Digital Photography - Summary

Sebastien Speierer

9 juin 2014

Table des matières

1 Image Formation Model

Color signal : $C_i(\lambda) = E(\lambda) \times S_i(\lambda)$
The sensor response : $\rho = \int_\lambda E(\lambda)S(\lambda)R(\lambda)d\lambda$
- $C_i(\lambda)$ is what we see
- $E(\lambda)$ is the illuminance
- $S_i(\lambda)$ is the surface reflectance (i is the different colors channels(RGB))
- λ is the wavelength

1.1 Human visual system

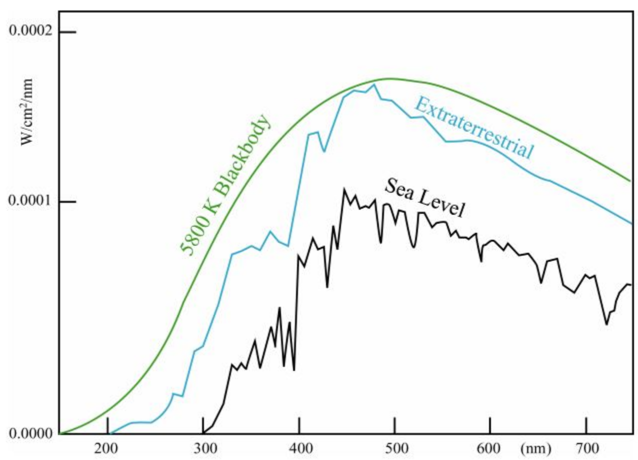
$$(L, M, S) = \int_\lambda E(\lambda)S(\lambda)(l, m, s)(\lambda)d\lambda$$

1.2 Camera measurements

$$(R, G, B) = \int_\lambda E(\lambda)S(\lambda)(r, g, b)(\lambda)d\lambda$$

1.3 Illuminant

Spectral Power Distribution are used to describe light source (normalized to 1)

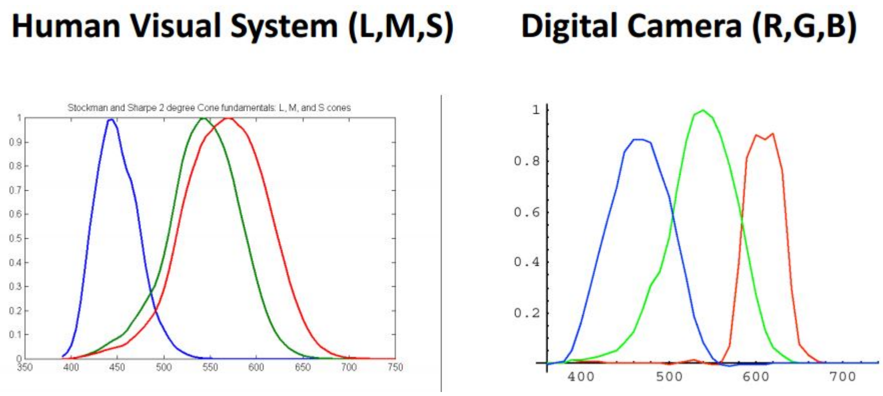


1.4 Reflectance

The color of an object is characterized by is surface reflectance $S(\lambda)$

1.5 Lambertian reflectance

Lambertian reflectance is the property that defines an ideal "matte" or diffusely reflecting surface. The apparent brightness of such a surface to an observer is **the same regardless of the observer's angle of view**



1.6 Spectral Sensitivity

Any physical sensor can have a number of filters (or channels) with different sensitivities $R_k(\lambda)$.
We usually call the three responses :
- **RGB** : for sensors that have their peak sensitivities in the red, green or blue wavelength part of the visible spectrum.
- **XYZ** : when the sensors correspond to the CIE color matching functions
- **LMS** : when the sensors correspond to cone fundamentals

2 Light and Light Sources

Light is a stream of mass-less particles.

Light bends when it changes the medium it travels through ; it slows down when it enters a medium denser than vacuum.

