The human vision system

Light and electromagnetic spectrum

The electromagnetic spectrum can be expressed in terms of frequency, wavelength, and energy:

$$\lambda = \frac{c}{v} = light\,\frac{speed}{frequency}$$

$$E = hv = plank\ constant * frequency$$

$$c = \ 2{,}998 * \frac{10^8 m}{s}$$

$$h\ Plank\ constant$$

The spectrum of the visible light goes to 400 to 800 nm of wavelength.

A curiosity is that our eyes have a different sensitivity in base of the wavelength of the light that we had perceived: the green cones could reach the maximum sensitivity of our eyes, but we can note that there is an overlap between the sense. So, we can represent the absorption of light by the red, green and blue cones in the human eyes as a function of wavelength.

Image characterization

Images ca be characterized as a two-dimensional distribution of intensity (bright):

$$f = (x, y)$$

This is a continuous function, and when we catch the light using a sensor, we transform it in a discrete function.

We can describe the brightness of light using the luminous flux that is a quantity that tells us how much power a certain wave carries so it's a measure of the perceived power of visible light.

The unit of luminous flux are known as lumens; their value depends on light wavelength and is related on photometric measure because is related on sensitivity of our eyes.

So, we can define:

$$1\ Watt\ output\ at\ 555{,}02\ nm\ (specific\ wavelenght) => \ \Phi = 683\ Lumen$$

Some definitions:

Transmitted energy W [w x s = Joule] which corresponds to the light energy Q [Lm x s]

Power (energy flow) Φ $= \frac{dW}{dt}\ [watt]$ which corresponds to the luminous flux measured in lumens [lm] (1 Watt = 683 lumen)

Radiant intensity $\boldsymbol{I} = \frac{d\phi}{d\Omega}[w * sr - 1]$ which corresponds to the light intensity (lm/sr = candle)

A candle is equal to the luminous intensity of a source emitting at a frequency of $540*10^{12}$ Hz and the radiant intensity of 1/683 watts per steradian.

We can define two kinds of quantities:

- **Radiometric** quantities characterize the physical phenomenon (not the sensation produced by the visual system. Radiometry deals with the measurement of energy per time (power in watts) emitted by the light sources or impinging (interferire) on a particular surface. Thus, the units of all radiometric quantities are based on watts (W).

- ⇨ **Irradiance (irraggiamento):** the radiant flux (power) received by a surface per unit area
- ⇨ **Radiant exitance (emittenza):** the radiant flux emitted by a surface per unit area
- ⇨ **Radiant intensity (intensità):** the radiant flux emitted, reflected, transmitted or received per unit solid angle
- ⇨ **Radiance (radianza):** the radiant flux emitted, reflected, transmitted or received by a given surface, per unit solid angle per unit projected area

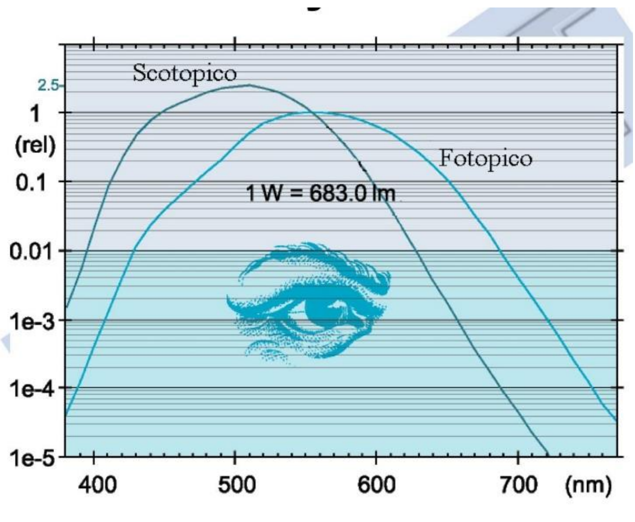
Grandezza	Definizione	Unità
Irradiance reflex Irraggiamento	$d\Phi/dS_r$ area of reflex	watt/m^2
Radiant exitance emitted Emittenza	$d\Phi/dA_s$ area of the source	watt/m^2
Radiant Intensity Intensità	$d\Phi/d\Omega$	watt/sterad
Radiance Radianza	$d\Phi/(d\Omega dS\cos\theta)$	watt/(sterad m^2)

- **Photometric** quantities relate the physical reality of the phenomenon related to the sensitivity of the human eye
  - ⇨ 1 candle (cd) light output per unit of solid angle (1/683 w per steradian at 555nm)
  - ⇨ Total flux (lumen) = cd * 4π
  - ⇨ 1 cd = 1 lumen/steradian

Quantity	Definition	Unit
Illuminance *Illuminamento*	$d\Phi/dS_r$	Lumen/m^2
Luminous emittance *Emittenza luminosa*	$d\Phi/dS_s$	Lumen/m^2 (lux)
Luminous intensity *Intensità*	$d\Phi/d\Omega$	Lumen/sterad
Luminance *Luminanza*	$d\Phi/(d\Omega\ dA_{p)}$	Lumen/(sterad m^2)

The brightness function

Radiometric measurements can be related with the photometric measurements using the luminosity function (experimental). This represents the human eye response to the varying of the wavelength.



There are two adjacent light field, one is constituted by a source of know intensity, the other from a source that emits on a slightly different wavelength. The observer varies the brightness of the second source until he feels the same intensity from the two areas. The brightness function is obtained by plotting the ratio (rapporto) of the intensities of the two light fields as a function of wavelength of the unknown source.

- Scotopic: evaluate (valutare) for low intensity light
- Photopic: evaluate for strong intensity light

An image can be characterized as the spatial distribution of radiant energy produced by a light source:

$$f = f(x, y, \lambda, t)$$

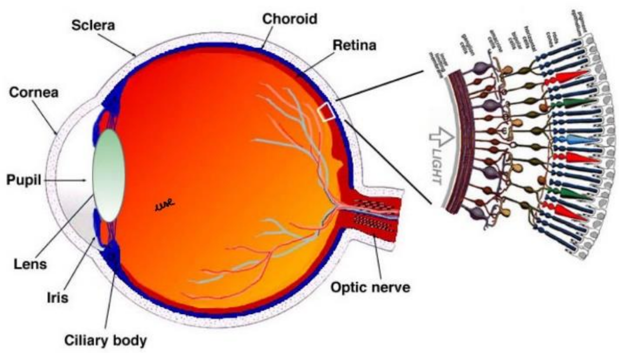
f is real, not negative, defined and limited (spatial and temporal variables).

The image perceived (by a human observer) or acquired (by means of a transducer) is the function f modified by viewer's or transducer response, typically according to time and wavelengths averages.

Human eye structure

The eye is made up of three coats:

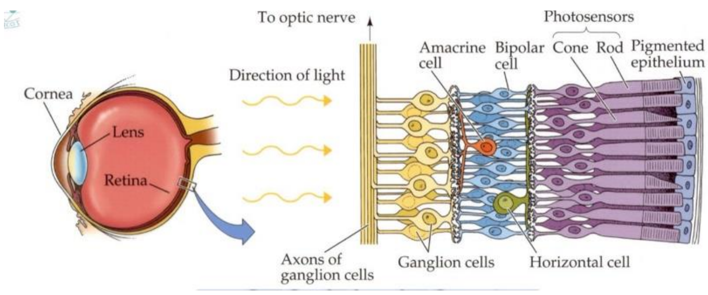
1) The outermost layer, is composed of the cornea and sclera (very pigmented, to reflect the light, and resilient)
2) The middle layer, consist of choroid (dark membrane to avoid the light), ciliary body (are connected by ligaments to the lens so they can change the curvature of it) and iris (reduce or increase the amount of light)
3) The innermost is the retina (made by sensors that are not uniform distributed, there is a point in which the concentration is elevate, over this specific point the intensity of sensor decrease, utile an area in which there aren't sensors that is call blind spot area). The retina is composed by a pigment layer and the sensor layer



Light falling on the light sensitive cells of the retina is converted into electrical signals that are carried to the brain by optic nerves.

The sensor layer is made by different photoreceptive cells:

- Image-forming receptors: rods, sensible to low light so they are responsible for the night vision or scotopic, and cones (green, blue and red), sensible to hight and medium light so they are responsible for the daytime vision or photopic.
  They compose the sensor layer, and they produce only information about the quantity of light that the eye receives. Cone and rod are related to image-forming visual information. Cones are mainly concentrated in the central portion of the retina, called the macula, and are highly sensitive to colours. The rods, in a much greater number are distributed more evenly over the entire surface of the retina and are practically insensitive to colour. Distribution over a wider area and the fact that several rods are connected to a single nerve fibre decreases the detail resolution capabilities, so rods give a global image and less accurate.
- Small subset of retinal ganglion cells in mammals is also intrinsically photosensitive



The main function of the inner retina photoreceptors is to generate and transmit non-image-forming visual information. The distribution of receptors on the surface of the retina is radially symmetric relative to the central point of the fovea (so we have in this area the maximum resolution of an image): they decrease near the optic nerve in where there is the blind spot. Thanks to the continuous movement of the eyeball during viewing, avoids the lack of perception in this area.

The visual perception of intensely blue objects is less distinct than the perception of objects of red and green. Blue cones are outside of the fovea where the close-packed cones give the greatest resolution.

The refractive index for blue light is enough different from red and green that when they are in focus, the blue is significantly out of focus (chromatic aberration).

Pupillary light reflex (PLR) is a reflex that controls the diameter of the pupil, in response to the intensity (luminance) of light that falls on the retinal ganglion cells of the retina in the back of the eye, assisting adaptation of vision to various levels of lightness/darkness.

When an object or scene are focused by the lens, the optical image is projected on the retina, allowing light sensitive receptors to absorb the energy of electromagnetic radiation and convert it into electrochemical signals that are conveyed to the brain through nerve fibres which form the optic nerve.

Light stimulates receptors, which send information to the ganglion cells; each ganglion cell responds to direct light to a specific area of the retina, called the cell's receptive fields. Ganglion cells respond only weakly to uniform illumination, they give information more on contrast rather than on absolute intensity.

The information required to detect object is contented mainly in the variations in light intensity through the visual field. The retina measures the difference and transmit a signal proportional to it (not directly signal but a sum of all information by sensors in order to reduce errors). In the fovea region the light rays arrive directly on the photoreceptors.

Space-variant image sensors

In human, central foveal vision has higher acuity compared to the periphery: the pattern of sensors is very irregular and the information that we have to processing is only from the fovea. It is possible thanks to the continuously movement of the eye.

This space-variant scheme enables a large field of view, while allowing efficient visual processing: eye movements allow to resolve fine details if necessary. But not all mammals possess a fovea because the ganglion cells distribution depends on the symmetry of the perceived world so is correlated with species behaviour and habitat.

So we can define three different kind of focus vision distribution:

1. Visual streaks => are elongated regions of higher cell density and can be horizontal or vertical.
   - ⇨ The horizontal streak provides a panoramic view of the environment without the need for a high degree of eye movement.
2. Area centralis
3. Fovea

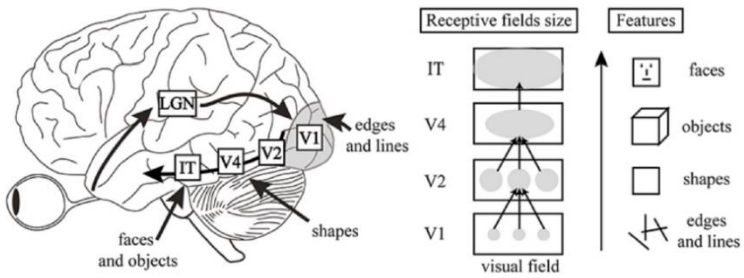
Visual Pathway and Visual cortex Areas

Ganglion cell neurons carry the visual signals generated by the retina to the optic nerves in the back of the eyes. After exiting eyes, the neurons carry the visual signals to the primary visual cortex (DX eye gives the signal to the right visual cortex, and SX eye gives the signal to the left visual cortex).

Visual cortex, based on the structural and the function characteristics it is divided into different areas:

1. Primary visual area => responds only to specific simple signals
2. Visual association area
3. Higher visual association area

To the primary to the higher area, the complexity of information that they can processing is getting more:



We can also say that there is a topographical map of cortex that mapping each element that our eyes see, the visual.

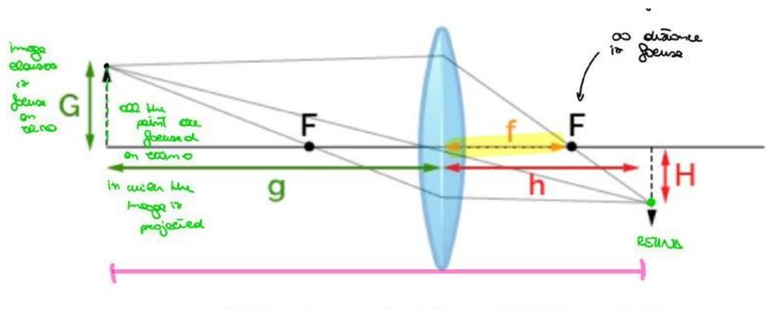
We can define a hierarchical information processing:

- Primary Visual Cortex or V_1: It preserves spatial location of visual information i.e. orientation of edges and lines. It is the first one to receive the signals form what eyes have captured.
- Secondary Visual Cortex or V_2: Its function is to collects spatial frequency, size, colour and shape of the object.
- Third Visual Cortex or V_3: It receives inputs from V_2. It helps in processing global motion and gives complete visual representation.
- V_4: It also receives inputs from V_2. It recognizes simple geometric shapes and also forms recognition of object. It is not tuned for complex objects as Human Faces.
- Middle Temporal (MT)Visual Area or V_5: It is used to detect speed and direction of moving visual object i.e. motion perception. It also detects motion of complex visual features. It receives direct connections from V_1.
- Dorsomedial (DM) Area or V_6: used to detect wide field and self-motion stimulation. Like V_5 it also receives direct connections from V_1. It has extremely sharp selection of the orientation of visual contours

Image forming by human eye

The lens of the eye has a greater radius of curvature in the anterior surface than in the posterior one. It is flexible and its shape is controlled by the fibres of the ciliary body (flat lens far object, object near thick lens). The distance between the centre of the lens and the retina (focal length) varies from 17mm to 14mm.

