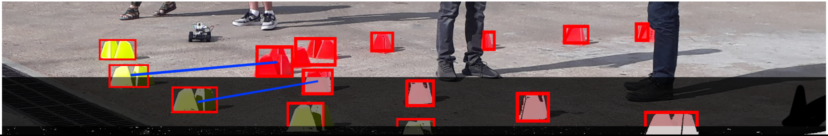
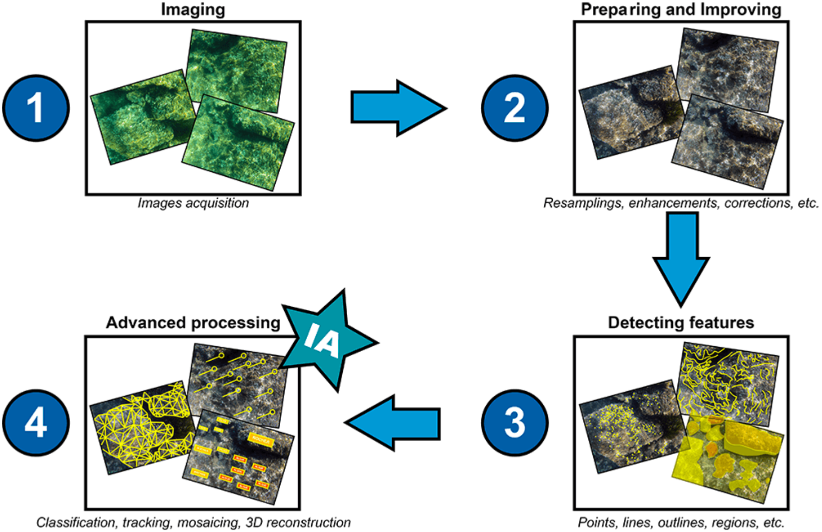
[IPRO] Computer vision – Imaging

Explorers' journey – First Camp

Laurent Beaudoin & Loïca Avanthey



Computer Vision: overview



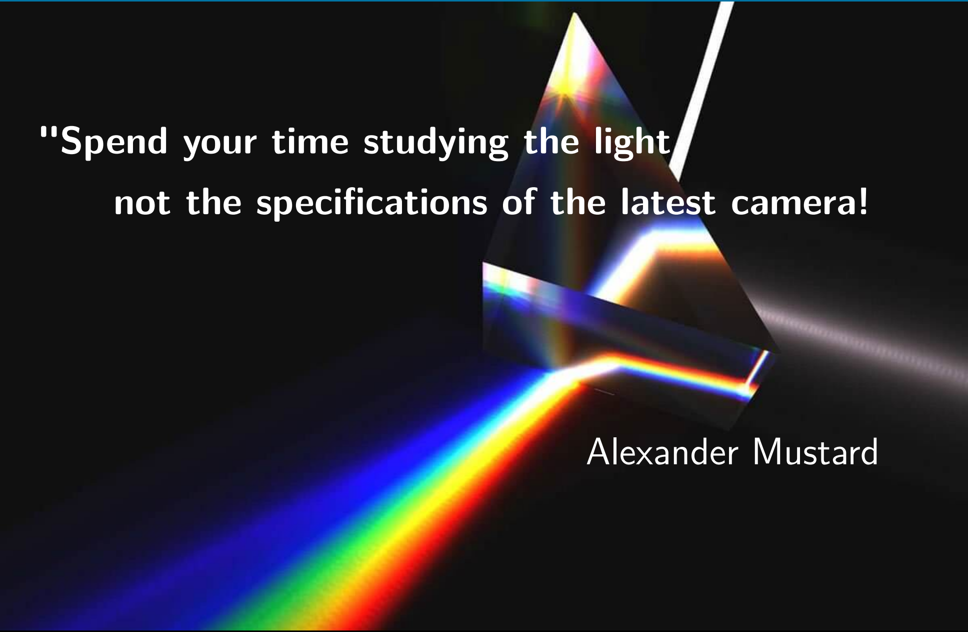
[Avanthey et al., 2018]

Prelude

Computer vision starts with images

Understanding **how images are formed**, or **how the sensors we use work**, is **essential** for developing **effective processing strategies**.

So we will **first** go back to **physics**, **biology** and **technology**, **before** moving on to **algorithms**.



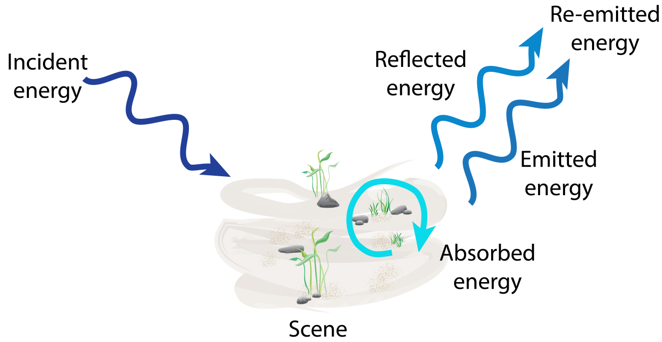
"Spend your time studying the light

not the specifications of the latest camera!

Alexander Mustard

Reminder: energy & objects

The objects of a scene **receive energy**: they **absorb** some of it and **reflect** the rest in addition to the energy they **produce**



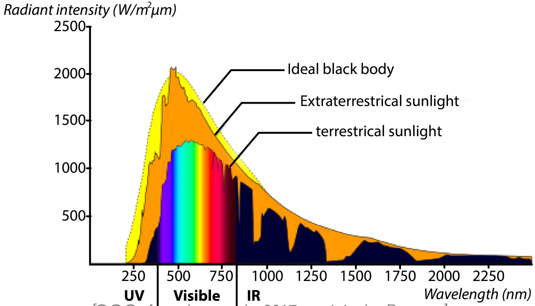
[Avanthey et al., 2018]

Reminder: electromagnetic spectrum

The **higher** the amount of energy transported, the **shorter** the wavelength. The distribution of this quantity over all wavelengths forms the **electromagnetic spectrum**.