In quantum theory : everything is discrete and comes in package (packets of energy = particles) :
- packet of electromagnetic energy = photon
- number of photons = intensity of radiation
- "color" of a photon = wavelength of radiation

The **energy** of the emitted radiation is :

$$Q = hc/\lambda$$

with λ the wavelength, c the speed of light and $h = 6.6 \cdot 10^{-34}$ a constant.
The **radiant power (flux)** is radiant energy emitted in a time interval (Watt)
The **number of photon** is defined by

$$N_v = \frac{P_v}{hv} = \frac{1}{hc^2} P_\lambda \lambda^3$$

2.1 Black Body

Every object emits energy from its surface when heated (> 0 K). A Black Body Radiator is an ideal solid object that can absorb and emit electromagnetic radiation in all spectrum so that all incident radiation is completely absorbed and emission is possible in all wavelength and directions.
Planck's formula : The Plank's formula say that the amount of energy released is dependent on the frequency of radiations :

$$E = hv$$

Stefan-Boltzmann's law : As the temperature increases, the amount of radiation at each wavelengths increases.

2.2 Color Temperature

The color temperature of a light source is the temperature of a black body that would emit radiation of substantially the same spectral distribution in the visible region as the radiation from the light source.

2.3 Color Temperature of Common Light Sources

Candle	1800 K
60-Watt Tungsten Lamp	2800 K
Halogen Lamp	3800 K
Direct Sunlight	4800
Photographic Daylight	5500 K
Electronic Flash	6000 K
Sky light (blue)	10'000 - 20'000 K

2.4 Photometry

Measurement of the electromagnetic radiation between 380 and 750 nm and weighs the measurements according to the luminosity function of the human eye.
The luminous flux (lumens) :

$$\Phi_L = 683 \int_{380}^{750} v(\lambda)\Phi_E(\lambda)d(\lambda)$$

with $v(\lambda)$ the luminosity function and $\Phi_E(\lambda)$ the radiant flux per wavelength (Watt/nm).
The Inverse Square Law :

$$E = \frac{I}{d^2}$$

with E the illuminance, I the intensity and d the distance
Finally, we can define the luminance has the intensity over the area : $L = \frac{I}{A}$.

3 Human Vision

There are the different parts of the human eye :
- **The Cornea** provides 3/4 of the optical power necessary to focus light on to the retina.
- **The Lens** : The shape of the lens is controlled by ciliary muscles : when focusing on nearby object, the lens gets fatter. When focusing on far away objects, the lens becomes ”flatter”.
- **The Iris** is a ring of muscles that regulates the amount of light entering in the eye by adjusting the size of the pupil.
- **The Pupil** is the adjustable opening at the center of the iris taht allows varying amount of light to enter the eye.
- **The Retina** Contains visual systems photosensitive cells and initial signal processing and transmission circuitry. The **blind spot** is a small area in the retina where there is no visual sensation (there is the optic nerve).
- **The Fovea** is the central part of the retina that provides best spatial and color vision.
- **Rods** : primarily located outside the fovea (120 million) and used for low luminance vision (monochromatic vision)
- **Cones** : primarily located in the fovea (7 million) and used for high luminance level. There are three types of cones (L,M,S) with different spectral sensitivities.

3.1 Dark and light adaptation

The adjustement to lighting situation is not immediate.
The HVS doesn't retain absolute luminance values of the scene, but only relative values.
Weber-Fechner Law says that as luminance increases, larger changes in luminance are needed to perceive a just noticeable difference : $\frac{\Delta L}{L} = constante$

3.2 Color Imaging

- **Additive** : used for capture, image processing, and additive display encodings (RGB, LMS, XYZ)
- **Substractive** : used for print (Cyan, Magenta, Yellow and black)
- **Opponent** : used for compression, color ordering systems and calculating visual difference metrics (YCrCb, HVS, Lab).

Lateral Inhibition : Light on one part of retina inhibits adjacent regions to react to light.
Sharpening : The HVS increases the local contrast at an edge.
Uniform Field Contrast : The response of the HVS depends on the local luminance variation to the surrounding luminance.
Threshold Contrast : the minimum contrast necessary for an observer to detect a change in intensity.
Gamma : Because of the non-linearity of HVS, we use the gamma correction. The gamma-characteristic of displays is an approximate inverse of the nonlinearity of human brightness perception.
Contrast sensitivity function : The HVS ability to discriminate fine detail has been determined by its contrast sensitivity function :

$$Constrast = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$

The relative low sensitivity of the HVS to high frequency (spatial of temporal) is exploited in compression (we can reduce the quality of the "color part" of the image).