Focal length and image magnification



f= focal length

d=g+h= focus distance

h= the distance between the real image and the lens

g= the distance between the object and the lens

When an object distance is infinity, or when is close to the lens, is always focused on retina, where the image is projected.

The magnification of a lens is an absolute measure of how much the height of a real image differs from the object's height so we can calculate as:

$$Magnification = h/g$$

The optical centre of a lens is a point inside the lens on the principal axis: a ray of light passes through without any change in is direction. For a fixed sensor size, decreasing the focal length, increases the Field of View FOV. In the case we want to see the same image that we see using a large sensor with a larger focal length, we have to use a smaller sensor with a shorter focal length.

If we use different sensor by using the same lens, this condition doesn't change the magnification: if we use a full frame sensor we obtain a large image; if we use a cropped sensor with the same number of pixel (of the full one) we can obtain better resolution, and it seems that the image is larger but isn't.

Retinal Sampling

When we observe a line constituted by dots, we need in order to distinguish the point from each other, 3 sensors, 2 for point and 1 for understand the distance between the two points.

The distribution of sensor assume an hexagonal pattern: cone density decreases with retinal eccentricity, higher densities along the horizontal as opposed to the vertical meridians.

When we are sampling an image, we can be occurred on an Aliasing effect when the signal frequency exceeds its Nyquist limit: the frequency of sensors is to low respect to the one of input image:

$$f_c \geq 2f_{S\_max}$$

The result image appears to have lower frequency and a different orientation.

Visual acuity is a measure of the spatial resolution of the visual processing system: the density of cones in a human eye implies a Nyquist limit for the foveal centre od almost exactly 60 cycles/degree. The critical assumption of Shannon's theorem is actually the spatial regularity of the sampling array.

Spatial regularity in the human foveal cone mosaic is preserved only over distances on the order of one-tenth of a degree. Visual resolution decreases rapidly outside the foveal centre:

- resolution is worse than the cone spacing predicts and better matches the sampling limit of retinal ganglion cells. (in the centre one sensor is directly connected to a ganglion, instead outside more sensors are connected to one ganglion).
- Cones do not form a perfect spatial lattice, so the images are recovery from irregular samples

There are two sampling stages:

1) Rods and cones transduce light into voltage and electrochemical signals

2) Retinal ganglions cells sample the output of the photoreceptors

There are about 130 million photoreceptors tiling the retina, and this information is summed into only one million ganglion cells.

Eye movements

The eyes are never completely at rest: they make frequent fixational eye movements even when fixated at one point. This movement can be three different kinds:

- saccades: rapid ballistic eye movements that change the point of foveal fixation to scan a visual scene and they range in amplitude based on the image that we observed
- vergence movements align the fovea of each eye with targets located at different distances from the observer; they involve either a convergence or divergence of the lines of sight of each eye to see an object that is nearer or farther away
- vestibulo-ocular movements stabilize the eyes relative to the external world, thus compensating for head movements
- smooth pursuit eye movements: slow tracking movements of the eyes designed to keep a moving object aligned with the fovea

Purkinje effect

Purkinije noticed that his favourite flowers appeared bright red on a sunny afternoon, while at dawn they looked very dark. This is due the tendency for the peak luminance sensitivity of the human eye to shift toward the blue end of the colour spectrum at low illumination levels as part of dark adaptation The effect occurs because the colour-sensitive cones in the retina are most sensitive to orange yellow light, whereas the rods, which are more light-sensitive, respond best to green-blue light.

Colour Vision

The colours that an individual can see are defined by the number of different colour receptor in his retina. The response of our visual system is effectuated by the presence of only three types of cones, it is not possible to analyse the spectrum in detail. Colour is not a property of electromagnetic radiation but a feature of visual perception by an observer. Ther is an arbitrary mapping between wavelengths of light in the visual spectrum and human experiences of colours.

Metamerism

The metamerism is a phenomenon that happens when two colours seem to match under an artificial lighting: changing the composition of light, change also the perceived colour.

Metamerism occurs because each type of cone responds to the cumulative energy from a broad range of wavelengths, so that different combinations of light across all wavelengths can produce an equivalent receptor response.

Several different combinations of wavelengths of the spectrum of the source light can produce apparently different colour respectively the real one.

For each crossing point in the graph of the sensitivity function, the cone target has the zone reflection but for more of 3 crossing points we have metamerism.

Colour constancy

The apparently changing of colour can also happen thanks to a brain process, in this case the phenomenon is called colour constancy. A feature of the human colour perception system which ensures that the perceived colour of objects remains relatively constant under varying illumination conditions. Objects tend to always appear the same colour even though the lighting conditions change.

This provides a much easier recognition of the colour combination in order to recognize immediately an object. According to the Land's Retinex (retina and cortex) theory, the eye and the brain are both involved in the colour vision: research suggests that colour constancy is related to changing in retinal cells as well as cortical areas related to vision.

So, the Land's Retinex theory says:

"the visual system includes **three separate eye-brain systems called retinexes**, each with a peak sensitivity to long-wave light, medium-wave light, or short-wave light and inhibitory effects on the other systems. **These assign a colour** to each spot in the visual field **according to the ratio, for each of the three retinexes, of light reflected from that spot to the average of the light reflected from its surround.** The resulting triplet of ratios uniquely defining the colour at each spot."

Colour constancy also depends on:

- **Spatial contrast**: The Retinex compares the 'red' light reflected from a surface with the spatial average of 'red' light reflected from surrounding surfaces, and then does the same for the green and blue
- **Colour memory**: If you recognise a particular object as a banana, you may remember its typical yellow colour, and use any deviations from this memory colour to determine the illumination colour and then correct the colours of other objects

To mateine the colour constancy is necessary to have at least two different coloured surfaces available within the observed scene. The visual system can evaluate the reflectance ratios between the various colours. Given two objects, their spectral ratio remains constant even when light change.

To evaluate the reflection of the light we can use two different instruments:

- **Spectroradiometer:** provides the reflectance spectrum of the observed surface. So we can take a signature of the reflex light of the spectrum that a specific object reflect.
- **Colorimeter:** reflected light is analysed by three different photocells sensitive to short, medium and long wavelengths. It does not detect metamerism.

Dyschromatopsia (color blindness)

Is a condition in which the ability to perceive colours is not fully normal. There are 3 different types:

- *Protanopia*: there are no working red cone cells.
- *Deuteranopia*: there are no working green cone cells
- *Tritanopia*: lack blue cone cells

Daltonism, the total confusion between red and green is the most common form of colour blindness.

Subjective brightness

The range of intensity levels at which the eye can adapt is enormous, of the order of 10^{10}, from the scotopic threshold to the limit of glare (where there is saturation). The subjective brightness is a logarithmic function of the intensity of the light incident on the eye. It is important to note that the eye does not work simultaneously (we can just see one of the two area correctly) on the full range of levels. Rather, the system goes through a series of levels of adaptation to the intensity of light.

If the eye is, under certain environmental conditions, at the level of adaptation $\boldsymbol{B_a}$, the range of the levels that it can discriminate is only that indicated by the short curve, which has $\boldsymbol{B_b}$ as the limit below which no stimulus produces a different sensation from black. The dashed portion indicates that at levels higher than $\boldsymbol{B_a}$ the eye actually moves to a higher level of adaptation. **The ability of the eye to discriminate between changes in brightness at different levels of adaptation is very important.** A useful quantity, in quantitative terms, is the so-called **Weber ratio.**

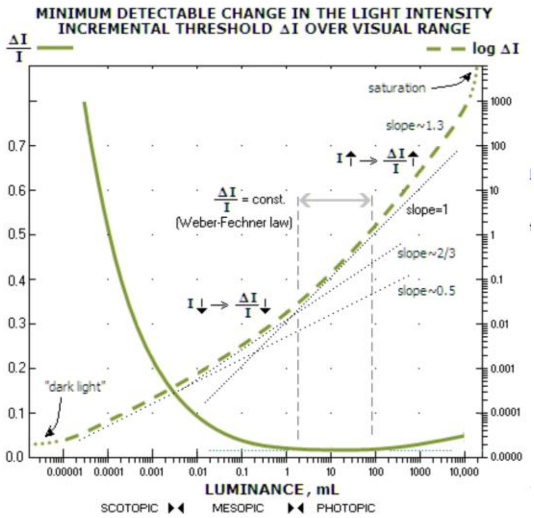
The weber ratio it’s an experiment to determine the ability of the eye to discriminate levels of brightness: At the centre of a uniformly illuminated area (intensity I) that covers the entire visual area, an intensity increment (ΔI, can increase or decrease the intensity) is applied in the form of a short-term circular spot. ΔI is increased until it is perceived 50% of the time by the observer. (it means that we can recognize this level of intensity). So, I is the intensity to which the eye is adapted and ΔI is the increase in that intensity.

A small value of ΔI / I mean that a small percentage variation of intensity is discriminable (therefore a good ability to discriminate variations) Differently, high values of Weber's report indicate the need for strong variations in brightness, so that the variations themselves are noticeable.

The discrimination capacity improves as the lighting level increases (from rods to cones). So. Woth lower intensity our sensitivity is less able to distinguish the variation of intensity.

The perceived brightness is not just a function of intensity **The human visual system tends to "get confused" at the border between zones of different intensity** Although in strips the intensity is constant, the sensation is of a pattern of variable luminosity, particularly at the border between the stripes.

- **Corn sweet Effect** => if we observe two block that are of the same colour but they are separated with a central line of the same colour but darker, we seem that the two blocks have different shade of colour, one darker and the other brighter. We can note that we have a lower perceived in left side and a higher perceived in right side. This is a psychological effect.
- **Simultaneous contrast** => our perceived target colour also depending on his background, so the context contributes to determining the brightness of a region
- **Lateral inhibition =>**The most active cells "switch off" the sensitivity of the neighbouring ones, making them less active. The result is that the signal of one side of the border between black and white is amplified, while the one on the other side decreases, creating a greater contrast than what actually exists. Your retina is partially composed of many small nerves, which function as receptors of light. It was discovered that if you illuminate a single receptor (A) you will get a large response; however, when you add illumination to A's neighbours, the response in A decrease. In other words, the illumination of the receptors "inhibits" the activation of nearby receptors.



- **Colour contrast =>** the colour depends on the context: if the background is darker, the colour becomes lighter, if the background is clear, the colour becomes darker. Kirschmann formulates a theory of the colour contrast that say:
  1. The smaller the test area, the larger the effect

2. Colour contrast occurs even if there is a spatial gap between the two field, but larger the gap is, the smaller the effect is.
3. The effect is maximum when brightness contrast is absent or weak
4. The larger the size of the inducer, the larger the effect
5. The higher the saturation of the inducer, the larger the effect

Optical illusions

A visual stimulus that is perceived by the eyes and then comprehended by the brain in a way that is different from reality. A **cognitive illusion** happens when the brain perceives an object based on prior knowledge or assumptions. The brain wants to understand an image based on other images it has seen before, it creates its own version of that image.

- **Literal optical illusions** => is an image created by smaller images that are not related to the larger, overall image created. When the mind receives visual information, it will fill in details or gaps that do not actually exist. The eye and brain will choose and focus on specific objects which causes on part of the image to appear one way or the other. Depending on what the brain chooses to focus on, it can perceive two different images in one.
- **Physiological illusions** => the theory is that stimuli have individual dedicated neural paths in the outer visual cortex of an organism for the early stages of visual processing; repetitive stimulation of only few channels misleads the visual system

Troxler's fading

If we could perfectly fixate on some point in our visual field by suppressing saccadic movement, a static scene would slowly fade from view after a few seconds due to the **local neural adaptation of the rods, cones and ganglion cells** in the retina. **Any constant light stimulus will cause an individual neuron to become desensitized to that stimulus**, and hence reduce the strength of its signal to the brain.

Micro saccades cause the pattern of activity which forms the retinal image to shift across hundreds of photoreceptors at a time, providing a constant refreshing of the image.

Gestalt theory of Perception

This theory tries to describe and analyse how we perceive the world that it's depending on the stimulus that we perceive. Our brain organizes and processes the stimulus and the information in order to give an interpretation of what we see.

The Gestalt Principles are a set of laws arising from 1920s' psychology, that **describe** (but do not explain) how humans typically see objects by grouping similar elements, recognizing patterns and simplifying complex images. These principles **try** to explain when and **how our minds perceive different visual components as being part of the same group: "the whole is different from the sum of its parts".** Our brain tends always to see simple images that he knows and also tend to group information and find something in with he is familiar. So, our perception is not depending only in stimulus of our eyes but also from the brain processing.