EM spectrum	γ	X	UV	Visible	IR	μ	Radio
Wave length (nm)	0.01 $<$	0.01 - 10	10 - 380	380 - 750	750 - 10^6	10^6 - 10^8	10^8 - $10^1$6

The **main source** of ambient energy is **solar radiation** which is mainly expressed in the **visible**, **infra-red** and **ultra-violet** parts of the spectrum

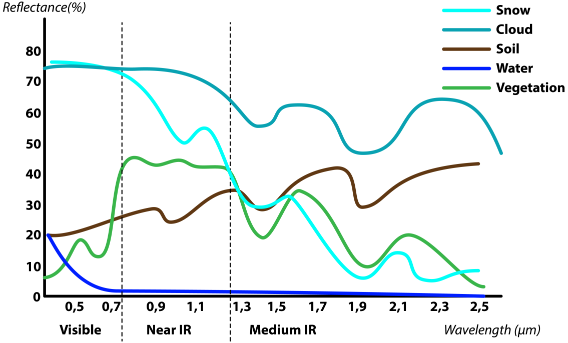


[Avanthey et al., 2017 – original : Degreen]

Reminder: spectral signature

Each **quantity** of energy **re-emitted** by an object over the entire spectrum constitutes a **spectral signature**

Objects of very **different natures** will have very **different spectral signatures**



[Avanthey et al., 2017]

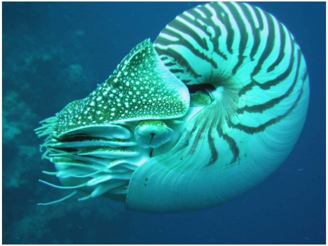
Reminder: imaging sensor

An **imaging sensor** (biological or artificial) must be able to:

1. **Capture** the **energy** emitted by illuminated objects so as to **form** an **image**
2. **Convert** this **energy** (photons) into an **interpretable signal** (electric current)

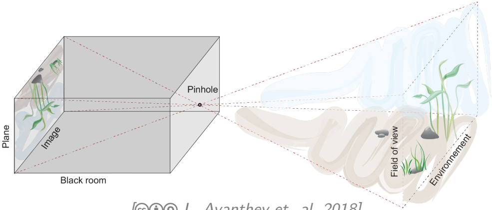
Capture the energy: simple imaging system

[cc by sa Hans Hillewaert, 2008]



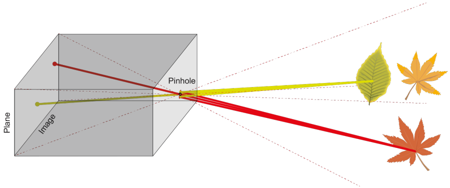
[cc by sa Lee R. Berger, 2007]

The eye of the Nautilus: model of the pinhole and dark room



[cc by sa L. Avanthey et. al, 2018]

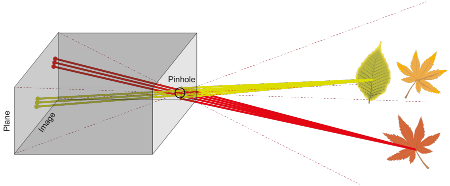
Simple imaging system: problematic



⇒ **Very small hole**

The number of **photons** passing through the hole is **small**: the resulting image is **dark**

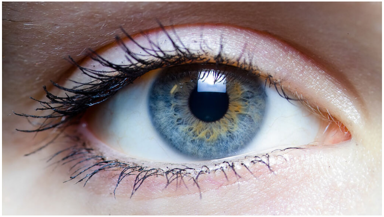
(except very long exposure time)



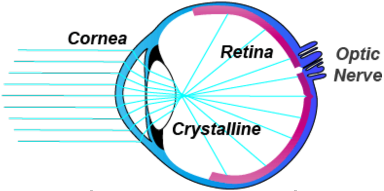
⇒ **Wider hole**

Several rays coming from the **same point** pass through the hole: the resulting image is well exposed but **fuzzy**

Capture the energy: complex imaging system

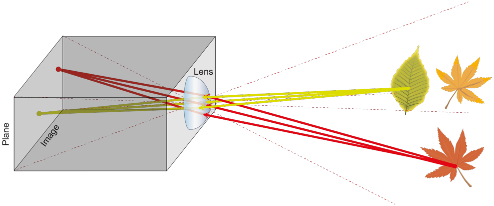


[Laitr Keiows, 2004]



[Avanthey et al., 2018]

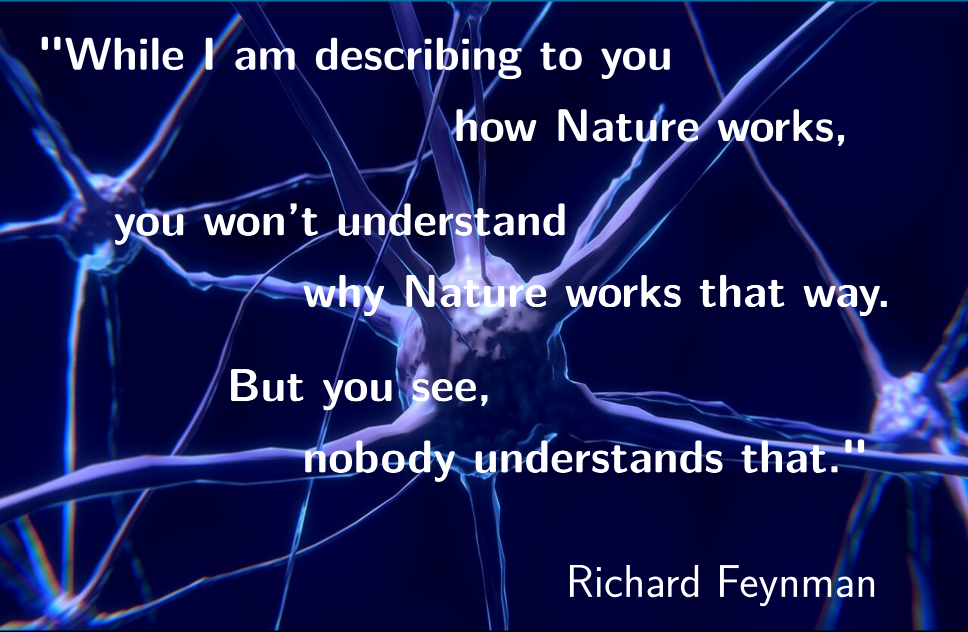
The human eye: wide hole + concentration of light



[L. Avanthey et. al, 2018]

Convert energy into an interpretable signal

Now that we have seen the **biological and artificial systems** capable of **capturing** an image, in the following we will focus on those capable of **transforming** these **amounts of energy** into an **electrical signal**.



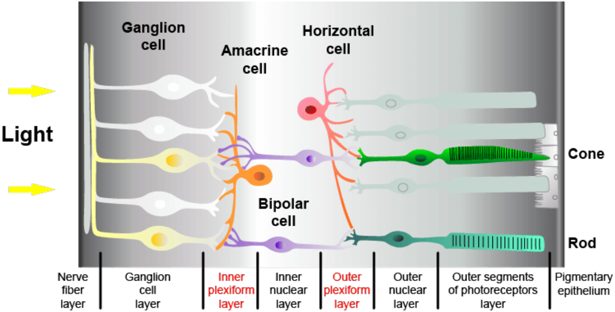
"While I am describing to you how Nature works, you won't understand why Nature works that way. But you see, nobody understands that."