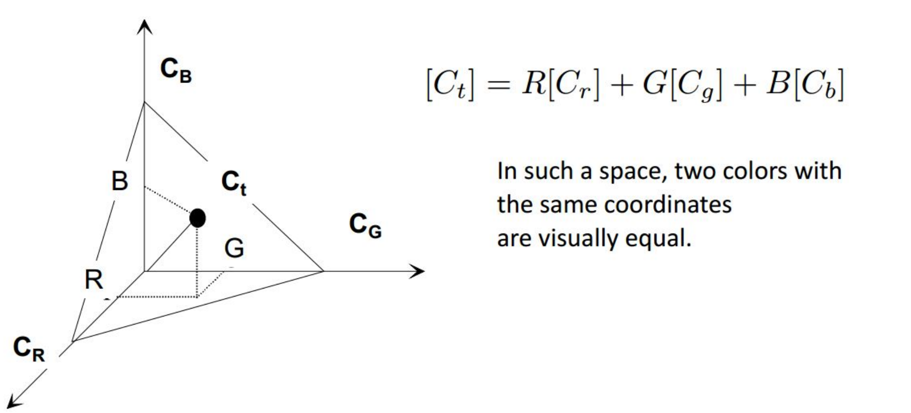
4 Color and Colorimetry

Colors have three main perceptual attributes :
- **Hue** : refers to the appearance of a color
- **Colourfulness** : denotes the extent to which the hue is appartent.
- **Brightness** : denotes the extent to which a color apperas to exhibit more or less light.

Chroma is defined as the colourfulness of an area judge in proportion to the brightness of a similarly illuminated area that appears to be white. Chroma is a relative colourfulness.
Lightness is defined as the brightness of an area judged in proportion to the brightness of a similarly illuminated area that appears to be white.
Saturation is defined as the colourfulness of an area judged in proportion to its own brightness.

We observed that reddish-greenish or yellowish-bluish hues cannot be perceived simultaneously, so we proposed that their are three types of visual receptors with bipolar responses to light-dark (L+M+S), red-green (L-M+S) and yellow-blue (L+M-S). The neuron of the retina encode the colors and this transformation serves to disassociate luminance and color information.

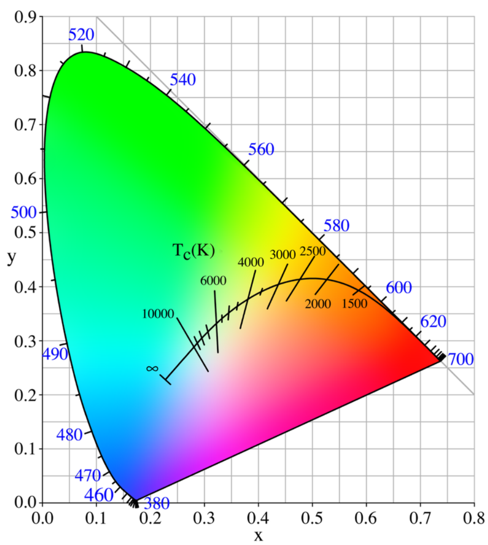
We can represent a color in a three-dimensional space, where each axis represents the color matching primaries.



The HVS has the ability of the visual system to adjust its colour sensitivity according to the colour illumination called **chromatic adaptation**. This allow it to keep the white point in a scene more or less consistently white, independently of the color of the illuminating light source.

4.1 Chromaticity diagrams

Chromaticity is an objective specification of the quality of a color regardless of its luminance. Chromaticity consists of two independent parameters, often specified as hue (h) and colorfulness (s). This number of parameters follows from **trichromacy** of vision of most humans, which is assumed in most models in the color science. In color science, the white point of an illuminant or of a display is a neutral reference characterized by a chromaticity ; all other chromaticities may be defined in relation to this reference using polar coordinates. The hue is the angular component, and the purity is the radial component, normalized by the maximum radius for that hue.



4.2 Color Constancy

Color constancy is a feature of the human color perception system which ensures that the perceived color of objects remains relatively constant under varying illumination conditions. A green apple for instance looks green to us at midday, when the main illumination is white sunlight, and also at sunset, when the main illumination is red.