The main idea is that when we perceive the world there are many different signals coming in at the same time, but we organize them as unitary forms or groups. **Your mind makes sense of several features as a whole.**

The ability to perceive an object is related to the ability of the nervous system to "process" and not to a banal image focused by the retina

- Phi phenomenon = an observer perceives movement in stationary objects.

The perception is organized, in an unconscious way, through processes of imitation, learning and sharing and some rules of organization of the perceived data.

Laws of Perceptual organization (leggi della forma)

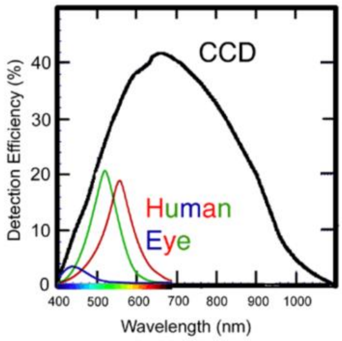
The main laws of organization of perceived data are:

1) **Simplicity** => Every stimulus is perceived in its most simple form
2) **Proximity** => object or shapes that are close to one another appear to form groups.
   We can use the Proximity principle in UI design for grouping similar information, organizing content and decluttering layouts. Its correct use will have a positive impact on visual communication and user experience. Items that are related should stay close to each other, while the unrelated items should stay further apart.
3) **Similarity** => Elements sharing similar visual characteristics are perceived to be more related than those not sharing similar characteristics. There are different ways of making elements perceived as being similar, and thus, related. These include similarity of colour, size, shape, texture, dimension, and orientation. Using elements, colors or symbols that visually link information to another helps to make a site with large amounts of content accessible and easily navigable.
4) **Common Fate (destino comune)** => lines with the same direction or orientation or movement, tend to unify according to the most consistent trend, in defence of the simplest and most balanced forms.
5) **Law of closure (chiusura)** => Things are grouped together if they seem to complete some entity. Our brains often ignore contradictory information and fill in gaps in information. People tend to identify elements first in their general outlined form. Our brain recognises a simple, well-defined object quicker than a detailed one.
6) **Law of continuity (continuità di direzione)** => Points that are connected by straight or curving lines are seen in a way that follows the smoothest path.
7) **Figure-ground** =>The “figure” is the element in focus, while the “ground” is the background behind the figure. two related principles:
   - ⇨ **Area**: The mind often perceives the smallest object in the composition as the figure, and the larger as the ground.
   - ⇨ **Convexity**: Convex elements are associated with figures more often than concave

Sensor and acquisition

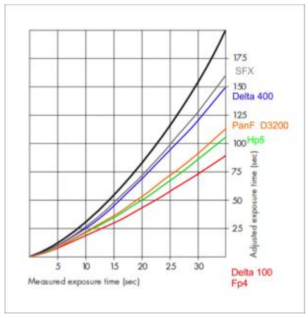
Quantum efficiency

A CCD (Charge Coupled Device) sensor transforms the photons that hit it during the exposure into electrons. The ratio between the number of electrons produced and the number of incident photons is called **quantum efficiency.**



More efficient is the device, more will be the quality of the image in dark conditions, so with less noise.

By analyse the graph of the spectral sensitivity and the quantum efficiency, we can observe that CCD can acquire also information form larger wavelength respect to human eyes sensors, it has larger sensitivity band. This is way there is always an infrared filter in front of the CCD sensor to avoid longer wavelengths.



Reciprocity failure

The quantum efficiency of the CCD remains constant regardless of the shutter speed, while that of the photographic emulsion decreases with increasing exposure time this it’s call **reciprocity failure**.

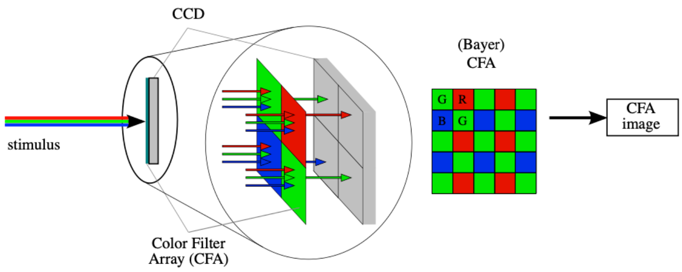
Logarithmic compensation

The eye photoreceptors have a sensitivity that varies with intensity in a logarithmic way => the perception of a very wide range of signal levels.

Instead, CCDs have a linear behaviour, sensors response on brightness with a linear function: we have to compensate, with a logarithmic compensation, in order to obtain a familiar image. Before compensation, the image is always very dark, so she needs to be compensated using a logarithmic function.

Sensor for colour image

The majority of colour cameras are equipped with a single sensor array. The surface of such a sensor is covered by a colour filter array (CFA), which consists in a mosaic of spectrally selective filters, so that each element samples only one of the three-color components Red (*R*), Green (*G*) or Blue (*B*). To estimate the colour (*R*, *G*, *B*) of each pixel in a true colour image, one has to determine the values of the two missing colour components at each pixel in the CFA image. This process is commonly referred to as CFA demosaicing, and its result as the demosaiced image: it calculates the other components of one pixel, considering all the pixels around and do the average of values for each missing component. This demosaicing technique is calling Bilinear interpolation. **Bilinear interpolation**, acceptable only in areas of uniform color, introduces many errors along the contours and in high frequency areas, this effect is calling **Zipper Effect.**

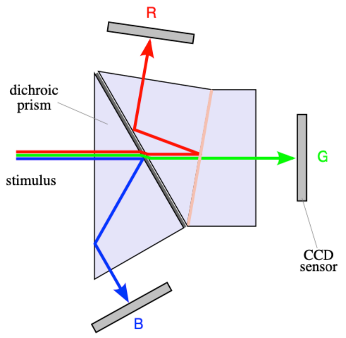


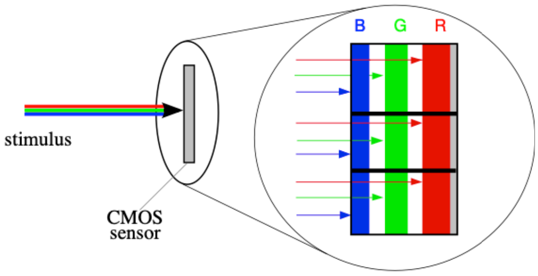
Each pixel of the matrix is a sensor that can acquire only the information of one component, this I why we need to discover the other components with the interpolation, but for sure we can have some artifact because of it. We can also observe that the number of green pixels is always more than the other this is a way to simulate the human eye perception, because we have a lot of green and red sensors on the fovea that allow us to have more sharper vision. Our eyes are more sensible to green light, so the sensor try to reproduce its behaviour.

This kind of matrix is calling Bayer Pattern.

To do acquire all the components of each pixel, in order to obtain a good quality image, there are different possibilities of sensors:

1) **Splitting by a trichroic prism** => there are 3 prism that select for each sensor only one colour component, the others are reflected. Each sensor has a specific colour that he can acquire, so they produce 3 images at the same time



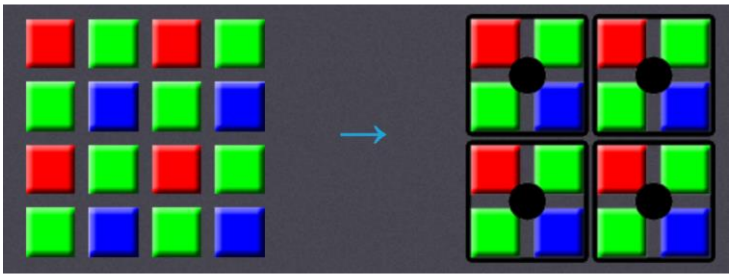


2) **Foveon X3 sensor** => the light hit the surface of a CMOS sensor that is composed by 3 different layers, each of them is sensible to a different wavelength, so we don't need the interpolation to produce a very accurate colour image

3) **Bayer based Sensor**

Demosaicing

Another technique could be to consider 2x2 array of red, green and blue that can be considered as a single full colour pixel halving the resolution in both the horizontal and vertical directions but in this way, we produce a smaller image respect the sensors size.



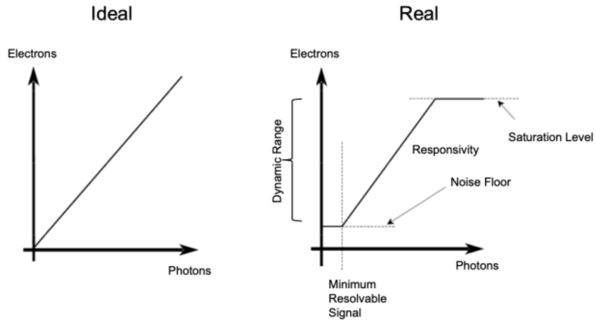
Foveon VS Bayer based sensor Each of the three

- **Foveon** layers responds to the whole visible light spectrum albeit non-linearly. The "RGB" designations refer to the point of maximum QE for each layer. There is no colour filter, Foveon requires quite severe raw-to-RGB conversion algorithms (slow and noise).
  Foveon stacked layers are not transparent and lead to a light loss. To compensate for this drawback, Foveon sensor pixel is larger than pixel in high end Bayer sensors.
  Sensor requires a good exposure because with under or over exposure we obtain poor results. The quality is more "accurate" simply because the Foveon conversion to RGB does not involve interpolation "demosiacing". There is no AA filter over the sensor because there is no colour aliasing. However, there is still **luminance aliasing** that could be quite visible, if well-focused with a sharp lens. *Foveon X3 Quattro* sensor has a ratio of 01:01:04 between the number of pixels for red, green and blue, respectively. The top layer captures brightness and colour information (luminance and chrominance), while the last two layers acquire only colour information.
- **Bayer** type sensor is inherently more sensitive to light for all colours. This means that these cameras will be more sensitive to light and inherently less noisy. Sensors have a single sensitivity; higher ISO levels are created by amplifying the signal. A noisy signal can't be amplified as much as a signal that is cleaner, so this limits the effective maximum ISO (to amplify the sensitivity) that can be used.

CCD sensor noise

Characteristics of ideal Image Sensors

- Infinite number of pixels
- No sensor noise
- Infinite dynamic range
- High (spectral) responsivity
- Infinite high frame rate
- Zero energy consumption
- Low manufacturing cost



In the ideal case, electrons and photons are linear depending; in the real case, we have a minimum resolvable signal, before of it we don't acquire nothing, a saturation level that increase with the surface of the sensor (smallest pixel, lower saturation, small sensor), noise floor that is due to the fact that our camera product electrons even when it's exposed to light.

How does Noise effect Dynamic Range? => ratio of maximum signal to minimum detectable signal

What is the maximum signal? => depending on Full-well capacity (electrons) and Saturation signal (mV)

What is the minimum signal? => depending on Noise floor electrons, Noise equivalent illumination, Minimum illumination.

For CCD sensors, and image sensors in general, is considered as noise everything which does not correspond to the image Many and various origins: conversion of the charge into voltage, thermal effects, overflow of the wells on the neighbours, etc.

Perception of Noise

Perception based differentiation of noise in image sensors:

- Random noise: Location and time of occurrence not predictable, usually are not too bad because our eye can average the noise if it isn't too hight
- Fixed pattern noise (FPN): decrease the quality of the image

Perception in still imaging:

- Underexposed images contain more visible noise due to proximity to noise floor
- Random noise averaged by eye
- Stationary noise more noticeable
- Random noise of 5% RMS barely visible
- Noise less noticeable at high temporal frequency
- Noise less noticeable at high spatial frequency

1) Fixed Pattern Noise

FPN (also called nonuniformity) is the spatial variation in pixel output values under uniform illumination due to device and interconnect parameter variations (mismatches) across the sensor. It is fixed for a given sensor but varies from sensor to sensor.

Consists of

- **DSNU - Dark Signal Non-Uniformity:** offset from the average across the imaging array at a particular setting (temperature, integration time) but no external illumination.
  Is seen as an offset between pixels in dark and can be corrected by subtracting a dark frame (the Dark Frame should have the same exposure time, integration time as the Raw Image).
  Measured in the absence of light. Is defined as the peak-to-peak difference between the minimum and maximum measured values for all active pixels in the array (measured with the sensor in darkness).
- **PRNU - Pixel Response Non-Uniformity**: ratio between optical power on a pixel versus the electrical signal.
  Deviations between pixels with a fixed gain applied. This is caused by the fact that individual pixels have different sensitivity curves. A responsivity variation between pixels under illumination. Corrected by offset and gain for each pixel. Increases with illumination but causes more degradation in image quality at low illumination.
  Flat Field Correction (FFC) is a technique for reducing the effects of sensor pattern noise (FPN Fixed Pattern Noise, and PRNU – Photo-Response Non-Uniformity) and lighting non-uniformities.
  - ⇨ the differences of light sensitivity between the pixel sensors of a camera (each sensor can have maybe a different sensitivity layer)
  - ⇨ the differences in the transmission of light through the lens (for instance: vignetting); usually the corner of an image is darker than the centre of it.