Richard Feynman

Photosensitive surface: the retina of the human eye

The **retina** is a **neurosensory** layer that transforms **light** into **electrical signals**

⇒ It is a **complex** process that involves a **cascade of cells**. To simplify, we will retain only the **photoreceptor cells** called **cones** and **rods**.

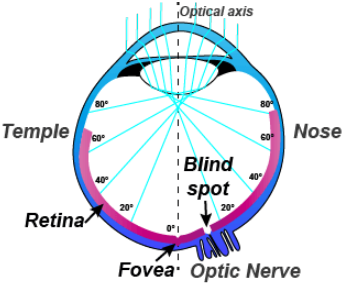


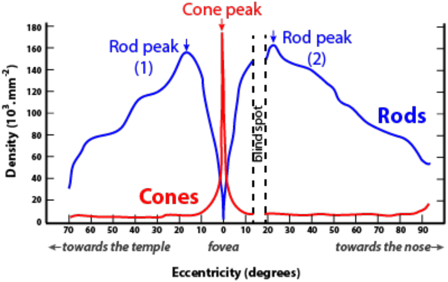
[Pancrat, 2011]

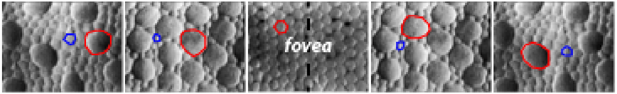
The **electrical signals** are then transmitted to the visual cortex trough the **optic nerve**

Retina of the human eye

The **retina** is not **symetrical**







[ⓒⓘⓢ L. Avanthey, 2018]

⇒ Presence of a **blind zone** (without cones or rods) on the nasal side where the **optic nerve** is located

⇒ The **cones** are located **near** the **fovea** while the **rods** are **farthest** from the **optical axis**

Frontal vision vs. lateral vision

Turn your head 20° to the left and try reading this text without moving your eyes

What do you notice?

Why?

Retina of the human eye: sensitivity

Retina works in **wide lightning conditions**

- **Day** vision *aka* **photopic** vision

  (from $\phi\omega\tau o\zeta$ light and $o\psi\iota\zeta$ view)

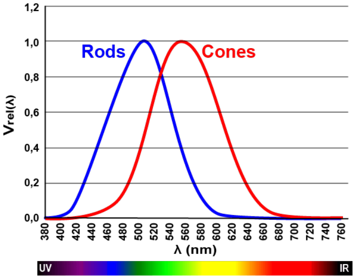
$\Rightarrow$ **trichromatic image** created by **cones**

- **Low light** vision *aka* **scotopic** vision

  (from $\sigma\kappa o\tau o\zeta$ darkness and $o\psi\iota\zeta$ view)

$\Rightarrow$ **monochromatic image** created by **rods**

(x100 more sensitive to photons than cones)



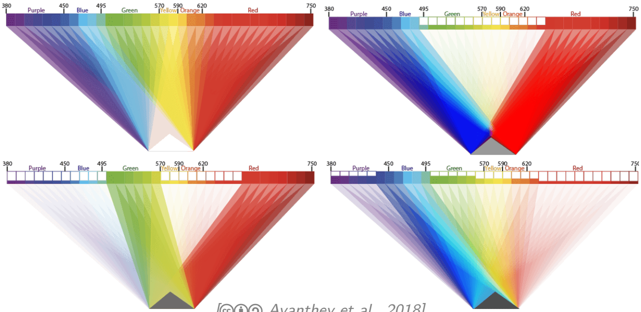
Normalized spectral sensivity of a human eye

Retina of the human eye: colors

We do **measurements** (discrete events) on the **spectral signature** (continuous phenomenon)

We thus **perceive** only the **integral** and not the **distribution**

$\Rightarrow$ If we **add** the **energy** of the whole spectra below, we obtain **different shades**: from white (maximum energy) to black (zero energy) through grays

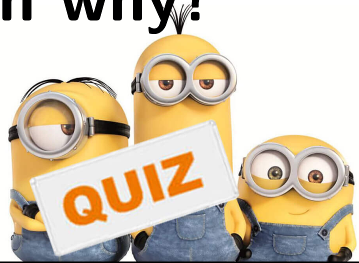


[Avanthey et al., 2018]

Retina of the human eye: colors?

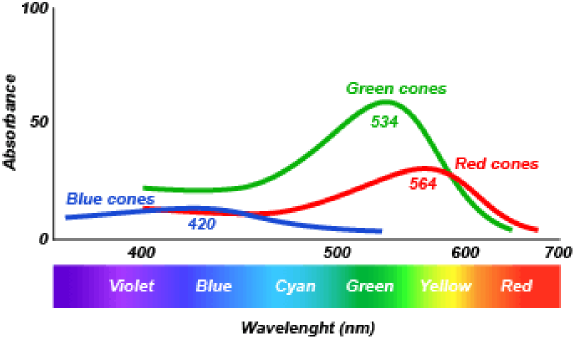
Except we don't see in grayscale...

Can you explain why?



Retina of the human eye: colors!

Color perception is done with **3 different** kinds of **cones** that have spectral sensitivity **centered** on red, green and blue

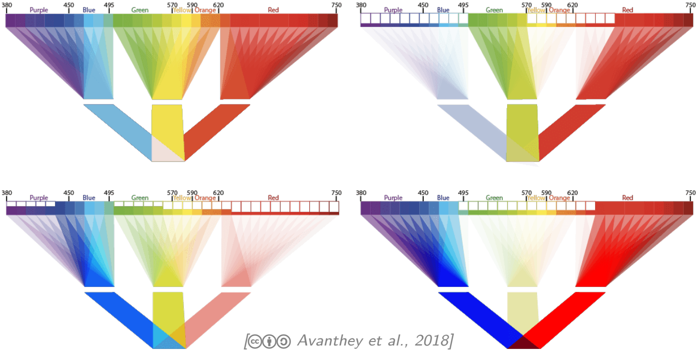


[cc by sa L. Avanthey, 2018]

⇒ The human eye is **more sensitive** to **green** (58.7%) and red (29.9%) than to blue (11.4%)

