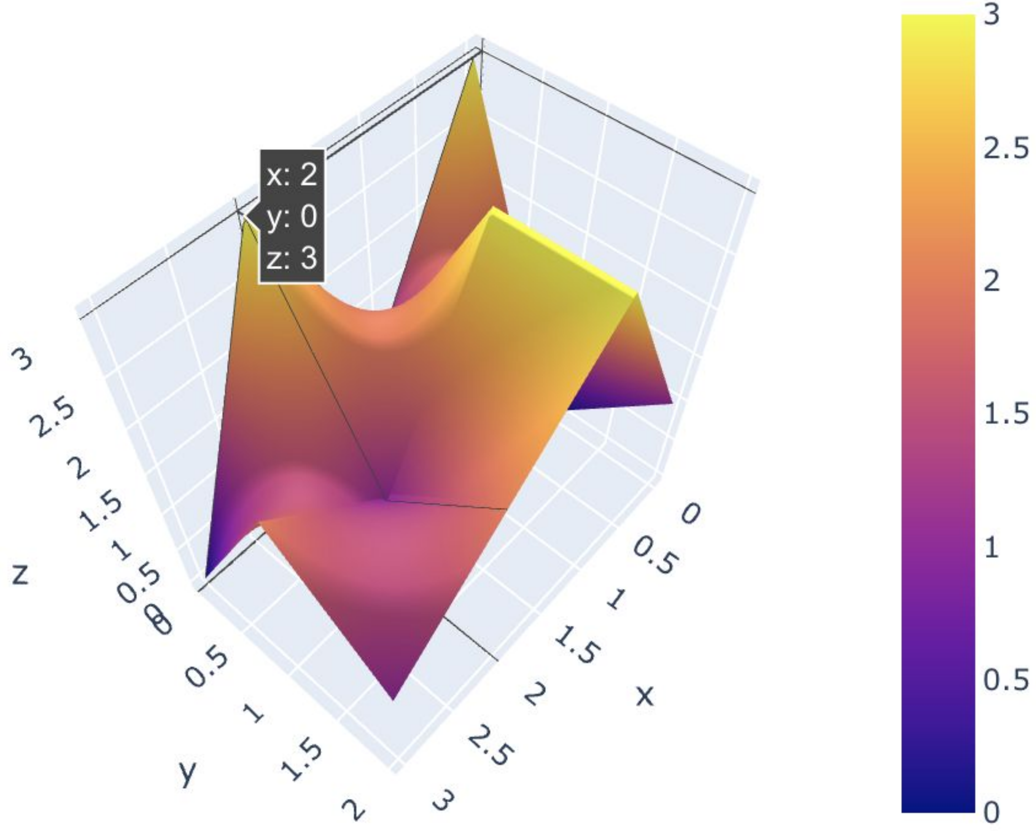
- Functions are more generic/universal than values (e.g. matrices).
- Each value can be represented as a function – but not the other way round.

Widely adopted convention:

- The X axis oriented to the right.
- The Y axis oriented downwards.

Image as a function

Rendering of the image from the previous slide as a 3D relief (the z axis is the value axis).



Raster image: the codomain

The codomain $\mathbb{V}$ can hold arbitrary values. In practice, the most common codomains are:

- $\mathbb{V} = \mathbb{R}^m$
- $\mathbb{V} = \mathbb{N}^m$

Where m is the number of *channels* (*image depth**):

- m = 1: scalar:
  - binary codomain: black-and-white (B&W) image, bitmap
  - multi-valued codomain: monochrome image (greyscale).
- m > 1: vector:
  - multi-channel image; values often interpreted in terms of colors (e.g. RGB, CMY, HSI),
  - multimodal images (e.g. satellite imaging, MRI),

*Not to be confused in depth in 3D CV and stereovision.

Raster image: the codomain

Technical implementation: the components of $\mathbb{V}$ are usually values of type:

- byte [0, 255]
- float [0.0,1.0]

In absence of counterarguments (e.g. computational performance), we should prefer float due to greater precision (256 levels often insufficient).

Other way of characterizing $\mathbb{V}$: the number of *bits per pixel* (bpp).
Common values:

- 1 (bitmap),
- 8, 12, 14 (monochrome images),
- 24 (color images),
- 32 (color + transparency/alpha channel).