  This correction is achieved by applying the following operation to each pixel of the raw image:
  CorrectedPixel = (RawPixel - Offset) * Gain where both Offset and Gain coefficients are specific values for each pixel evaluated by a calibration procedure:

  1. **Dark Image Acquisition:** acquire several dark images and compute the average response for each pixel of the array. The **Offset** coefficient for each pixel, is actually the pixel value of the Dark Image (dark-frame subtraction).
  2. **Flat Image Acquisition:** Place a uniform target wide enough to cover the whole field-of-view Adjust the lens aperture and the illumination intensity to obtain the brightest possible image providing that no pixels are saturated. Acquire several images and compute the average response for each pixel of the array
  3. **Gain coefficients calculation:**

$$\overbrace{average(Flat)}^{\text{average value for each flat pixel}} - \overbrace{average(Dark)}^{\text{average value for each dark pixel}} = \underbrace{(FlatPixel - DarkPixel)}_{\text{we want they are equal}} * Gain$$

obtain Gain by inverted formula

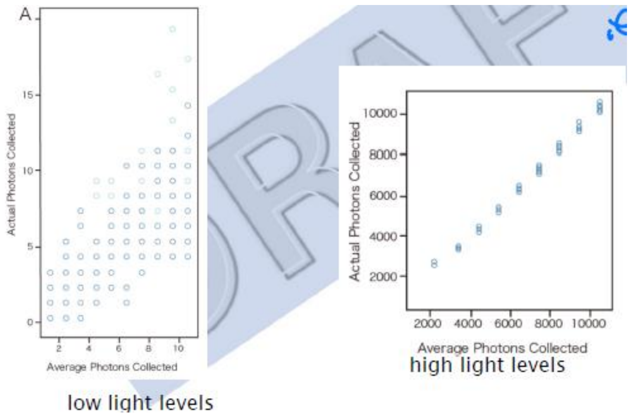
$$\textbf{Gain} = (average(Flat) - average(Dark)) / (FlatPixel - DarkPixel)$$

FPN is quantified by the standard deviation of the spatial variation in pixel outputs under uniform illumination. It is typically reported as a % of voltage swing (or well capacity) Experimentally, FPN is measured as follows:

1) Set a constant uniform illumination level (including no illumination)
2) Take many images
3) For each pixel compute the average output value (to average out temporal noise)
4) Estimate the standard deviation of the average pixel values
5) Repeat the procedure for several uniform illumination levels

2) Shot Noise (photon noise)

A source for sensor noise is the light itself. The detection of photons by the CCD is a statistical process. If images are taken over several (equal) time periods, then the intensity (the number of photons recorded) will not be the same for each image but will vary slightly. The deviation in intensity found for each image follows the Poisson distribution (the square root of the signal intensity measured). The number of photons that hit the sensor have a Poisson distribution and increasing the number of photons, the shot noise is less visible:



This noise doesn't depend on camera but only on light.

the uncertainty in the number of photons collected during a given period of time is simply:

$$\sigma_{shot} = \sqrt{S}$$

where σ_{shot}, is the shot noise and S is the signal, both expressed in electrons.

3) Dark current, thermal current

The CCD output is proportional to the exposure, *ER* x T_{int} (exposure time/ integration time).

The output signal can be enhanced by increasing the integration time, this permit to see dark object, but increase also dark current, thermal current and noise. Long integration times are generally used for low light level applications. This approach is limited by the generation of dark current, which is integrated such as photocurrent.

CCDs spontaneously produces and accumulates electrons, even when their surface is shielded by incident light. Dark current is the result of imperfections or impurities in the depleted bulk silicon or at the silicon-silicon dioxide interface that provide a path for valence electrons to sneak into the conduction band, adding to the signal measured in the pixel. The generation of dark current is a thermal process wherein electrons use thermal energy to hop to an intermediate state, from which they are emitted into the conduction band.

The Dark Current produced depends on the temperature of the sensor: it generally decreases by a factor of 2 for every 6 ° C The simplest way to reduce the dark current is to cool the CCD as dark current generation is temperature related. Reduce temperature, reduce the thermal current.

The dark current is perfectly reproducible. In identical conditions of temperature and duration of an exposure, a given sensor always generates the same number of electrons (less than a statistical dispersion factor, Thermal Noise). Dark noise is statistical variation in the number of electrons

$$Dark\ noise = \sqrt{(dark\ current)\ (integration\ time)}$$

thermally generated within the pixel in a photon-independent fashion It is the electron equivalent of photon shot noise.

The Subtraction of the Dark frame permit to remove only the average value of the dark noise: camera acquires a second photograph right after the exposure with similar conditions (ISO, shutter speed) without light exposure so it can detect noisy pixels and adjust them.

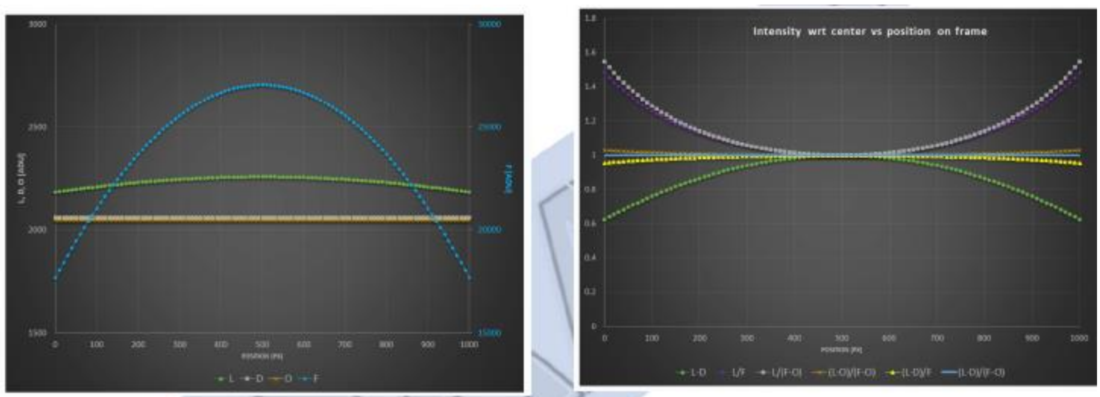
4) Readout noise

Read noise is created within the camera electronics during the readout process as the electrons are subjected to the analogue to digital conversion, amplification and processing steps that enable an image to be produced. Readout noise is the accuracy with which we can read the number of photons that were captured. This is a constant number, which depends on the sensor Read-out noise increases in proportion to read-out speed.

So, this noise is a characteristic of our sensor, a constant value for each sensor and increasing so much ISO cause the noise grow up to hight to see something in the image.

Sensor calibration

Bias: if objects are often very weak, their signal at the end of the exposure could still be overwhelmed by electronic noise and thus lead to a truncation of the result by the A/D converter (zero sum between signal and noise). Adding offset to the signal before conversion, we obtain a number that will always exceed that of the electronic noise. The offset value must be high enough to produce the desired effect (avoid data truncation), but not excessive as it would actually reduce the dynamic range of our camera



https://siril.readthedocs.io/en/stable/preprocessing/calibration.html

L for Lights, **D** for Darks, **F** for Flats and **O** for Bias.

$$L_c = \frac{L - D}{F - O}.$$

Light image corrected

Image stacking

Taking photos of the starry sky (fixed mount), if the exposure time is short there are no star trails:

- increase the sensor gain (ISO) increasing the noise too
- increase the exposure time with a lower ISO and thus with lower noise levels, but the sky will be smeared out due to the rotation of the Earth.

The rule of thumb is that the maximum exposure time for full frame cameras should be "600 / focal length". With this technique we acquire more noise but the signal increase to hight level.

Stacking is a term for adding/averaging multiple images together to reduce apparent noise (improve the signal-to-noise ratio) of the combined image. The signal-to-noise ratio, or S/N, increases by the square root of the number of images in the stack.

We consider and uncorrelated noise with zero mean Gaussian distribution, if the standard deviation is higher the noise of the image increase so we can see more colour spot on the image and consequentially the quality of the image is reducing. But increasing the number of stacking frame we can obtain a signal similar to the original image without noise and after we can process the image in order to increasing the details and the sharpness of the image.

Adding images *does* increase the noise, but the signal increases more, resulting in a better image.

A particular kind of stacking is median stacking in which we take several pictures of the same scene and then we calculate the median value of pixel (the value that divide the vector of pixels in 50% over and 50% under) and so obtain the final image. In this case, if we have a moving target, it's more probable that it cannot appear in the final image.

Whenever we measure a signal, there is always noise associated with it. **That noise is equal to the square root of the signal. Averaging exposures, we reduce the noise by the square root of the total number of exposures. This higher signal-to-noise ratio will produce a much better image even though, technically, it has more noise.**

In averaging exposures, the signal stays the same in the final image, but the noise is reduced. **An image with a high signal-to-noise ratio can be more aggressively processed to bring out faint details (faint deep-sky objects).**

- ⇨ When **random (noise) signals** are added, the resulting total signal sum is the same as the **sum of the power of the signals** => **incoherent gain of a system**
- ⇨ When **identical signals** are added, the **signal levels add directly** => **coherent gain of a system**

Image stacking S/N

The signal to noise ratio quantifies how well an object like a star can be distinguished from the noise in an image

age.

$$\text{Signal to Noise Ratio (SNR)} = \frac{\overset{\text{integration time}}{I\,QE\,t}}{\sqrt{\underbrace{I\,QE\,t}_{\text{shot noise}} + N_d t + N_r^2}}$$

incoherent signal => the noise

I = Photon flux (photons/pixel/second)
QE = Quantum efficiency
t = Exposure time (seconds)
Nd = Dark current (electrons/pixel/sec)
Nr = Read noise (electrons)

$I\,QE\,t$ is the signal produced by the star

Quantum efficiency => number of electron produced

$I\,QE\,t$ is the signal produced by the star
Noise = shot noise + dark noise + read noise

shot noise = $\sqrt{I\,QE\,t}$

dark noise = $\sqrt{N_d\,t}$

read noise = N_r

summing a number (F) of shorter exposure frames to produce the same total exposure.

$$\text{Signal to Noise Ratio (SNR)} = \frac{I\,QE\,t}{\sqrt{I\,QE\,t + N_d t + N_r^2}}$$

I = Photon flux (photons/pixel/second)
QE = Quantum efficiency
t = Exposure time (seconds)
Nd = Dark current (electrons/pixel/sec)
Nr = Read noise (electrons)

$$\text{signal} = \frac{I\,QE\,t}{F}$$

$$\text{shot noise term} = \frac{I\,QE\,t}{F}$$

$$\text{dark noise term} = \frac{N_d\,t}{F}$$

readout noise (unchanged) = N_r^2

noise component of a single frame

The total single frame noise therefore becomes $\sqrt{\left(\frac{I\,QE\,t}{F}\right) + \left(\frac{N_d\,t}{F}\right) + N_r^2}$

The SNR (short exposure frame) = $\dfrac{\frac{I\,QE\,t}{F}}{\sqrt{\left(\frac{I\,QE\,t}{F}\right) + \left(\frac{N_d\,t}{F}\right) + N_r^2}}$

If we now sum F short exposure frames:
Sum signal = $I\,QE\,t$ (the same as for a single long exposure)
To calculate the total noise we sum F of each term inside the square root

The SNR (stack) = $\dfrac{I\,QE\,t}{\sqrt{I\,QE\,t + N_d\,t + F N_r^2}}$

Dynamic range

In a scene the dynamic range is the ratio between the maximum value and the minimum brightness value. May be very large: great variation of the values of illuminance for incident light (scene with direct and indirect light), or there are string variations in reflectance.

In a photo the dynamic range is the difference between the maximum and minimum value of the brightness that can be acquired (the difference between the darkest and lightest tones). All values outside this range are not acquired correctly. Every device that we can use has its own dynamic range.

Dynamic range is measured in "stops". An increase of one stop equals a doubling of the brightness level. The **human eye** can perceive about **20 stops** of dynamic range (ideal circumstances). This means that the darkest tones we can perceive at any time are about 1,000,000 times darker than the brightest ones in the same scene. Modern **digital cameras** can achieve just under **15 stops** of dynamic range.

For values outside the dynamic range: saturation for the highest values (all assume the same value representing the maximum) and **suppression** of the lowest values (all assume the value 0).

The different dynamic ranges of the devices can greatly influence the representation of the same scene.

For every amplification gain, intensity below the minimum value is not acquired. A possible solution: increase the exposure time. This reduces the minimum value that can be acquired but increases the saturation for the highest values.

Due to the limited dynamic range of the sensors, it is not possible to reproduce all the brightness levels of a scene with a single photo.

We have to choose correctly the integration time; the saturation level mainly depends on the sensors size: larger sensor is better than smaller one because the bigger has also bigger pixels (even if they have the same resolution) that can acquire higher number of photons at the same time respect to the number of photons that can acquire a smaller one.

In CCD the intensity of a pixel is a function of the number of photons collected by the sensor. Representing the sensor as a bucket, this can collect water up to the edge (saturation). Further poured water is not collected but overflows in adjacent areas. This is called **Blooming effect**.

Larger sensors can collect more photons before filling (**saturation**), so they have a larger dynamic range. **Image quality:** the best sensitivity for a sensor is the one that involves the lowest amplification level (minimum). **Extended low ISO:** the minimum sensitivity is halved by overexposing by a factor of 2, the noise is minimized but lost in dynamics.

The range of signal amplitudes which the ADC can resolve.