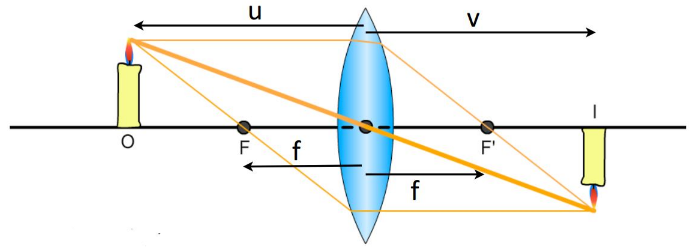
5 Optics and Lenses

A perfect lens is capable of focusing all light rays from infinity in a single point. But it doesn't exist and we have several problems with real lens :
- **Chromatic aberration** are caused by the camera lens focusing different wavelengths of light onto different focal planes.
- **Spherical Aberrations** appear when the focal length of a lens will vary depending on how far a point is from the centre of the lens.

There are the different lens formulas :

5.1 Photographic Lens

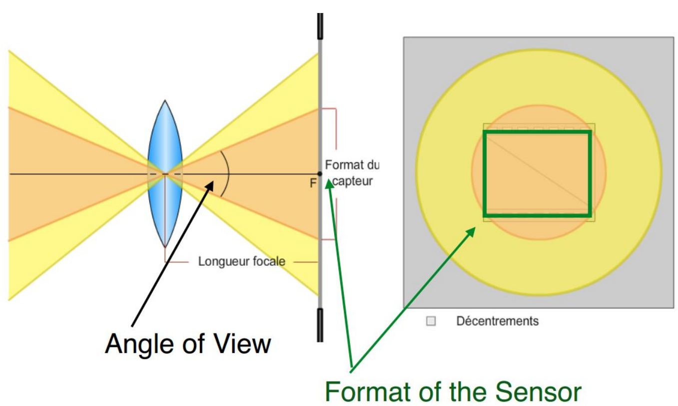
A photographic lens consists of more than one lens element. It can be characterized by its :
- **Focal Length** defines the distance between the lens and the sensor.



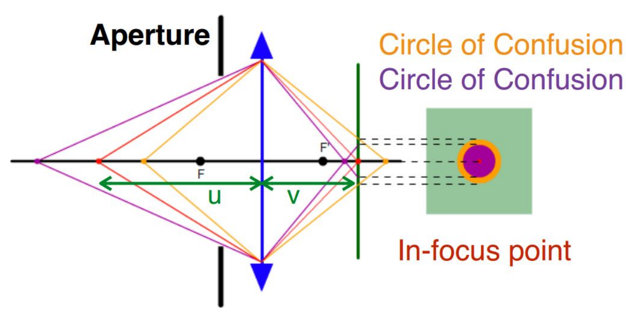
$$
\begin{aligned}
1/f &= 1/u + 1/v \\
m &= v/u = I/O \\
m &= (v-f)/f \\
m &= f/(u-f)
\end{aligned}
$$

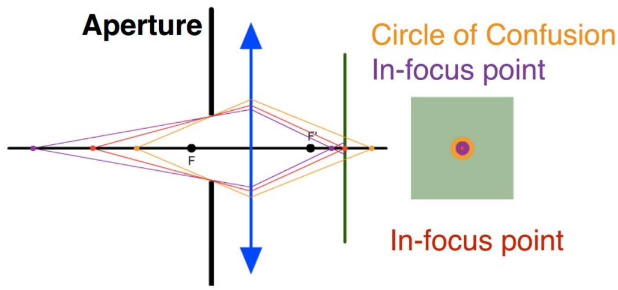
- f = focal length
- u = object distance
- v = image distance
- I = Image Size
- O = Object Size
- m = scale of reproduction

– **Angle of view** depends on the size of the sensor and on the focal length. This ratio can produce a eye-fish effect. If the lens doesn't fit the sensor, we can observe a effect called vignette. Finally, the angle of view define the "zoom".

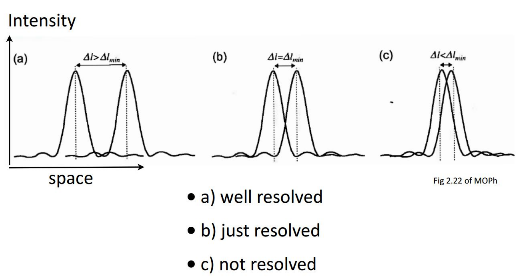


– **Aperture** : The ratio between the focal length and the diameter of entrance pupil for infinity focus is called the **relative aperture** (f-number). The more this ratio is high, the more the depth of field will be long. Increasing the f-number → decreasing aperture size.