On the discrete nature of raster images

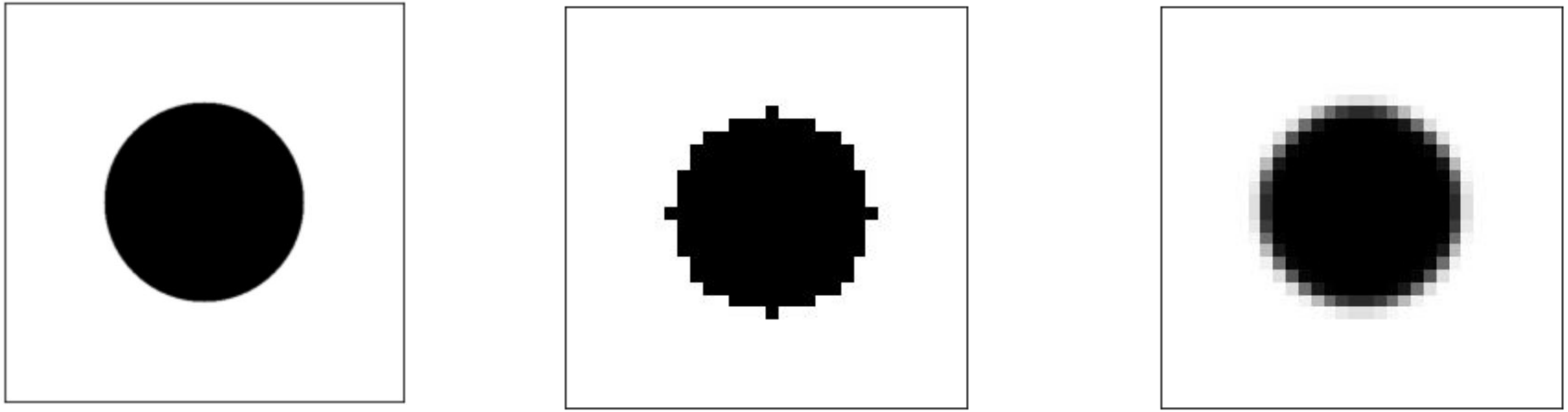
Spatial discretization (domain):

- Results from the limited resolution of the sensor, expressed e.g. in
  - points per mm or per inch (dots-per-inch, DPI),
  - points per degree angle (angular resolution),

Signal value discretization (codomain): may arise from:

- Finite precision of analog-to-digital converters
- Adopted/assumed precision for $\mathbb{V}$
- The floating point representation of $\mathbb{V}$
- In some applications: the discrete nature of the original signal
  - E.g. particle physics

Example: Effects of spatial discretization



Color representation spaces

The concept of color

Color image is a special case of multi-channel image.

Color is a subjective percept, which belongs to the category of *qualia* (singular: *quale*).

- Results from the specific interpretation of wavelength that takes place in visual perception and the central nervous system of primates.
- A famous voice in the discussion of qualia: *What Is It Like to Be a Bat*? [1]

Nevertheless: color does matter in CV, due to its importance for humans in many application areas.

[1] Thomas Nagel, What Is It Like to Be a Bat? The Philosophical Review LXXXIII, 4 (October 1974):435-50, http://www.philosopher.eu/others-writings/nagel-what-is-it-like-to-be-a-bat/