- Ideal case: the ratio between the maximum number of photons that can be measured by a sensor and the minimum detectable value (1 photon), defines the dynamic range of the sensor.
- Real case: ratio between the maximum measurable intensity value (saturation) and the minimum measurable intensity value (above the reading error, photon noise)
- Stops: an increase of one stop equals a doubling of the brightness level (1024=2^10 ==> 10 stop)

$$DR_{CCD} = \frac{\text{full well capacity}}{\text{rms noise}_{\text{dark}}}$$

⇒ no more photons can be collect by the sensor

$$DR_{CCD} = 20 \cdot \log\left(\frac{\text{full well capacity}}{\text{rms noise}_{\text{dark}}}\right) [\text{dB}]$$

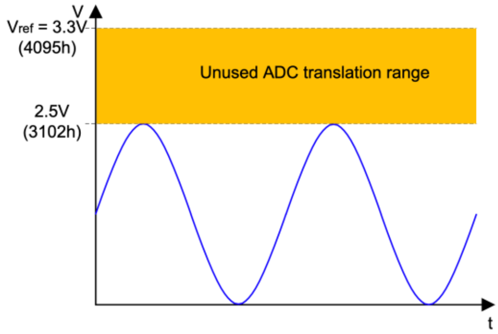
Averaging exposure, we reduce the noise by the square root of the total number of exposures. Averaging N images the increase in dynamic range is of:

$$20 * \log \sqrt{N} = 10 * \log N$$

Combining the homologous pixels of N images and quantizing the values with a number of levels higher than the dynamic range of the sensor, we obtain an image with a larger dynamic range.

Analog to digital converter S/N

To have the maximum ADC conversion precision, it is preferable that the ADC dynamic range matches the maximum amplitude of the signal.



A common definition of the dynamic range is expressed by the following formula:

$$DR = 20 * \log_{10} \frac{Maximum\ Voltage}{Minimum\ Voltage}$$

Higher the number of bits that we have to convert, higher the dynamic range that can be converted.

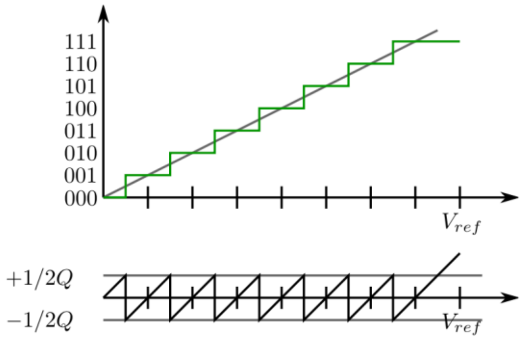
The A/D conversion operation associates a single binary code to a range of input voltage values. In the inverse operation, each code is associated with a voltage value which is the central value of each interval. The maximum amplitude of the error is equal to half the voltage range:

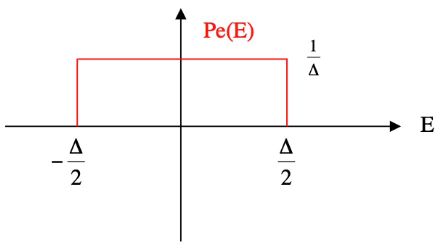
$$E = \frac{1}{2} * \left(\frac{Vmax}{2^N}\right) = \frac{Vmax}{2^{N+1}}$$

When an ADC converts a continuous signal into a discrete digital representation, there is a range of input values that produces the same output (quantum Q) The difference between input and output is called the quantization error (between ±Q/2).

Increasing the quantization level, the error decrease.

The quantization error has a uniform distribution from -Q/2 to +Q/2





it can be considered a quantization noise with RMS:

$$v_{qn} = \sqrt{\frac{1}{Q}\int_{-Q/2}^{+Q/2} x^2 dx} = \sqrt{\frac{1}{Q}\left[\frac{x^3}{3}\right]_{-Q/2}^{+Q/2}} = \sqrt{\frac{Q^2}{2^3 3} + \frac{Q^2}{2^3 3}} = \frac{Q}{\sqrt{12}}$$

assuming an input sinusoidal with peak-to-peak amplitude V, its RMS value is:

$$V_{rms} = \frac{V_{ref}}{2\sqrt{2}} = \frac{2^N Q}{2\sqrt{2}}.$$

to calculate the signal-noise ratio, we divide the RMS of the input signal by the RMS of the quantization noise:

$$SNR = 20\log\left(\frac{V_{rms}}{v_{qn}}\right) = 20\log\left(\frac{\frac{2^N Q}{2\sqrt{2}}}{\frac{Q}{\sqrt{12}}}\right) = 20\log\left(\frac{2^N\sqrt{12}}{2\sqrt{2}}\right)$$

$$= 20\log\left(2^N\right) + 20\log\left(\frac{\sqrt{6}}{2}\right) = 6.02N + 1.76(dB).$$

the final value is the ideal case.

To improve the dynamics of the converter we can: reduce the noise (ideal case only quantization noise) by increasing the resolution of the converter. In digital cameras, converters use more bits that those strictly necessary to represent the maximum dynamic range of the sensor.

HDR photography

Our eyes can perceive an extraordinary range of contrast in a scene, a range far greater than any camera's sensor can capture. We can see into a scene's brightly lit areas, and we can also tell what's going on in the shadows. The camera is going to have trouble capturing the ends of that drastic range. If you choose to meter for the highlights (the bright areas), you'll lose all the detail in the shadow areas. If you choose to meter for the shadows, you'll lose all the detail in the highlights.

The human eye, through adaptation of the iris and other methods, adjusts constantly to adapt to a broad range of luminance present in the environment.

Exposure Fusion

HDR photography captures and then combines several different, narrower range, exposures, keeping only the “best” parts in the multi-exposure image sequence. Think the input sequence as a stack of images. The final image is obtained by collapsing the stack using weighted blending.

Weighted Blending

Many images in the stack contain flat, colourless regions due to under and overexposure. Such regions should receive less weight, while interesting areas containing bright colours and details should be preserved.

Interesting areas: can be selected by evaluation of Contrast, Saturation and Well-exposedness

- ⇨ Contrast: to assign a high weight to important elements such as edges and texture.
- ⇨ Saturation: (saturated colours are desirable) is computed as the standard deviation within the R, G and B channel, at each pixel.
- ⇨ Well-exposedness: intensities should not be near zero (un-Der exposed) or one (overexposed). Pixel intensity is weighted based on how close it is to 0.5 using a Gauss curve $\exp\left(-\frac{(i-0.5)^2}{2\sigma^2}\right)$

Tonal Mapping: to adapt images to a display, this process calculates a compression of the high dynamic range. The final image is always in a hight dynamic range but compressed.

Photo stacking for long exposure

By long exposure it is possible to visualize scenes with much more softness and harmony respect to a standard exposure. Long exposure problem: digital noise, due to sensor overheating Solution: high quality neutral density filter and photo stacking.

Sensors CCD and CMOS

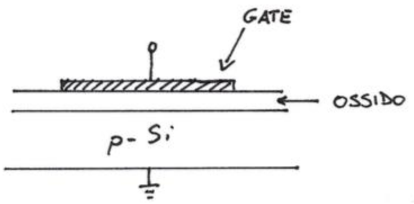
Sensors can be classified by their technologies (CCD or CMOS), or by size, noise, umber of pixels and kind of shutter time (rolling shutter time or global shutter time)

What is an Image Sensor

An image sensor is a photosensitive device that converts light signals into digital signals (colours/RGB data). Typically, the two main types in common use are CCD and CMOS sensors and are mainly used in digital cameras and other imaging devices.
CCD stands for Charged-Coupled Device and CMOS stand for Complementary Metal-Oxide-Semiconductor.

CCD

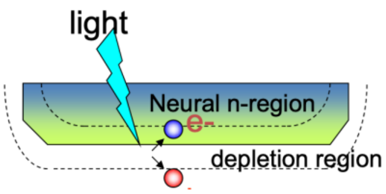
The MOS Capacitor



Overlap of three materials: the metal called gate, which forms one of the two faces of the condenser, an insulator, commonly SiO_2 placed between the two faces, and lastly a p-type silicon doped substrate (enriched with electron acceptor atoms, called bulk). if the gate voltage V_g is greater than a threshold voltage V_t the electron holes (lacune) from the surface are removed until the surface concentration becomes lower than the value relative to the internal regions of the silicon substrate (emptying layer). The positive charge accumulated on the gate is balanced by the negative charge of the ionized acceptor atoms in the depletion region.

A percentage of the photons reaching the silicon substrate is absorbed as a function of the energy of each photon E

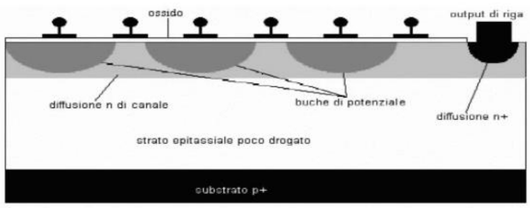
$$E = hc/l$$



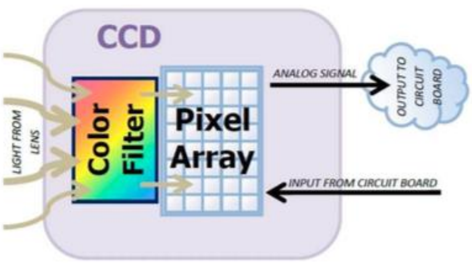
This energy is used to free the electron of an acceptor atom thus **creating an electron-hole pair.** Under the effect of the applied electric field, the hole is removed in the substrate, while **the electron moves towards the gate, accumulating under the oxide. The electrons** close to the surface **are directly related to the intensity of incident light.**

There is an accumulation of electrons under the gate that depends on the intensity of light: how much we are near to light.

The charge under each MOS gate is directly proportional to the distance of the gate from the point of incidence of the photon; A direct reading of the gate contacts is not possible, as they are shielded from the charge by a layer of insulating material (oxide). **Linear shift-registers formed by a series of MOS gates.**



CCD sensor



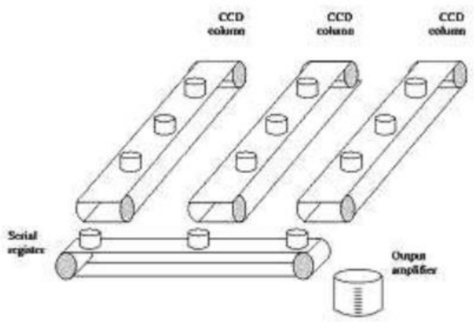
Charge Coupled Device (CCD 1969) transforms a light signal into an analog electronic signal. two fundamental parts: the **pixel matrix** and the **light filter**. The pixel matrix consists of metal-oxide-semiconductor (MOS) capacitors that under certain conditions (photoelectric effect) can release electrons by exploiting the energy of the photons of the incident light.

The voltage is amplified and transform to analogy signal to digital signal, but all these activities are usually made outside the pixel array.

The time during which the CCD sensor is exposed to light is referred to as the **integration period**. The **filter** allows on every single pixel only the passage of certain frequencies of light (red, green or blue).

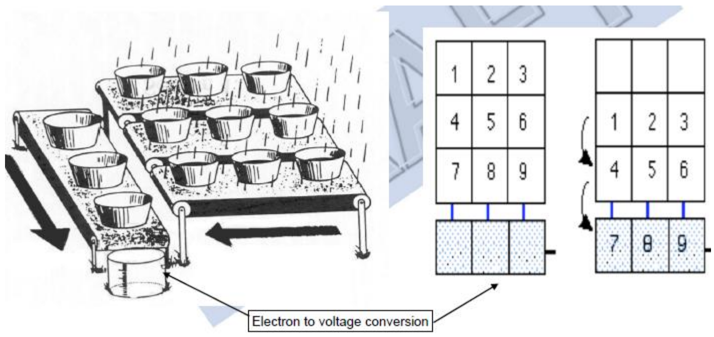
CCD reading

Each row of the matrix "empties" its contents in the next empty row, until a row used for reading is reached (a shift register made along the columns).



Each colour of the sensor is switching until the are on the reding row that it’s also a shift register and convert the RGB value into voltage.

Pixel are buckets filled by electrons (produced by photons:



During the **integration time** (exposure), **the intensity of the light signal is converted into an electrical charge**. **The transformation of the electric charge into an analog voltage signal** takes place only when it reaches the reading line (**shift register**) at the end there is an amplifier. **The signal conversion, from analog to digital is executed by an A/D converter that is generally external to the sensor.**

Full Well Capacity

The maximum number of photoelectrons that a single pixel is able to collect is called **electronic capacity**. If each individual pixel can be thought of as a **well of electrons**, then saturation refers to the condition where the well becomes filled. The amount of charge that can be accumulated in a single pixel is determined mainly by its area.

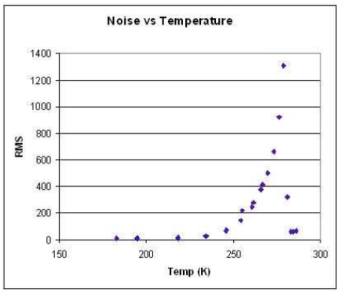
A high electronic capacity implies larger integration times to saturate pixel, with the same incident light intensity. At **saturation**, pixels lose their ability to accommodate additional charge. This additional charge will then spread into neighbouring pixels, causing them to report erroneous values or also saturate. This spread of charge to adjacent pixels is known as **blooming** and appears as a white streak or blob in the image.