A spectrum ⇒ a color

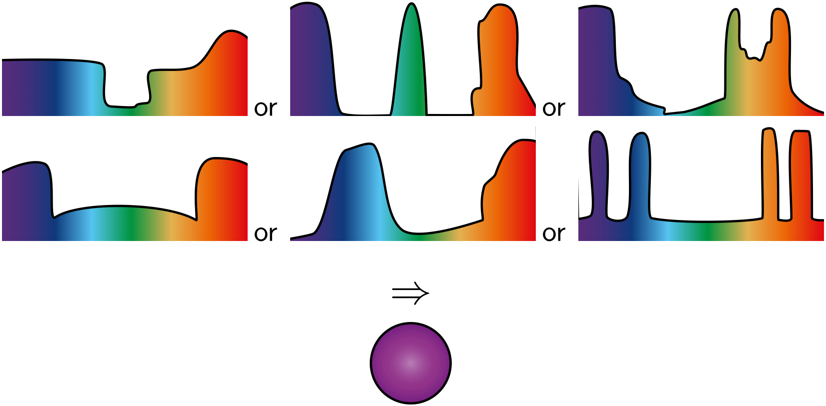
The **sum** therefore takes place in a **three-dimensional** space, the **brain** then **interprets** the result of these three integrals in a wide **variety** of **colors**



[Ⓒ Avanthey et al., 2018]

1 Spectrum ⇒ 1 Color ⇔ 1 Color ⇒ 1 Spectrum ?

We sum quantities of energies: **several configurations** of spectral signatures can therefore lead to the **same color**

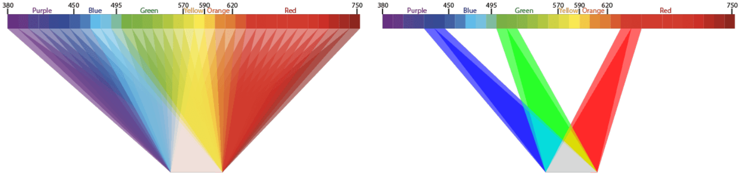


[Ⓒ Avanthey et al., 2018]

Young's Trichromatic Color theory

Due to the presence of our **3 types of cones**:

*"**Any color stimulus** (spectrum) can be **reproduced** by the **proportionate mixing** of **three monochrome stimuli** (spectra): red, green and blue"*



[Avanthey et al., 2018]

These **three primaries** are therefore **necessary** and **sufficient** to **reproduce any color** perceptible by a human eye

Perception of colors: additive synthesis

To resume : A **color** results from the interpretation of the brain of the three **sum of electromagnetic waves** in the visible part of the spectrum made by the red, green and blue cones

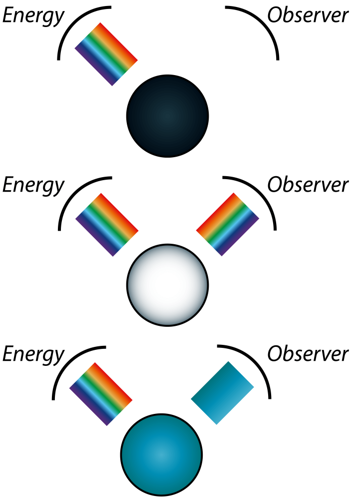
As electromagnetic waves are added, the process is called **additive synthesis**

But colors in **painting** result from the mix of the cyan, magenta and yellow colors, not the red, green and blue ones.

Why is it different?

The color of an object

- An object that appears **black** to us **absorbs** all incoming energy and does not emit any
- An object that appears **white** to us **re-emits** all the energy of the visible spectrum
- An object that appears **blue** to us **absorbs** all the wavelengths except those in the blue that it **reflects**



[Avanthey et al., 2018]

Colors in painting: substractive synthesis

In painting, each **pigment** used **absorbs** and **reflects** specifically the **incomming spectrum**

- In other words, the electromagnetic waves that have **not been absorbed** by the painting pigments produce the resulting **color**

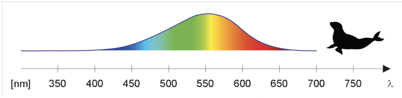
$\Rightarrow$ So, **adding pigments** means **adding** electromagnetic **filters**

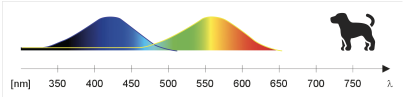
In **painting**, the **resulting color** is produced as the original spectrum **minus** the electromagnetic waves absorbed.
The process is called **substracting synthesis**.

Three types of cones in B, G and R: universal model?

Three types of cones in B, G and R: universal model?

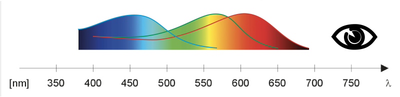
Monochromates (aquatic mammals) and **Dichromates** (mammals excluding primates + some fish)

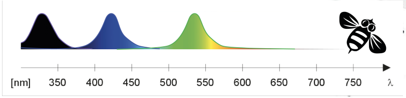




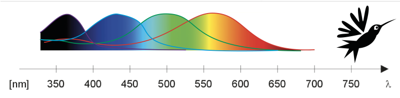
Trichromates (primates + insects + some fish)

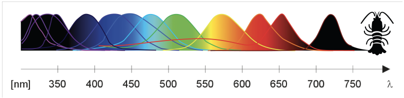
QUIZ





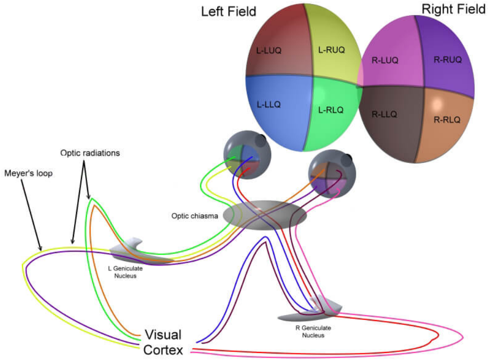
Tetrachromates (day birds + fish + reptiles) and **more** (Mantis shrimp)





 Avanthey et al., 2018]

Human visual perception : information processing



[Ratznium, 2007]

- **Electrical signals** from the retina are **transmitted** by optic nerves to optic chiasma and lateral genicular body
- Then, they dispatch them to the **visual cortex area** located in the back of the **brain**