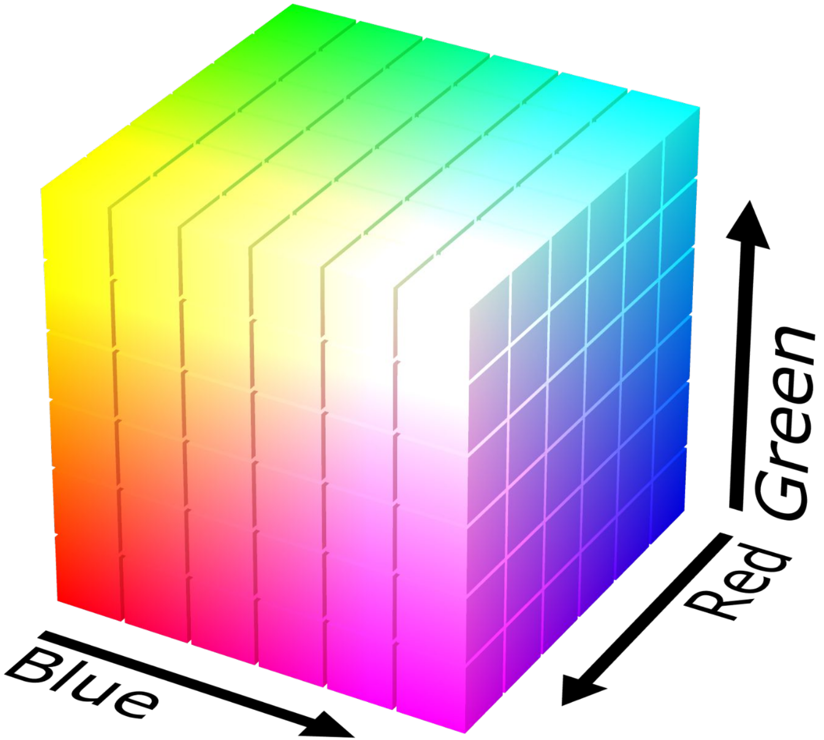
Color representation spaces

- RGB
  - sRGB
- CMY
- CMYK
- HS*
  - HSI
  - HSV
  - HSB
- …

The complete list is much longer.

- Many `targeted’ color representation spaces have been proposed in industry and application areas (see colorimetry)

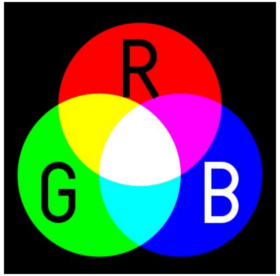
The RGB/CMY cube

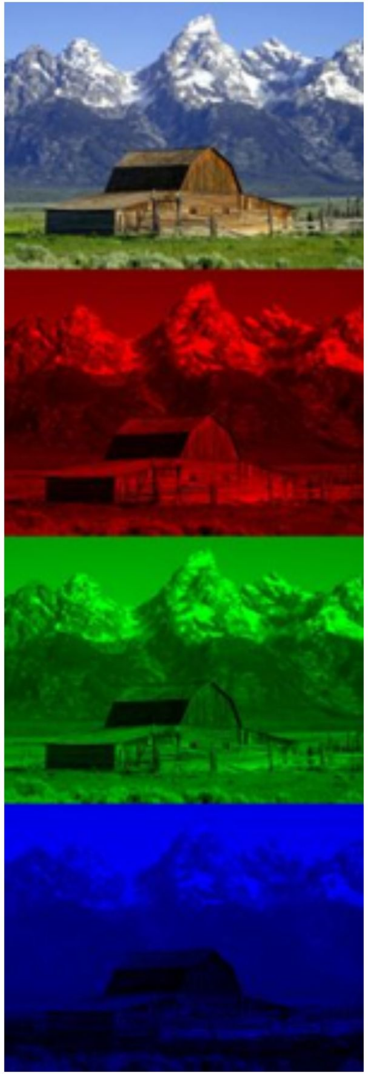


https://commons.wikimedia.org/wiki/File:RGB_color_solid_cube.png

RGB: Red, Green, Blue

Additive model. Common in <u>projective</u> devices (e.g. LED screens, beamers):
Color originates as the result of emission/projection of light.