Dark Current

The dark current is the main source of noise, and we cannot remove it at all: we can remove only the average value-

Statistically in the matter there are particles with **kinetic energy** (produce electrons eve if there is no light hitting the sensor) greater than the average. This also happens in the pixels of the CCD sensor.

It may happen that some electrons are able to free themselves and are accumulated despite the sensor being shielded from light or in the dark thermal type of noise, the quantity of electrons produced spontaneously is closely linked to the temperature in which the sensor operates.



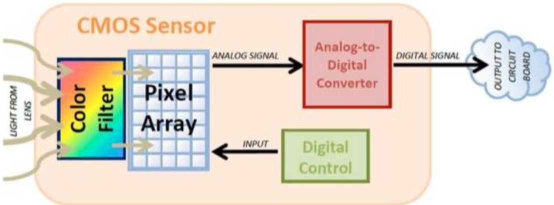
To minimize this phenomenon:

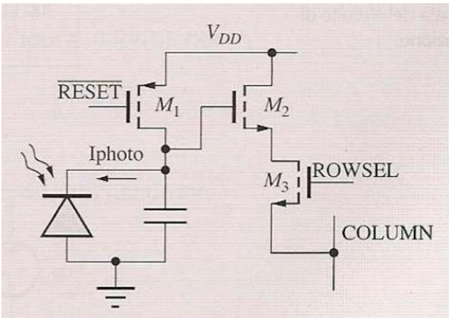
- Lower the sensor temperature.
- Subtract the dark frame from the light frame.

The thermal noise cannot be eliminated completely, because it is random.

CMOS

C-MOS, or Active Pixel Sensor, is an image sensor that, unlike the CCD, contains much of the functionality needed in every single pixel: for each pixel we have an amplifier and a converter. The sensor consists of an integrated circuit and a pixel array. Each pixel contains a light sensor and a signal amplifier. Furthermore, a digital analogue converter and a digital controller are also allocated in the same integrated circuit.





The photodiode in direct polarization behaves normally, when it is inversely polarized and is hit by light, it allows the passage of current from cathode to anode in relation to the intensity of light.

$$Vlux = VDD - CIphototesp$$

when ROWSEL is high M3, (acts as a switch), will turn on and will transmit the VS signal to the analogy digital converter.

C-MOS and noise

It may depend on external factors such as **supply voltage fluctuation**, **dirty digital control signals**, **electromagnetic disturbances** or features intrinsic to the sensor (resistive / thermal causes).

Dark Current: current that contributes to the discharge of the capacitor thus providing a data no longer truthful on the intensity of the light. Effect is much more evident when making long exposures than short ones (dark current has more time available to discharge the condenser). Increasing the exposure time, increasing the thermal effect.

Each high-performance CCD camera carries a dark current specification. Dark current noise is the statistical variation of this specification.

Hot pixels

Pixels that have a **higher-than-average dark current** are known as hot pixels. The hot pixel has a higher thermal noise value and it’s visible also in absence of light. With the aging of sensor, the number of hot pixels increases but there are different techniques used for reducing this effect.

A hot pixel has an output given by: $\boldsymbol{output} = \boldsymbol{m}(\boldsymbol{I}_{photo}\boldsymbol{t}_{int} + \boldsymbol{I}_{dark}\boldsymbol{t}_{int})$

Where:

- I_{photo} => is the current induced proportional to the intensity of illumination
- m => is the sensitivity of the single pixel
- t_{int} => is the period of integration
- I_{dark} => expresses the dark current

$$\boldsymbol{m * I_{dark} * t_{int}} \text{ is the } \textbf{dark signal} \text{ that we want to remove from the image}$$

Hot pixels even with very short exposure times are quickly saturated (they completely discharge the photodiode's capacity). The pixel is no longer able to detect further light (the information can be completely covered by the anomalous leakage current). **In the final image the hot pixels are identifiable as white or particularly bright spots.**

Hot pixels correction

dark frame: it is not very effective, if during the measurement of the Dark Frame the pixel is saturated, more importantly it will saturate during the Light Frame measurement, obtaining by subtraction in the final image a **black dot instead of white. Any saturation during the Light Frame causes all information loss on the actual light signal.**

To mask the hot pixels by software:

- identify them through a Dark Frame,
- interpolate via software with adjacent ones => this produce a fake information because in the reality there is not true value for that pixel it's only created by interpolation

This however involves an error on the information that would have been collected in that pixel.

Hardware correction: The defect that causes this phenomenon is very often **limited to a small region of the photodiode**. **Fault-tolerant APS (Active Pixel Sensor)** splitting the APS into two halves operating in parallel, if one half of the pixel is faulty, the other half can be used to recover the entire output signal. We have two equal device that corresponds only to a pixel: sum the voltage of both of them to obtain an average value.

The process consists of:

1) Pixel is divided into two sub-pixels with minimal overlapping of areas
2) During the detection of light, the two sub-pixels work independently in such a way that any defect of hot pixels remains limited to only one half of the pixel.
3) The respective outputs are added together to form a combined output.

Hardware correction + software detection and correction

$$I_{out} = f\left(Q_{photo} + Q_{dark}\right) + f(Q_{photo})$$

Where Q_{photo} *and* Q_{dark} are the charge accumulated after light exposure and the dark current $f(Q)$ it is the answer to the light exposure on a single sub-pixel.

$f(Q)$ of a good sub-pixel must be characterized at the factory and stored on the camera.

In a hot pixel:

$$I_{out} = f\left(Q_{photo} + Q_{dark}\right) + f(Q_{photo})$$

The correction process begins by identifying and locating all hot pixels.

A dark frame is captured, and hot pixels are identified as bright pixels in this image. At each identified hot-pixel:

$$I_{picture} = f(Q_{photo} + Q_{dark}) + f(Q_{photo})$$

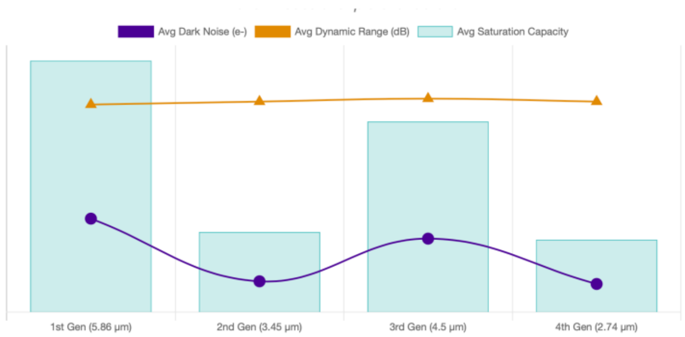
$$I_{darkframe} = f(Q_{dark})$$

The correction process proceeds as follows:

1) Build the illumination response look-up function, $f(Q)$
2) From the captured image record the output $I_{picture} = f(Q_{photo} + Q_{dark}) + f(Q_{photo})$
3) From the dark frame, record the output $I_{darkframe} = f(Q_{dark})$
4) Estimate $Q_{dark} = f^{-1}(I_{darkframe})$
5) Estimate Q_{photo} such that $I_{picture} = f(Q_{photo} + Q_{dark}) + f(Q_{photo})$ is satisfied
6) Map Q_{photo} onto the range used by the image format and store to the image file

C-MOS generations

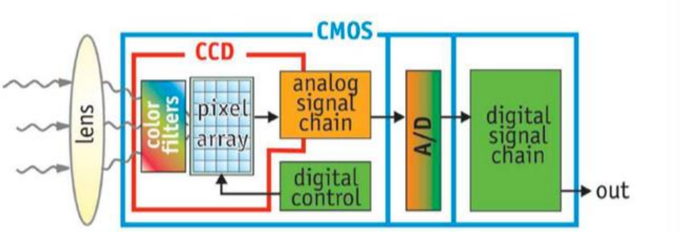
Much of the difference between generations is attributed to pixel size and pixel technology. The pixel size will not only determine how many pixels can fit into a specific sensor size, but it will also result in different image quality performance characteristics. In this example the 4th generation sensors have the smallest pixel size and therefore the lowest saturation capacity but also the lowest temporal dark noise: its dynamic range is the same of the 1st generation.



DPAF

Each pixel on the CMOS imaging sensor has two separate, light-sensitive photodiodes, which convert light into an electronic signal. Independently, each half of a pixel detects light through separate micro lenses. Each half of one pixel have a lens and acquire an image so we have 2 image at the end: if they are correctly focused they are perfectly superimposable in the other case they are not.

CCD vs CMOS

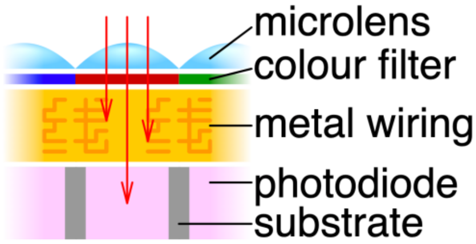


Both types of imagers convert light into electric charge and process it into electronic signals.

1) In a CCD sensor, every pixel's charge is transferred through one output nodes (bottleneck) to be converted to voltage, buffered, and sent off-chip as an analogy signal.
2) In a CMOS sensor, each pixel has its own charge-to-voltage conversion, and the sensor often also includes amplifiers, noise-correction, and digitization circuits, so that the chip outputs digital bits.
1) In CCDs all pixels have the same amplifier in common.

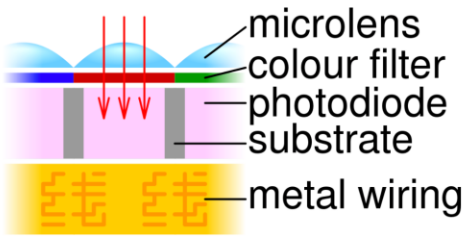
2) In the CMOS, each pixel has its own amplifier. =>A high number of amplifiers means a greater possibility of differences in gain resulting in greater inhomogeneity in the image.
solution:
- ⇨ **equalize** the amplifiers
- ⇨ accept the **noise pattern**, every time you "shoot" a photo you actually acquire two images: subject and noise pattern (closed shutter), subtracting this you get a "clean" image

Front illuminated CMOS



Because each pixel on a CMOS sensor has several transistors located next to it (charge-to-voltage conversion, amplifiers, digitization), **the light sensitivity of a CMOS chip tends to be lower**. Many of the photons hitting the chip hit the transistors instead of the photodiode. These circuits reduce the area available for light capture

Back illuminated CMOS



all the wiring and circuitry that's used to carry the electronic signals from each photo site or pixel is located at the back of the sensor instead of the front. This technique imitates the cephalopod eye, and we can obtain the 95% of quantum efficiency.

1) CCD sensors are affected by **less noise** CCDs are affected by the **blooming** effect
2) than CMOS sensors, and consequently their dynamic range is slightly wider, in CMOS the blooming effect is absent

about power consumption:

The CMOS sensor operates with quite low supply voltages and the devices are increasingly miniaturized, CCDs, on the other hand, require operating with normally higher voltages, which inevitably leads to greater dissipation of power. **CCDs consume as much as 100 times more power than an equivalent CMOS sensor** CCDs lose a lot of time in emptying the pixel registers, it is not possible to reach very high frame rates, which is instead possible in CMOS sensors thanks to the fact that each pixel autonomously processes its own voltage signal.

CCDs, have greater spectral range sensibility, are able to sustain long integration times with low noise, are the preferred sensors for electronic astronomical telescopes to acquire very low light images. CCD sensors are preferred in high performance cameras and in many scientific and industrial fields CMOS sensors (low power consumption) are mainly used in commercial applications (eg mobile devices, phones or cameras which require low power consumption).

Sensor classification by Noise

Noise is composed of two elements: fluctuations in colour and in luminance.

- **Luminance Noise** is colourless and can vary based on the size of the pixels and the ISO selection being used. Increase the ISO sensitivity of a digital cameras actually increase the signal that pixels emit amplifying the brightness of the image. Luminance noise can also be caused by long exposures because of the heat produced by the image sensor itself, “hot pixels”.

- **Chromatic noise** (chroma noise or simply colour noise), is the fluctuations of colour tone between pixels. This type of noise is very unsightly and can arguably be more of a nuisance than luminance noise. This type of noise becomes more apparent in the very dark or very light areas of digital images. We acquire 3 different noise levels for each of the RGB components, and this difference is very visible in underexposed images.

With an amplification of the ISO, we increase also the noise, and we can split it into its two components.

Noise occurs with greater intensity in the shadows area than the lights area. The noise increases exponentially when trying to recover underexposed photos. Never underexpose too much, always opt to read overexposures. Colours and tones will be more easily recovered while noise will be reduced.

We have to avoid an image underexposed: sometimes it’s better an overexposed image because we can easily improve the image with a better post processing to obtain a good quality image (always avoid the saturation of pixels).

Very long exposure times create more noise (chrominance): this happens because the sensor remains powered for all the time of exposure, overheating. If you take 100 images at a short distance from each other, all parameters being equal, the last picture will have more noise than the others due to thermal stress.

Sensor Classification by kind of shutter time

The sensor can have two kinds of shutter time:

- Global shutter time => all pixels are reading at the same time. This configuration is better to take photo of a specific instance of a moving object but it’s very expensive and usually produces more noise
- Rolling shutter time => the pixels are reading one row at time, so they are acquiring in different time, but it’s cheap and produce less noise

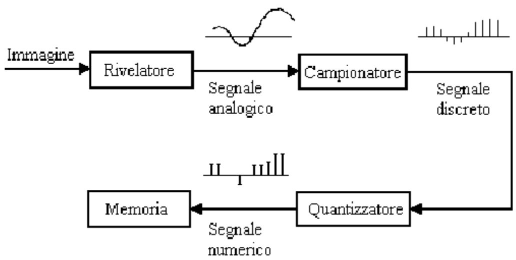
Image Acquisition

The general aim of Image Acquisition is to transform an optical image (Real World Data) into an array of numerical data (digital signal that we can process) which could be later manipulated on a computer.