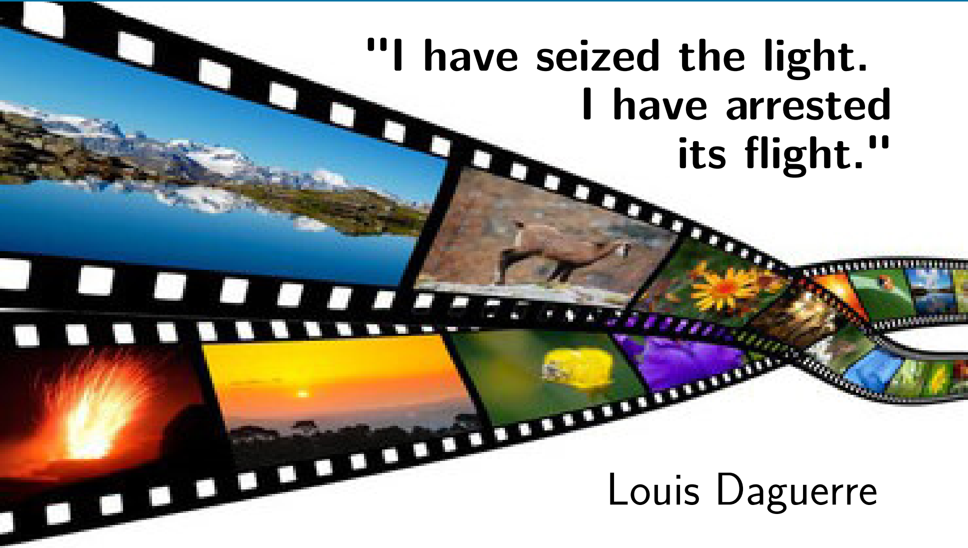
From human vision to artificial vision

The way of **how** visual cortex **interprate** these **data** is impossible to explain just by using **signal processing tools**!

⇒ Contributions from **other sciences** like linguistics, psychology, cognitive science, neuroscience, etc., are **essential** to **understand** the process

⇒ **Artificial vision systems** try to use **neural** or **deep learning technics** inspired by studies on human brain, but the **best results** are obtained when **isolating relevant image primitives**

Isolating relevant image primitives
is one **goal** of this course!



**"I have seized the light.
I have arrested
its flight."**

Louis Daguerre

Reminders: sensors, measurements and integrations

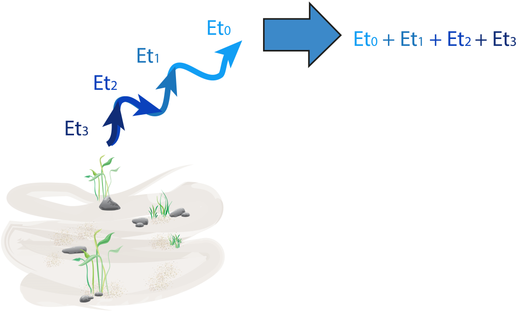
Energy is a **continuous phenomenon**, but a **measurement** is a **discrete event** (sampling)

$\Rightarrow$ To obtain a **measurement**, you need a **sufficient amount** of incoming energy (intensity)

We must thus **integrate** on one or more dimensions: time, space, frequency

Reminders: Temporal integration

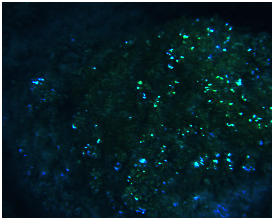
1. Either we **integrate** in **time** by increasing the **exposure time**



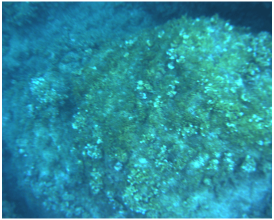
[🅭🅯🄎 *Avanthey et al., 2018]*

⇒ Sensitivity to **motion blur** if there is a relative movement of the scene with respect to the sensor

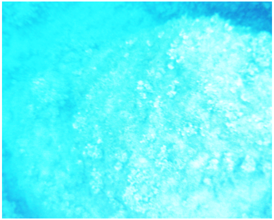
Temporal integration: illustrations



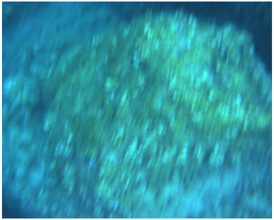
Underexposure



Good exposure



Overexposure

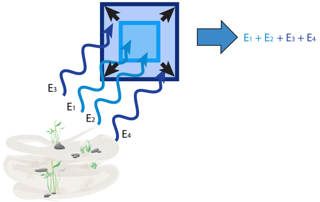


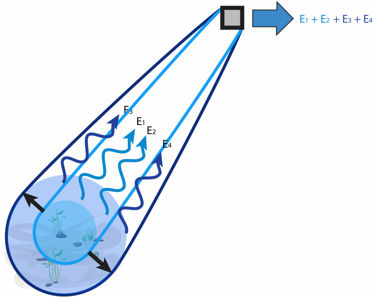
With movement (motion blur)

[Avanthey et al., 2013]

Reminders: Spatial integration

2. Either we **integrate** on the **spatial surface** by increasing the **size of the photosensitive elements** (grains or cells, at the construction) or by increasing the **size of the cones of analysis** (dynamically, during the use)

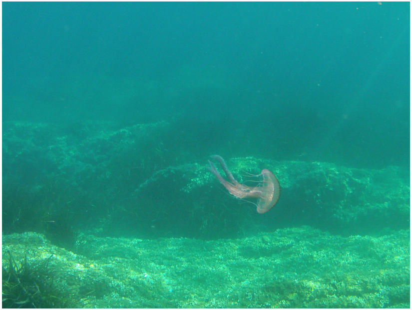




[Avanthey et al., 2018]

⇒ Decreased **spatial resolution** (loss of detail)

Spatial integration: illustrations



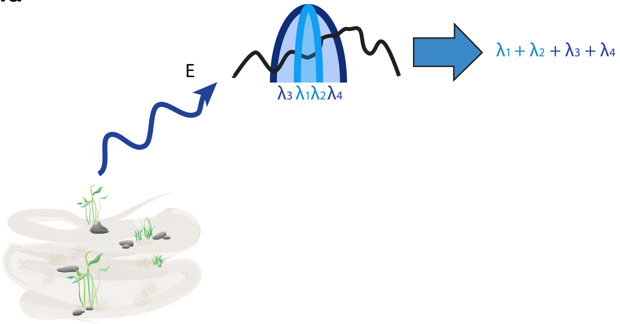
Thin

Wide

[Beaudoin et al., 2014]

Reminders: Spectral integration

3. Either we **integrate** over the width of the electromagnetic **spectrum** by increasing the **size** of the observed **spectral band**



[Avanthey et al., 2018]

⇒ Decrease of the **spectral resolution**: if $\sum_{i=n_0}^{n} \lambda_{iS_1} = \sum_{i=n_0}^{n} \lambda_{iS_2}$ then no possible distinction between the spectra S_1 and S_2