One of the first color photographs

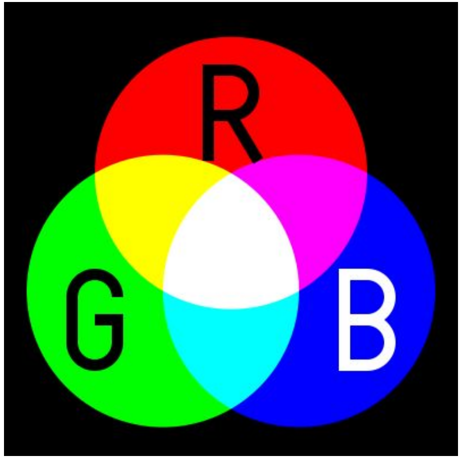
B, G, R

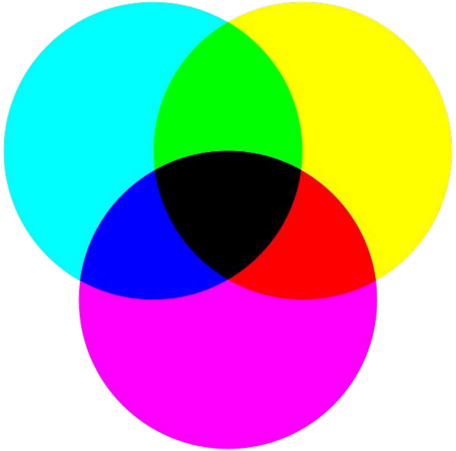
Sergey Prokudin-Gorsky, 1911

CMY: Cyan, Magenta, Yellow

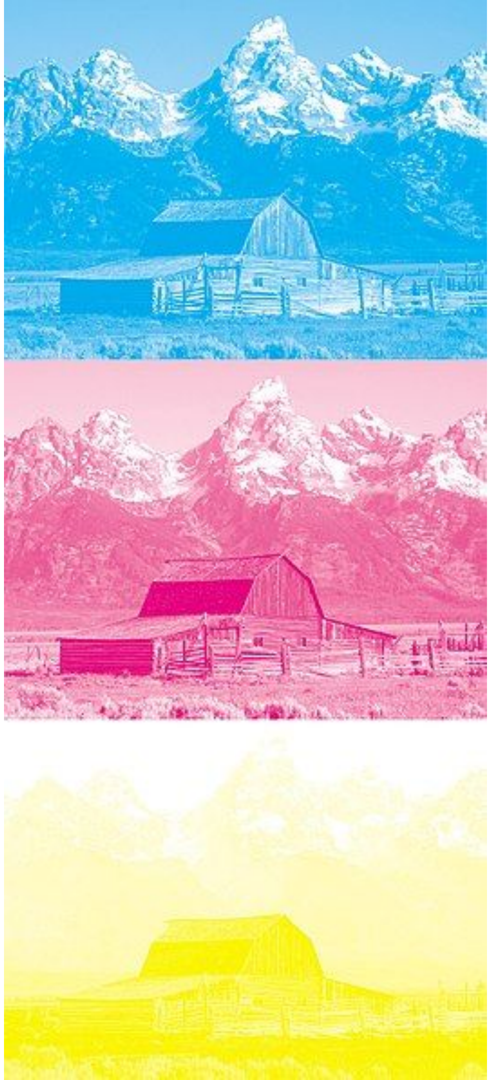
A ‘complement’ of the RGB space. Subtractive model.

- Used mostly in printing. Color originates in reflection of light from surfaces, with some parts of the spectrum being absorbed by the surface.
- C, M and Y are complements of R, G, B with respect to some maximal value (e.g. 1):
  - C = 1-R
  - M = 1-G
  - Y = 1-B





CMY: Example



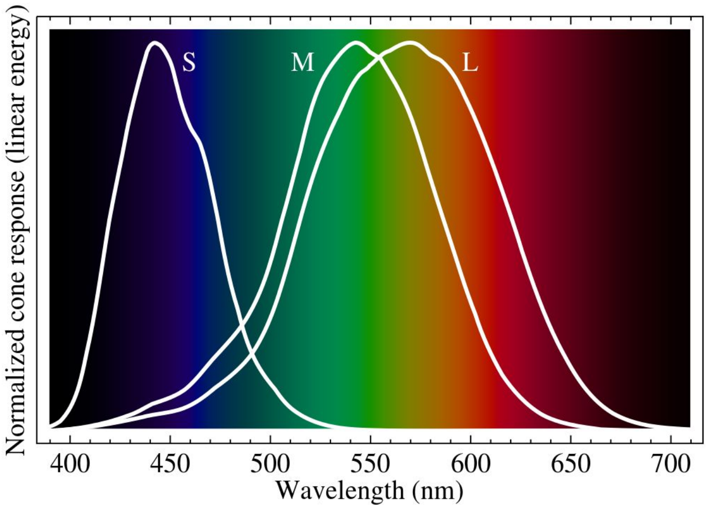
CMYK space

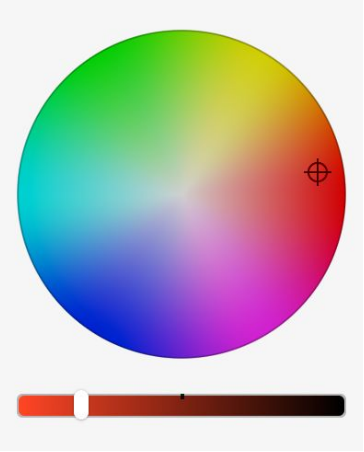
Extends the CMY space with the *key* channel (black)

- Leads to significant savings in printing materials (toner, paint, etc.) compared to CMY.