The Image Acquisition process is based on:

- Energy reflected from the object of interest
- Optical system which focuses the energy
- A sensor which measures the amount of energy

The process involves into 5 phases:



1) Image sensing: we use sensor according to the nature of energy (illumination)
2) Transduction: incoming illumination energy is transformed into voltage by the combination of input energy and sensor material that is responsive to the particular energy that is being detected
3) Sampling: process of digitizing the domain
4) Quantization: process of digitizing the range
5) Storing

Image generation

Images are generated by a source of "illumination" and by the reflection or absorption of energy of the source from the elements of the "scene". The reflected or transmitted energy is converted into voltage by appropriate sensors.

The sensors can be:

- Single
- Vector (more sensors arranged along a line)
- Matricidal (sensors organized as a matrix)

Acquisition

A monochromatic digital image is obtained from data acquired through sensors The **output of a sensor is normally a continuous variable** The creation of the digital image starts from the analog signal and needs two processes: sampling and quantization. **Sampling is the discretization of spatial coordinate values** (we talk about spatial frequency, increasing the number of pixels, we increase the sampling rate, the spatial frequency). **Quantization is the discretization of the intensity values.**

The quality of the digital image depends on the number of samples, the quantization and the image content.

Sampling

Each pixel represents the intensity in the corresponding position of the sampling grid. Actually, a pixel is not just a point in the image, but rather a rectangular region that corresponds to a grid cell. The value associated with the pixel represents the average intensity in the cell. The value of that pixels I an average: increasing the number of pixels, will produce a better-quality image also for this reason.

Spatial resolution

Spatial resolution is related to sampling: is the smallest distinguishable detail in an image.

When an evaluation of the resolution that correlates the number of pixels with the level of detail of the original scene is not necessary, it is usual to simply say that an image with dimensions M x N has a spatial resolution M x N.

With large pixels, not only the spatial resolution is poor, but the gray discontinuities at the border between the pixels are clearly visible. As the pixel size is reduced, the effect becomes less visible to the point where you have **the impression of a continuous image.** This happens **when the pixel size becomes smaller than the spatial resolution of the human visual system** This depends on the distance and on the other observation conditions, **the number of pixels necessary to guarantee a good image quality cannot be defined a priori** (depends on dimension of sensor, distance from the target etc.). Surely **the pixel size must be small in relation to the scale of the objects represented in the image.**

Grey Levels and Quantization

The resolution of grey levels refers to the smallest perceivable grey variation. The grey level is related to the number of bits using in the quantization.

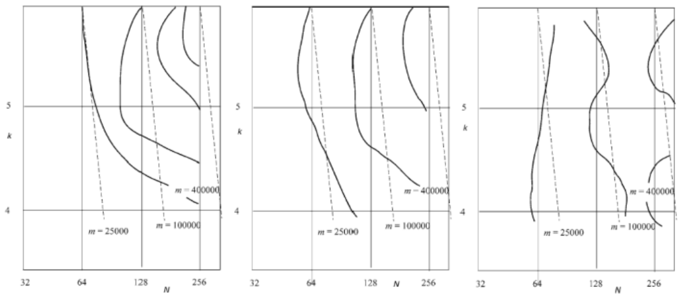
Reduction of the number of levels causes the reduction of image quality: false contours appear, and it is not possible to distinguish objects that differ by slow grey variations.

In other images you can see the **phenomenon of false contour in the areas of slow grey variation**, up to the limit case of two-level images. (the effect increases as the number of levels decreases).

Simultaneous variation of resolution and Quantization

The relationship between these two parameters and the effect of their contemporary variation can be **studied experimentally.** The experiment (Huang) consists of a set of subjective tests, performed by submitting to the judgment of a group of observers the quality of different versions (obtained by changing the values of L and N) of some sample images.

The sample images are chosen to contain a relatively **small number of details** (the face of a woman), or a **medium number of details** (the cameraman), or **a large number** of details (the crowd) The test results can be illustrated using the **isopreference** curves in the N-k plane. Each point in this plane represents an image with values of N and k equal to the coordinates of the point itself. **The isopreference curves link representative points of images characterized by the same subjective quality.**

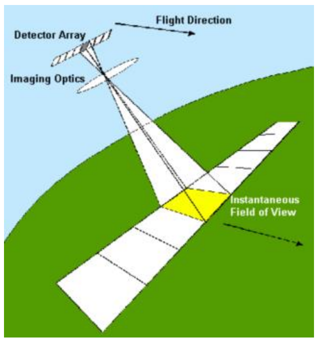


volto cameraman folla

In the vertical axis is represented the number of bits used for quantization, in the horizontal axis is represented the number of pixels: from the left to the right the curves represent images of increasing subjective quality. If we want to decrease the quantization level, we had to increase the sampling level in order to obtain the same quality image.

The dotted lines, that join points with constant bit number, move away sharply from the isopreference curves. The quality of the images obviously tends to grow with increasing k and N. In some cases, fixed N, the quality improves if you decrease k. This is probably due to the fact that **the decrease of k** generally **increases the apparent contrast of the image**. As the details in the image increase, the curves tend to become more vertical. Therefore, **only few gray levels are required for images with large amounts of detail.**

Pixel size and image resolution



Pixel size and spatial resolution are not interchangeable. The detail discernible in an image is dependent on the spatial resolution of the sensor and refers to the size of the smallest possible feature that can be detected. Spatial resolution of passive sensors depends primarily on their Instantaneous Field of View (IFOV).

- ⇨ Geometric resolution: it depends on the size of the pixel that change the visual appearance of an area. To produce a good image with narrow bandwidth, we can increase the integration time or increase the sensors size (so increasing the size of the single pixel)
- ⇨ Radiometric resolution: is the smallest change in the intensity level that can be detected. Is related on the number of bits used for the quantization for each pixel. It is limited by the number of discrete quantization levels used to digitize the continuous intensity value and by the signal to noise ratio of the detector.

Digital images Representation

The irradiance (intensity of our image) can be represented by the product of two terms, the illumination $i(x, y)$ (depends on the source of light) and the reflectance $r(x, y)$ (depends on the kind of surface that we consider):

$$f(x,y) = i(x,y)r(x,y) with\ 0 < f(x,y) < \infty, 0 < i(x,y) < \infty,$$

$$(doesn't\ reflect\ anythig)\ 0 < r(x,y) < 1 (reflect\ all\ waveleghts)$$

The image is constituted by a component due to the light coming directly from the lighting source and a component due to the light reflected by the objects present in the scene. The lighting component is responsible for the slow changes in brightness (low spatial frequencies), while the reflectance component gives rise to abrupt variations in brightness, for example in correspondence with the contours of objects (high spatial frequencies).

Intensity

The actual nature of $i\ (x, y)$ is determined by the light source, while $r\ (x, y)$ depends on the characteristics of the objects in the scene, and varies between 0 (total absorption) and 1 (complete reflectance)

Gray scale

It can therefore be assumed that $L_{min} \leq f(x,y) \leq L_{Max}$ where:

$$L_{min} = i_{min} r_{min}\ and\ L_{Max} = i_{max} r_{max}$$

(indoor $L_{min} = 0{,}005\ e\ L_{Max} = 100$).

For a monochromatic image, the interval $[L_{min}, L_{Max}]$ is called the gray scale, while the intensity $f\ (x, y)$ is also called the grey level of the image in the coordinate point (x, y). Practically we use a grey scale conventionally included in [0, L-1], where 0 corresponds to black and L-1 represents white. The discrete levels of grey represent the digital character of the f after the intensity quantization.

the continuous image after sampling and quantization is approximated by MxN equispaced samples along the two dimensions of the image:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix}$$

Sampling and quantization

A monochromatic digital image is a matrix f (x, y) of discrete values of light intensity (grey levels). It consists of M * N pixels, each of which has a value belonging to the interval [0, L-1] being L the possible levels of intensity. Typically, L = 2^k, where **k is the number of bits** used to encode each pixel (**pixel depth**).

Considerations

The response of the human visual system is logarithmic (non-linear), the eye is able to discriminate a huge number of brightness levels, much higher than the 256 previously hypothesized. A linear quantized digital image normally has a much smaller dynamic range than the human visual system. (it's due to the logarithmic response of the human visual system).

This is particularly evident high contrast scene: if the sensor is linear, the dark parts will be underexposed, or the light parts will be overexposed.

Gamma correction

Some specific elaborations (**homomorphic filtering**) are able to improve images with the highlighted defects, allowing at the same time to compress the dynamic range and increase the contrast in the darker areas. A more general solution to the problem of scenes characterized by too large dynamics is adopted in video cameras, in which **the quantized quantity is** not directly the intensity, but **its power f^{γ}, we apply the gamma correction to the intensity of the pixels:**
If *g* is the grey level and *f* represent the intensity, we have: $g = f^{\gamma}$, *not* $g = f$

An example of a grey level transformations is the **gamma correction**, this allows the correction of **the non-linearities introduced by the acquisition and / or visualization devices.**

Obviously, we must use $\gamma = 1$ in applications where it is essential to keep a linear relationship between intensity and grey scale.

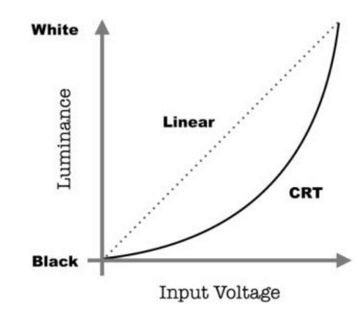
The non-linearities introduced by the acquisition and / or visualization devices depends on:

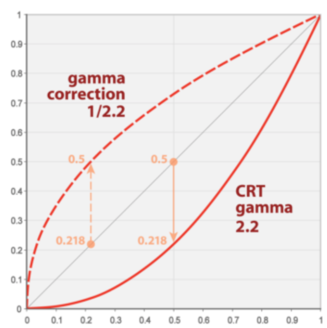
- ⇨ Different camera sensors: have different response to light intensity, produce different electrical signal for same input

How do we ensure there is consistency in:

- Images recorded by different cameras for given light input
- Light emitted by different display devices for same image

The reason we must gamma correct images lies in the historical need to deal with the exponential output response of the old Cathode Ray Tube (CRT) displays. The luminance would arc up from black to white as the input voltage increased. And because the data is linear, the final image would have mid-tones too dark.





Increasing the input in a linear way we obtain a ? response.

To correct the function, we can apply to the input signal, the inverse function of CRT

Characteristics of Gamma:

- Originates from analog photography
- Exposure function: relationship between logarithmic light intensity and resulting film density
- Gamma: slope of linear range of the curve
- The same in TV broadcasting

- **Gamma function:** a good approximation of exposure curve
- Inverse of a Gamma function is another gamma function with

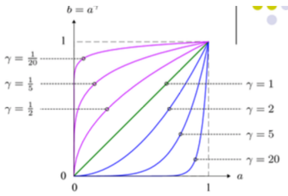
$$\bar{\gamma} = 1/\gamma$$

- Gamma of CRT and LCD monitors:
- 1.8-2.8 (typically 2.4)

$$s = B^{\gamma_c}$$ Output signal Raised by gamma

$$b = s^{1/\gamma_c} = (B^{\gamma_c})^{1/\gamma_c} = B^{(\gamma_c \frac{1}{\gamma_c})} = B^1$$

Correct output signal By dividing by 1/ gamma (called Gamma correction)



$$b = f_\gamma(a) = a^\gamma \quad \text{for } a \in \mathbb{R}, \gamma > 0$$

$$a = f_\gamma^{-1}(b) = b^{1/\gamma}$$

$$f_\gamma^{-1}(b) = f_{\bar{\gamma}}(b)$$

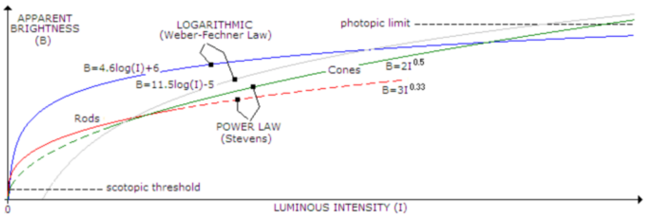
$$\bar{\gamma} = 1/\gamma$$

With the gamma correction we want to obtain a measurement b proportional to original light intensity B by applying inverse gamma function. Gamma correction is important to achieve a device independent representation. Any device has to be corrected using the specific gamma for that sensor.

An image appears brighter and darker for lower and higher gamma values, respectively, but does not change the black and white points. It strongly influences an image's apparent contrast.

Calibration: look-up Tables: Whenever the red, green and blue values are equal, an accurate monitor should display this as a neutral grey. Often this isn't the case. LUT is used to maintain neutral grey tones with the correct gamma. With the LUT an input value of R, G,B=159,159,159 is sent to your monitor as an output value of 145,155,162 (which is now perceived as neutral grey)

Gamma Encoding: Gamma encoding of images is used to optimize the usage of bits when encoding an image, taking advantage of the non-linear manner in which humans perceive light When a digital image is saved, it is "gamma encoded" — so that twice the value in a file more closely corresponds to what we would perceive as being twice as bright. gamma encoding redistributes tonal levels closer to how our eyes perceive them.