Spectral integration: illustrations

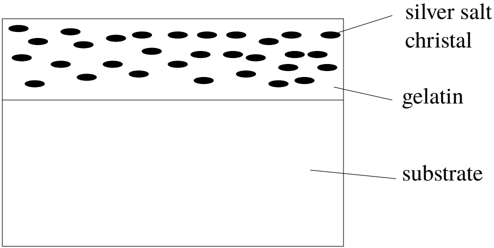
RGB

Panchromatic

[Ⓒⓘⓢ *Villard et al., 2018*]

Artificial photosensitive surface: photographic film

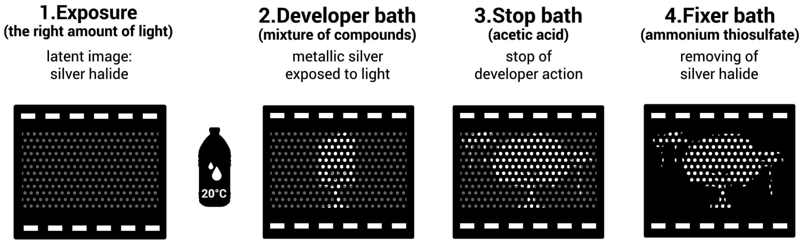
- A **panchromatic film** (black and white) is composed of a **silver salt emulsion** (light sensitive) coated onto a **substrate** (glass, cellulose, polyesther, paper...)



[CC BY SA Beaudoin et al., 2018]

The **more photons** are received by the film, the **more opaque** it becomes

Black and white negatives development

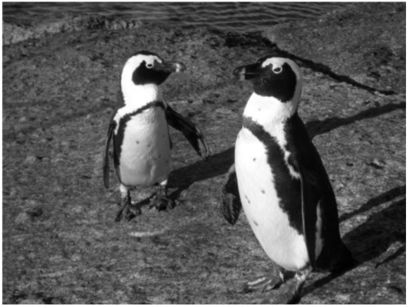


[🅭🅯🄎 *Beaudoin et al., 2018 – original: Marianna Caserta, DensityDesign Research Lab, 2014*]

From negatives to positives

⇒ The **previous process** produces an **image** where the **darkest parts** are the ones that have been **most illuminated**: we thus obtain a **negative**!

To get a **positive** (the final photograph), just make a **negative**... of the **negative**!



[Beaudoin et al., 2009]

Bonus: Latent image

Latent image for the first **special effects**!

The Man with the Rubber Head

[Méliès, 1902]

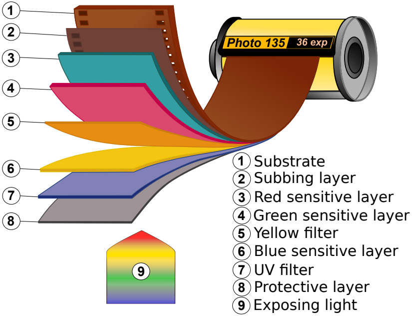
Color (trichromatic) photography

Same principle as for black and white photography

But:

⇒ chemical interactions are **more complex** (dye couplers)

⇒ color film is composed of **3 sensitive layers** centered on red, green and blue



[Beaudoin et al., 2018 – original: Voytek S., 2007]

Dematerialization: from analog to digital

- **Analog** photography has very **precise** spatial and colorimetric **resolutions**
- But it can only **fix** the measurements on a **physical medium**

⇒ The **new generation** of artificial imaging sensors, **digital sensors**, produce **dematerialized measurements**, which:

- can be **easily transmitted**
- are **natively** fully **compatible** with **digital processing**
- are **not as accurate** as analog measurements
  - but are **nowadays almost as accurate** as analog measurements

Digital CCD and CMOS sensors

The **most used digital sensors** are **CCDs** and **CMOS**

- **CCDs** (*Charge-Coupled Device*)

  (work of Smith and Boyle, Bell Lab., 1969 and nobelled in 2009)
- **CMOS** (*Complementarity Metal-Oxide-Semiconductor*)

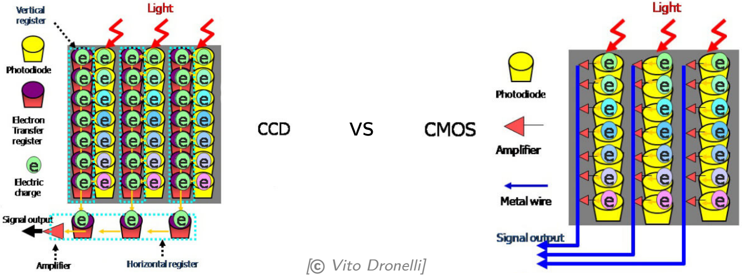
Both types of sensors are arrays of photoreceptors

Each photoreceptors converts the **photons** (quantity of energy received) into **electrons** and then into a **digital value** (0 to 255): the value of one **pixel** in the resulting digital image

CCD vs. CMOS

⇒ For **CCDs**, the electrons captured by each photoreceptor **migrate** step by step to a **high quality voltage converter**

⇒ For **CMOS**, each photoreceptor has its **own voltage converter** (no charge transfer) but of **poor quality**



[© Vito Dronelli]

CMOS sensors are **less expensive** and **consume less** than CCDs but are **less efficient** in **low light** and do not allow to recover the **real energy captured**

Shutter : global or rolling

There are **two ways** to **expose** the **photoreceptor grid**

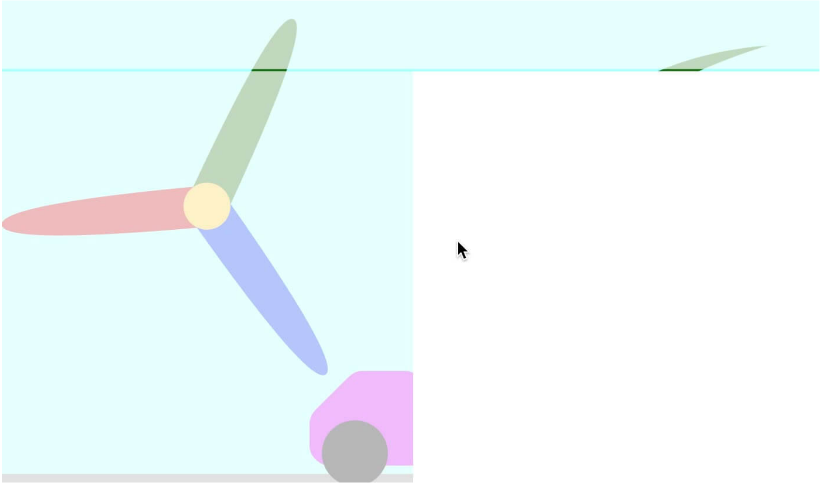
⇒ **All photoreceptors** are **exposed** exactly at the **same time**: the **shutter** is **global**

But with **miniaturization** (less buffer) and **improved resolutions** (more photoreceptors) it is **difficult** to **quickly** process the information: the shooting **frequency** is **lower**

⇒ A **line of photoreceptors** is exposed, the information is **processed** while the **next line** is exposed: the **shutter** is **rolling**

With a **rolling shutter**, there may be **consequences** on your **images** if the scene is very **dynamic**!