HSI: Hue, Saturation, Intensity

- Motivation: better alignment with human perception (compared to RGB and CMY)
    - Humans don't reason about color in terms of cone cells.
- Useful in CV,
    - though originated in computer graphics: color pickers
- Related spaces: HSV, HSL, HSB





HS* spaces: HSL, HSV [HSB]

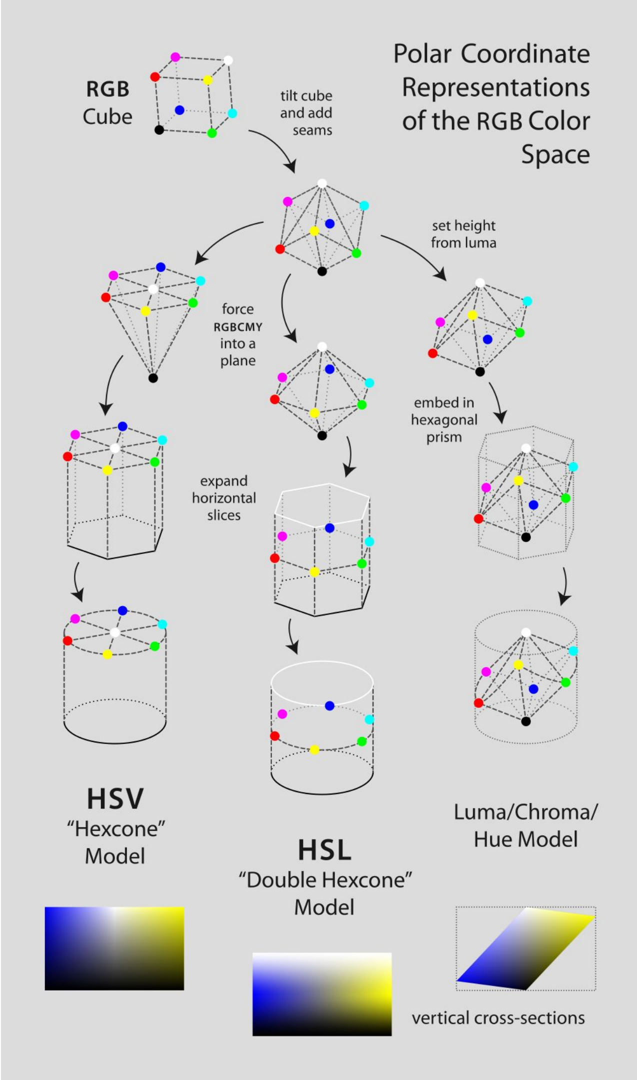
- Common objective: To ‘morph’ the RGB cube in a way that is perceptually natural for humans.
- Common features: delineation and orthogonalization of
  - Hue
  - Saturation
  - ‘Value’, which can be defined as:
    - Luminance: <u>a physical (photometric) quantity</u>, amount of light passing a unit of surface/area (measured in candela per square meter)
    - Lightness: <u>relative quantity</u> (multiple definitions in color appearance models)
    - Brightness: <u>absolute observed value,</u> however subjective per observer
      - Usually defined as the arithmetic mean of R, G and B.
- Most common:
  - Hue, Saturation, Value
  - Hue, Saturation, Lightness

Relations between HS* spaces

Relations between RGB and HS*

Notice:

- H is defined on a circular (modulo) scale.
- In most cases, RGB and HS* are stored as integers,
  - Transformations RGB ↔ HS* are irreversible.
- General remark: color transforms
  - are local (pixel-wise)
  - do not augment/enrich the available information.



https://commons.wikimedia.org/wiki/File:Hsl-and-hsv.svg

RGB → HSI

Raster image as a data structure

Two-dimensional array:

- rows, columns

Technical remarks:

- Preferred iteration scheme: rows, columns for consistency with memory organization (minimizes the likelihood of cache misses),
- In multi-channel images, the channels are often interweaved (BGRBGRBGR …)
    - Common operation: channel split
- Some libraries use 32bits/pixel anyway.
- Some libraries allocate extra ‘slack space’ to make sure that image rows start at memory addresses that are ‘convenient’ (e.g. even).