– **Resolving Power** depends on the diffraction and define the minimum resolvable detail (distance between two different point that we can still distinguish).



6 Fourier and Convolution

Any pattern can be created by super.imposing a large enough variety of sinusoidal patterns (different frequencies, amplitudes, orientations, phases). **Fourier Transfom formula** :

$$F(u) = \int_{\infty}^{\infty} f(x)e^{-2\pi iux}dx$$

6.1 Point Spread Function (PSF)

Real imaging processes are not ideal and so points in the original are not reproduced as points in the image. An object point is imaged with small concentric rings of light, given by the diffraction of the light. So we use a point spread function for every "pixel". The image of a point of intensity is the point spread function. Images are formed from the summation of an infinite number of overlapping PSF. The shape of the PSF determines the sharpness and resolution aspects of image quality produced.

6.2 Convolution

$$h(x) = \int_{\infty}^{\infty} f(e)g(x-e)de \Rightarrow H(u) = F(u)G(u)$$

7 Digital Image Processing

Capture : CCD and CMOS that convert irradiance to voltage
Sampling : discretization of domain (space). We can represent a sampling function by a row of impulses (regularly displaced Dirac delta function). The sampling intercal is the distance between adjacent two pixel centers.
Quantization : discretization of values (voltage). It is a projection of a continuous signal onto finite set of discrete values. The level (in bit per pixel) is 1 (B/W), 8 (after processing), 16 (typical at capture)

A histogram describes the number of pixels per quantization level.

7.1 Image Processing

In spatial domain : pixel operations (addition, substraction,...), neighbourhood operations, geometric operations,... There are applications like contrast enhancement, filtering, interpolation, rotation,...
In transform domain : coding, compression, filtering,...
There is the formula for the two-dimensional discrete convolution that is used in digital image processing :

$$f(i,j) \otimes h(i,j) = \sum_{m=-\infty}^{\infty} \sum_{\infty}^{\infty} h(m,n) f(i-m, j-n)$$

Knowing the convolution theorem, we know that linear spatial filtering can be implemented by multiplication in the frequency domain.
Low-pass filter : can be used for blurring the image or smoothing the noise.
High-boost filter : this filter doesn't remove low frequency, but just boost the frequency over a level. This will create the illusion that the image is sharpened.
An **image gradient** is a directional change in the intensity or color in an image. We can use image gradient for edge detection.
Denoising can be done by applying a low-pass filter (image high-frequencies are noise) or a median filter. We can also averaging over many images.
Median Filter is a non-linear filter which will just select the median value of the neighbourhood. This is an effective way to remove extreme values. It can be used to remove outliers or denoising in case of salt and pepper noise (B/W).

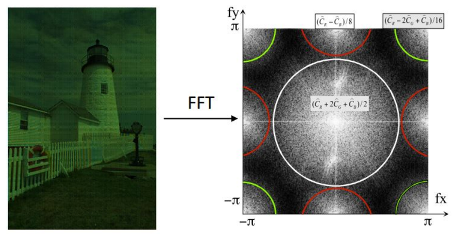
8 In-camera processing

8.1 Sensor

There are two types of sensor : CCD (Charge Coupled Device) and CMOS (Complementary Metal Oxide Semiconductor). They both use photosensitive material like semiconductor (Silicon).
Sensor response : The number of electons retained is proportional to the incident illuminance. The Sensor has thus a linear response curve.
The Charge Capacity is the max charge level where the sensor response is still liner (80% of saturation).
Dynamic range is the ratio of charge capacity to dark and read noise.
Pixel (photo site) size is small if the sensor is small. It varies between 1.5 to 8 microns. The distance between two adjacent pixels is called the pixel pitch and determines the sampling frequency.