Rolling shutter with a dynamic scene



Example images with rolling shutter

[🅭🅯🄎 Axel1963, 2010]

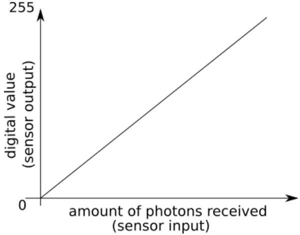
[🅭🅯🄎 Jonen, 2007]

[🅭🅯🄎 Richmilliron, 2011]

[🅭🅯🄎 Gademer, 2009]

From photons to electrons (analog) to a digital measure

Like the **silver salts** of analog **films**, each **photoreceptor** of a digital sensor will give a **response** (digital value) **proportional** to the quantity of light (radiometry) received



[CC BY SA *Beaudoin et al., 2018*]

A **zero** (respectively **high**) **amount of photons** gives a **zero** (respectively **high**) **digital value**

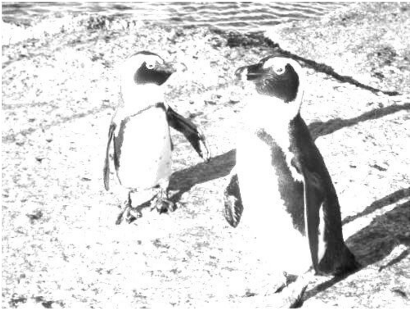
The function that defines for each sensor input a digital value is called a **LUT** (**Look-Up Table**).

But in practice...

⇒ There is **no reason** that at the time of acquisition, the **minimum quantity** of photons received is **zero** or that the **maximum quantity** corresponds to the digital value of **255**.

Consequently, the use of the **256 values** will **not** be **optimal**!

Sensor output max $<$ 255
(dark and low contrast image)



Sensor output max $>$ 255
(clear and satured image)

[Beaudoin et al., 2018]

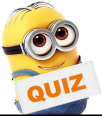
A little hindsight

Is it important

to try to get an image with a

good contrast?

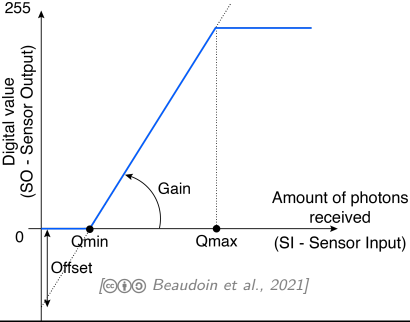
Why?



Gain and Offset

To obtain the best contrast, you need:

- set the **minimum digital value** (Sensor Output – SO) to the **minimum** amount of photon **received** (Q_{min}): $SO = 0$ if $Q < Q_{min}$
- set the **maximum digital value** to the **maximum** amount of photon **received** (Q_{max}): $SO = 255$ if $Q > Q_{max}$
- **vary linearly** between these two borns: $SO = Gain \times Q + Offset$



[Beaudoin et al., 2021]

Small computational break

Knowing Q_{min} and Q_{max},
what is the gain and the offset ?

Small computational break

**Knowing Q_{min} and Q_{max},
what is the gain and the offset ?**

$$\begin{cases} 0 = Gain \times Q_{min} + Offset \\ 255 = Gain \times Q_{max} + Offset \end{cases} \tag{1}$$

Small computational break

Knowing Q_{min} and Q_{max},
what is the gain and the offset ?

$$\begin{cases} 0 = Gain \times Q_{min} + Offset \\ 255 = Gain \times Q_{max} + Offset \end{cases} \quad (1)$$

$$\begin{cases} Offset = -Gain \times Q_{min} \\ 255 = Gain \times (Q_{max} - Q_{min}) \end{cases} \quad (2)$$

Small computational break

Knowing Q_{min} and Q_{max},
what is the gain and the offset ?

$$\begin{cases} 0 & = & Gain \times Q_{min} + Offset \\ 255 & = & Gain \times Q_{max} + Offset \end{cases} \tag{1}$$

$$\begin{cases} Offset & = & -Gain \times Q_{min} \\ 255 & = & Gain \times (Q_{max} - Q_{min}) \end{cases} \tag{2}$$

$$\begin{cases} Gain & = & \frac{255}{Q_{max} - Q_{min}} \\ & & \\ Offset & = & -\frac{255 \times Q_{min}}{Q_{max} - Q_{min}} \end{cases} \tag{3}$$

Gain and Offset in real life?

The above calculations **assume** that we **know in advance** what will be the **actual values** of Q_{max} and Q_{min} ! But the **lighting conditions** are **variable**...

In practice:

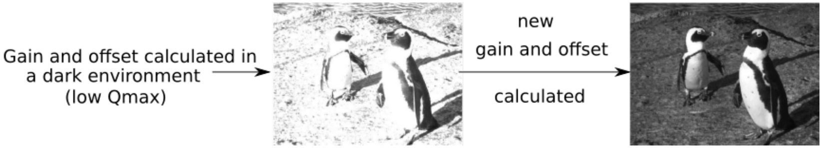
- ⇒ a **new acquisition** is done very **often** by the digital sensor
- ⇒ then the values Q_{max} and Q_{min} are **calculated**
- ⇒ and the values of the Gain and the Offset are **adapted** *a posteriori*

This **explains why** your phones, cameras or camcorders **need an adaptation time** during **sudden changes** in lighting conditions!