Metadata for raster images

Image metadata

Parameters that describe the data structure (and its origins), rather than its content.

Examples:

- number of channels
- dimensions
- resolution (dots per inch, DPI),
  - Can vary along different dimensions.
- line/row length in bytes (e.g. in the BMP image format)

Example metadata

Colour Model: RGB

Depth: 8

DPI Height: 72

DPI Width: 72

Pixel Height: 4 608

Pixel Width: 3 456

Profile Name: sRGB IEC61966-2.1

EXIF: Exchangeable Image File Format

A standard for description of image formats (also sound data formats) and their annotations ([meta]tags). Used in JPEG, TIFF, RIFF; not used in PNG, JPEG2000, GIF.

Example entries in an EXIF record :

- Aperture Value: 2
- Brightness Value: 0
- Colour Space: sRGB
- Components Configuration: 1, 2, 3, 0
- Date Time Digitised: 18 Apr 2017 at 16:02:44
- Date Time Original: 18 Apr 2017 at 16:02:44
- Exposure Mode: Auto exposure
- Exposure Program: Not defined
- Exposure Time: 1/991
- Flash: Off, did not fire
- FlashPix Version: 1.0
- FNumber: 2
- Focal Length: 3,57
- Focal Length In 35mm Film: 4
- Image Unique ID: 0fa7658f296a735b0000000000000000
- ISO Speed Ratings: 100
- Metering Mode: CentreWeightedAverage
- Pixel X Dimension: 3 456
- Pixel Y Dimension: 4 608
- Scene Capture Type: Standard
- Scene Type: A directly photographed image
- Sensing Method: One-chip colour area sensor
- Shutter Speed Value: 1/990
- White Balance: Auto white balance

File formats

Image compression

The objective: reduction of occupied disk space/memory.

- Lossless compression: perfect restoration possible.
- Lossy compression: quality of compression traded against compression rate.

Motivating example:

- A raster image from an average smartphone (2020):
  4032 * 3024 * 3 = 36,578,304 B = 34.9 MB
  A smartphone with 64GB memory could store only 1900 photos.
- Astronomy, e.g. Pan-STARRS telescope https://panstarrs.stsci.edu/
  - 1.4 mld pixels/image
  - 1 image = ~2GB

Compression vs. file formats

Compression algorithm ≠ file format

- Some compression algorithms are used in multiple file formats.
- Some file formats are ‘containers’ that can hold images compressed with different algorithms.

Some more exotic file formats and compression algorithms:

- TIFF/TIF (Tagged Image File Format) – can use LZW, PackBits (lossless) lub JPEG (lossy)
- JPEG2000: Better than JPG, but higher computational complexity of compression and decompression; used in certain professional settings (e.g. cinematography)
- Other: PCX, PPM, PGM

File formats (using a single compression algorithm)

Lossy:

- JPEG (Joint Photographic Experts Group) – name of a compression method, not of a file format
  - The formal name of the file format is JPEG File Interchange Format (JFIF).
  - Successive compressions and decompressions degrade image quality.

Lossless:

- GIF (Graphics Interchange Format) – 256 colors, multiframe,
  - patented (as it uses LZW), though the patent has already expired in most countries
- PNG (Portable Network Graphics) – the successor of GIF, opensource
- BMP – Windows bitmap
- RAW – more of a metaformat, non-standardized
  - Usually: two integers determining image dimensions, followed by the raw data ‘dump’.

Downsides of lossy compression

While not evident at cursory inspection, can have profound impact on:

- Color characteristic
- Precision of localization and dimensions of small object

Particularly dangerous in the context with ML (and deep learning): overfitting to image compression artefacts (see, e.g., [1]).

Next slides: a satellite image compressed with JPEG and its Saturation channel.

[1] Samuel Dodge, Lina Karam, Understanding How Image Quality Affects Deep Neural Networks, https://arxiv.org/abs/1604.04004

Exemplary image (satellite imaging)

The saturation channel of that image

Macroblocking

End of module