8.2 Raw sensor

A camera raw image file contains minimally processed data from the image sensor of either a digital camera, image scanner, or motion picture film scanner. (Wave pour les images) There are dozens if not hundreds of raw formats in use by different models of digital equipment and they are also sometimes called digital negative.
Demosaicing : in a mono-CCD digital camera, one color is encoded per spatial position. There are two interpolation methods once the masks applied : Bilinear Interpolation (simple filter) and the Hue based interpolation (use the natural correlation between color channels).
If we take the fourier transform of CFA image, the luminance and chrominance are localized in specific areas of the Fourier domain :
So we can apply a filter (not a simple low-pass this time) in order to get a better luminance estimation. After this process of color demosaicking, we still have a RAW format image.



8.3 Input-referred Image Encoding

We apply some camera compensation like exposure adjustments, white balancing, color transform, and sensor characterization to the RAW image to get the input-referred image. This image represents the colorimetric coordinates of the color elements of a source image.

8.4 Output-referred Image Encoding

Color Rendering : Produce a colorimetric description of the desired image on the reference output encoding in the specified viewing conditions.
Tone Mapping : the most important step for good reproduction quality is to map the lightness of a picture (black to black and white to white) This part really depends on the dynamic range of the scene and that of the display. Local tone mapping bring a better imitation of the HVS local adaptation because it takes advantage of the fact that the HVS has a greater sensitivity to relative than absolute luminance levels. It also compresses the global contrast while increasing or keeping intact the local contrast.
B and W point correction : Maps the white point of the image to the brightest display luminance and the black point of the image to the darkest display luminance. These points are defined by groups of dark or bright pixels and not just the min/max pixel value (can be outlier). To define these group, we can use a low-pass filter or work with the histogram.
Gamma correction is conventionally modeled by a power function $y = x^{\gamma}$. Used to get a non-linear relationship between the input voltage of a display device and the output intensity of the display. Luminance is non-linear related to the voltage.
Key is a photographic term used to describe the content of an image :
- Low-key : a low-key image is one that contains predominantly dark tones and colors.
- Normal image : Normal distribution of tones and colors.
- High-key : the image contains predominantly light tones and colors.

Holm's photographic tone reproduction is a more sophisticated global tone mapping method. It assigns tonal value to objects in the scene. It defines tone reproduction curves according to scene statistics (mean exposure, low-key, high-key) and also to the output media range and the viewing condition.
Then we get a output-referred image which represents the colorimetric coordinates of the color elements of a reproduction of the source image.
sRGB : it's the preferred color image encoding of most consumer digital cameras. It is optimized for CRT monitor output and it is based on a standardized transform from XYZ under D65.
Then we apply a **Color Component Transfer Function** in order to mimic the non-linear sensitivity of the human visual system. Colour component transfer functions are frequently used to account for the non-linear response of a reference device (printer or monitor) and/or to improve the visual uniformity of a colour space.

9 Image Encoding, Compression and File Formats

Coding : representation of information with symbols.
Source coding or compression : Reduction of redundancy/data rate
Channel coding : Adapting data to transmission channel

Lossless compression : perfect reconstruction possible
Lossy compression : extraction of most important data
Entropy : avergae information, maximum for symbols with equal probability and theorical limit for lossless compression

$$H = -\sum_k P(a_k) \log_2 P(a_k)$$

Huffman coding : (used in JPEG) is a variable-length entropy coding. It represent each symbol by codeword such that its number of bits is inversely proportional to the symbol probability.
Lempel-Ziv coding : (used in GIF/PNG) is a dictionary-based encoding. Is good when statistics aren't known. The scenario described encodes sequences of 8-bit data as fixed-length 12-bit codes. The codes from 0 to 255 represent 1-character sequences consisting of the corresponding 8-bit character, and the codes 256 through 4095 are created in a dictionary for sequences encountered in the data as it is encoded. At each stage in compression, input bytes are gathered into a sequence until the next character would make a sequence for which there is no code yet in the dictionary. The code for the sequence (without that character) is added to the output, and a new code (for the sequence with that character) is added to the dictionary.
A **region of interest** (often abbreviated ROI), is a selected subset of samples within a dataset identified for a particular purpose. It can be used to apply a non-uniform quality distribution encoding. ROI coded with higher quality than the rest.
JPEG use the YCbCr color space. It uses the discrete cosine transform that transforms an 8x8 block of input values to a linear combination of 64 patterns. The patterns are referred to as the two-dimensional DCT basis functions, and the output values are referred to as transform coefficients