Gain and Offset: some examples

- **Next acquisitions:**

  ⇒ after going **from** a **dark** area **to** a **brighter** area



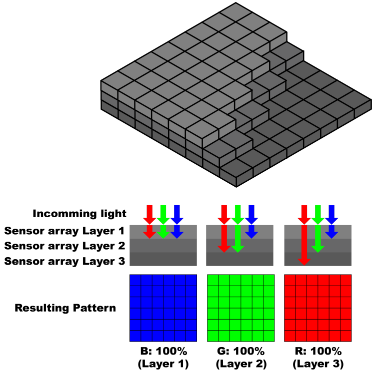
  ⇒ after going **from** a **bright** area **to** a **darker** area

[cc by sa Beaudoin et al., 2018]

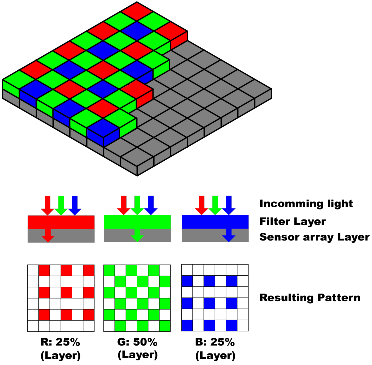
Color artificial sensors

Color sensors mainly use 2 techniques: **depth** or **mosaic**

Vertical color filter
Foveon X3



Lateral color filter mosaic
Bayer matrix



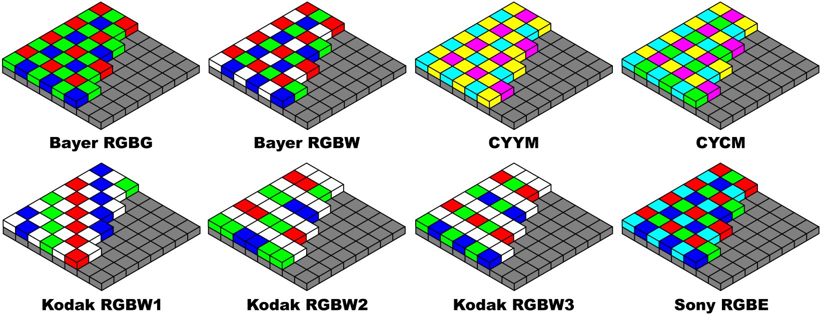
[CC BY SA *L. Avanthey et al., 2019*]

Color Filter Mosaic

The **most common** technology is **mosaic**

- A **colored filter** is placed in front of each **photoreceptor** to measure in different spectra (red, blue, green, etc.)

The **Bayer pattern** is the **most used**, but there are many **more**, here are some **examples**:



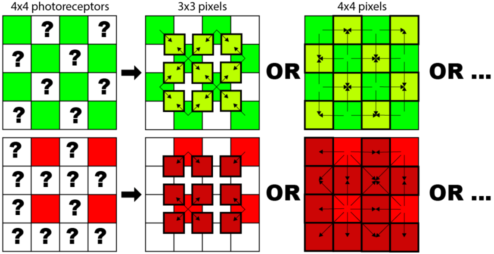
[CC BY SA L. Avanthey et al., 2019]

Color Filter Mosaic: image reconstruction

- Unlike vertical sensors that **directly** obtain the numerical values R, G & B for each pixel, lateral sensors need **demosaicing algorithms**

One of the **central challenges** is the **interpolation** of photoreceptor values to **combine** the channels of **nearby cells** into one pixel

This is not a trivial problem:



[L. Avanthey et al., 2019]

- Many algorithms exist, seeking to **minimize** the loss of resolution, the appearance of artifacts and the calculation cost depending on the matrix used

Color Filter Mosaic : color weighting

Two-thirds of a digital RGB image is actually a **reconstruction**!

Another **related issue** is **channel cell weighting**: how much of each color should be combined to form the RGB pixel?

- The sensor **material** (silicone) will be **more sensitive** to certain wavelengths. This effect must be **corrected** with the desired balance.

Blue is the **most problematic** color because it is **absorbed very quickly** and has a hard time **reaching** the photoreceptors.

⇒ The smaller the photoreceptors are, the less we will capture in the blue spectrum, the more we have to **amplifier** the result, the more we will have **noise**.

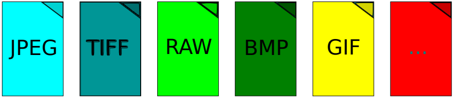
"Digital for storage and quickness.

Analog for fatness and warmth."

Adrian Belew

File Types

⇒ **Raster image**: **array** of **pixels** (*picture element*), quality **dependent** of the resolution



⇒ **Vector image**: **curved** based **graphics**, quality **independant** of the resolution

Only **raster images** will be used in this module

File types of raster image

Why is there not a **universal file type** for raster images?

File types of raster image

Why is there not a **universal file type** for raster images?

Each has its **benefits** and **drawbacks fitting** to a **specific use**: as there is **no universal use**, there is **no universal type**!

Among these uses, we can distinguish:

⇒ those that **preserve the original data** acquired (`raw`, `bmp`...)

- **Benefit**: raw data
- **Drawback**: large recorded file

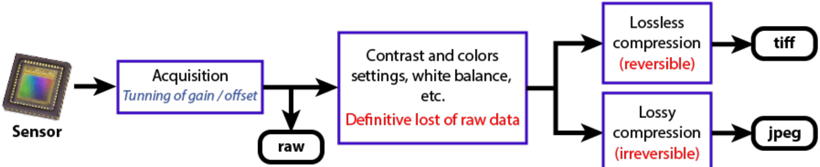
⇒ those that **optimize the size of the recorded data** (`jpeg`, `tiff`, `png` ...)

- **Benefit**: heavy recorded file able to be upload on web
- **Drawback**: definitive lost of raw data

Contents of a raster image file

Size of the image	Number of lines and colums
Number of channels	1 (grayscale image), 3 (color image) or 4 (color + transparency)
Depth of each channel	The number of bits per pixel, usually 1 byte (the sum of channel's depths makes the image depth, usually 24 bits)
"Magic" number	Identify the file type (jpeg, raw, tiff...) independently of its extension
Data	The value of each pixel in each channel
Optionnal data (EXIF data – Exchangeable Image file Format)	Date, hour, camera settings (iso, shutter speed, focal length, etc.), GPS position, etc.

Record of a digital photography: a pratical example



[Beaudoin et al., 2018]

There is **no standardization** of the `raw` format! So, it is not always possible to **read** its data with all **imaging softwares**!

How to create a video file?

The principle is the same as for cinema: a **succession** of **images** at a **rate** of 30 images per second

Computing **full resolution images** with a **high rate** implies **efficient** hardware or network ressources that are often not available!

The solution is to **COmpress** images in a file that is **DE-Compress** when it is displayed (CODEC)

Compression of video files

Two strategies are possible to **compress** video files

1. Compressing **each** single image
   ⇒ but the result could lead to a pixelization effect

2. **Temporally** compressing the sequence either by :
   1. simply removing some single images from the sequence
      ⇒ but then there is a lost in fluidity
   2. stocking some key images and then only their differences with the next images

As for images, there are **many different video file formats** depending on the specific **use** (AVI, MPEG, Quick Time, real Video, etc.)

End of the exploration of Imaging

We have completed our initiatory exploration of how images are formed!

⇒ On this way, we have discovered the birth of **images**, how evolution deals with **vision** and how humans try to **reproduce** those systems in the **digital world** to be able to perform **artifical vision**

Our next exploration?
Preparing and improving images!

To Be Continued...