The linear encoding (greyscale values are proportional with their corresponding physical light intensities) uses insufficient levels to describe the dark tones — even though this leads to an excess of levels to describe the bright tones. On the other hand, the gamma encoded gradient distributes the tones roughly evenly across the entire range ("perceptually uniform"). Gamma encoding allows to keep more of the useful information where we prefer while avoiding encoding more precision than necessary for bright values.

Encoding original image with a greyscale that is **perceptually linear**, but consequently **non-linear in terms of emitted light intensity. That non-linearity would match that of the human vision.**

Weber law: the perceived object brightness changes with the logarithmic of object's actual brightness

Gamma Workflow: encoding & correction

The purpose of gamma encoding is for recording the image — not for displaying the image. A gamma encoded image has to have "gamma correction" applied when it is viewed — which effectively converts it back into light from the original scene. Gamma correction is automatically performed by monitor or software.

Aliasing

The presence in an image of fine details (strong variations of the grey level in narrow spaces) is related to the presence of high spatial frequencies in the image. An image can be thought of as a two-dimensional spatial signal, decomposable into a sum of (infinite) sine and cosine terms of various frequencies (spatial). The highest of these frequencies determines the fastest variation in the gray level in the image, and therefore the finer perceptive detail.

If the maximum frequency present in a function is finite (**band limited function**), the **Shannon sampling theorem** guarantees that **the function can be completely reconstructed starting from a sampled version only if the sampling frequency is greater than or equal to twice the maximum frequency contained** in the function. The **sampling rate** of an image is given by **the number of samples per unit of distance** (in both spatial directions).

If the image is **under-sampled**, the sampled version **is corrupted** by a phenomenon known as **aliasing.**

The aliasing is always present in digital images and in order to remove it, we have to use a low pass filter, in front of the filter, that can remove the higher frequencies before the sampling.

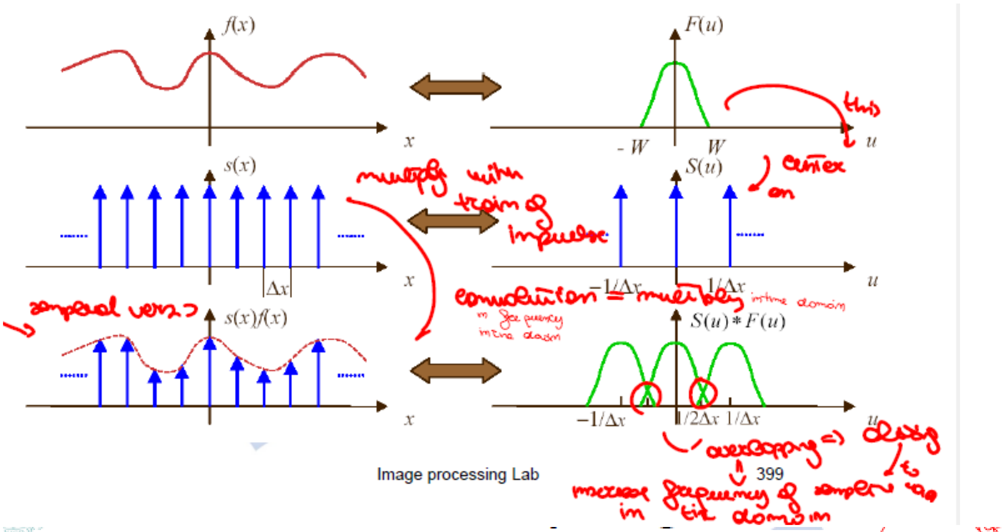
The sampled images we can obtain are always of finite extension (duration). **The sampling theorem cannot be satisfied in practice** (with the exception of periodic functions) **We can only work with sampled data that are finite in duration**. Aliasing consists of **adding high-frequency components** do not present in the original image (aliasing frequencies).

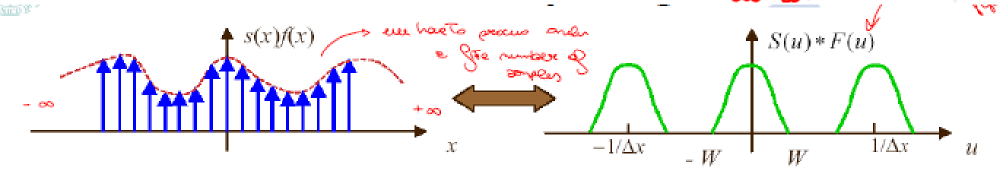
In this case, the reconstructed image shows a distortion with respect to the original, the **Moiré effect**.

A finite duration function can be obtained from a non-limited duration function, **multiplying by a rectangular window function** (all samples with weight = 1 into in the area of interest and zero elsewhere).

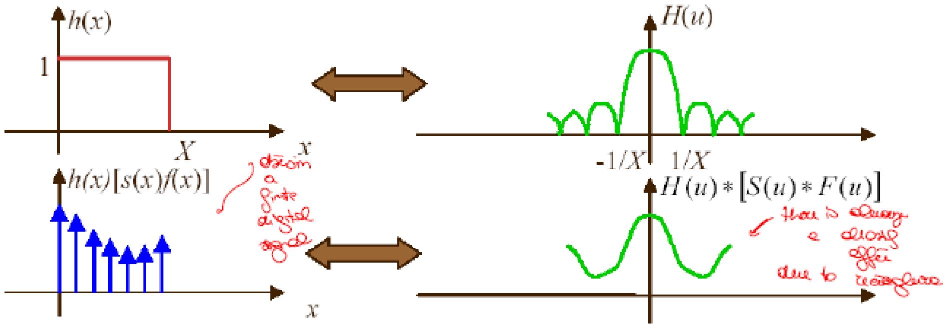
But the window function has a frequency content that extends to infinity and therefore we obtain an unlimited band function, **violating the basic condition of the sampling theorem**.

There is always. Aliasing in a sampled image, although you can try to reduce it by limiting the high-frequency content of the image by low-pass filtering (before sampling).





The sampling in the interval [0, X] can be represented mathematically by multiplying the sampled function for a window h (x) having amplitude equal to one in [0, X] and zero otherwise.



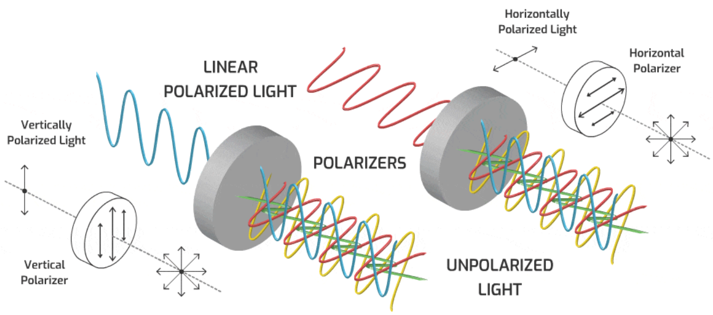
In our camera there is always an optical low pass filter before the sensor (after the red filter).

The optical low pass filter is composed of two layers of transparent birefringent material and inside of them there is another material.

How works the optical low pass filter

1. **Polarization**

Ordinary light wave that is vibrating in more than one plane is referred to as **unpolarised** light. If the electric fields vector oscillates along a straight line in one plane, the light is said to be **linearly polarized**. **Polarization** is the process to transform unpolarized light into polarized light. With a polarized filter we can select only one vibration plane and after



the filter we see only one plane. Rotating it we can change the plane chosen

2. Anti-Aliasing – Calcite crystal birefringence

The optical low pass filter uses the feature of a particular material the calcite crystal that is a **Birefringence material.**

Birefringence is the property of optically non-isotropic transparent materials whose refractive index depends on the polarization direction (direction of the electric field). An **unpolarized light beam** is splitter into **two** linearly **polarized** beams when hitting the surface of a birefringent crystal (**double refraction**). When an object, which is illuminated with unpolarized light, is viewed through a birefringent crystal (e.g. made of calcite), two images occur which are slightly displaced.

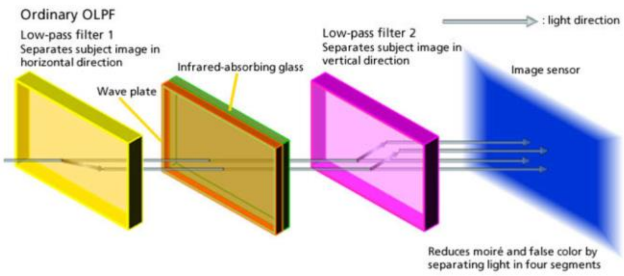
Using the polarized filter over the calcite crystal, we can select with the rotation of the filter, one of the two images under the crystal. In our camera there are two layers of these material.

3. Structure of a optical low pass filter

1°-layer birefringent material: split the signal into two horizontal output signal each of them is now linearly polarized

2° -layer Wave plate: transform the linearly polarized light to again a circular polarized light

3° -second layer birefringent: transform each one of these two circular polarized light in two polarized light in the vertical direction, so we obtain two outputs of light for each circular polarized light.



Filters made from birefringent materials, which bend or diffract light differently, depending on its polarization. The incoming light is randomly polarized, so is split into two half-images, shifted from each other very slightly in one direction. The so-called **"wave plate"** is technically a quarter-wave plate, which **changes linearly polarized light back into circularly polarized light**. This means that each of the half-images coming from the first LPF will be split again by the second LPF, this time vertically.

In this way we reduce the frequency of the original signal.

Removing the low pass filter allows to the camera to acquire image with more fine details using only the high resolution of the camera, high number of pixels on sensor. But this removing is very expensive so usually, producer just remove the effect of the filter with changing the direction of the second birefringent layer.

If nothing is done to the light leaving the first half of an LPF element, it just gets recombined when it passes through the second. If, on the other hand, its plane of polarization is rotated by the LCD layer, the light won't recombine, and the full low pass filter effect will be felt.

Moirè Effect

A periodic function sampled at a frequency that meets the sampling theorem, can be recovered is the sampling captures exactly a whole number of periods of the function. Moirè fringes are caused when a fine pattern in the subject (such as the weave in a fabric or very close, parallel lines in architecture) matches the pattern of the imaging chip. Moiré occurs when two patterns are overlaid and result in a new, third pattern. The frequency of detail exceeds the sensor's pixel pitch and ability to resolve “real” information.

Temporal Aliasing

The sampling rate (number of frames per second) of a scene is too low compared to the speed of objects inside of the scene.

Visual Acuity and resolving detail on Prints

Visual acuity is defined as the ability to read a standard test pattern at a certain distance, usually measured in terms of a ratio to “normal” vision.

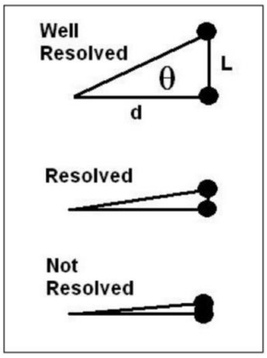
Visual acuity (Visus) is a measure of the spatial resolution on the visual processing system. It rates the examinee's ability to recognize small details with precision and depend on optical and neural factors

$$Acuity = \frac{1}{gap\ size\ [\text{arc min}]}$$

Several studies report the value of 1.7 for human visual acuity. The acuity of 1.7 corresponds to 0.59 arc minute PER LINE PAIR It is needed a minimum of 2 pixels to define a line pair, so the pixel spacing is 0.59/2 = 0.30 arcminute Each pixel must appear no larger than 0.3 arc-minute.

Visual angle is the angle, at the eye, under which the **optotype** (**Snellen** 1862) appears.

“**Standard vision**” is the ability to recognize an optotypes from a distance of **20 feet** when it subtended an angle of 5 minutes of arc (**features separated by 1 minute of arc must be distinguished**).



The angle separating two objects is simply tanθ = L/d. for angles much smaller than 1°:

$$\theta\ =\ L/d\ (\theta\ is\ measured\ in\ radians)$$

L = distance between the two objects

d = distance from the observer

One radian = 57.296o, and since 1°= 3600 arcseconds, therefore, **θ = 206265 L/d** arcsecond.

The formula for calculating the FOV:

$$FOV = 2\ arctang\left(\frac{h}{2FL}\right)\ where\ h = sensor?\ pixel?\ size\ and\ FL = focal\ lenght$$

The instantaneous Field of View Formula is:

$$IFOV = \frac{DES}{FL}$$

IFOV corresponds to theta.

Where:

- DES is the detector element size
- FL is the camera focal length

Focal plate, total size of the sensor (number of pixels * dimension of one pixel)

$$Focal\ plate = \ angular\ size\ (degree) * \ FL$$

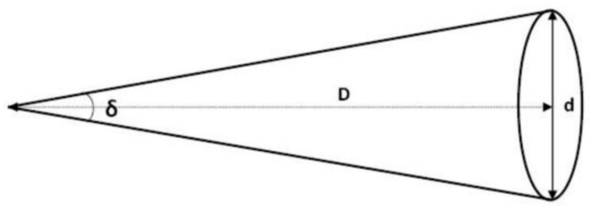
In modern studies, like Curcio et al. (1990), acuity is measured in cycles per degree. Again, you need a minimum of 2 pixels to define a cycle, so **the pixel spacing is 0.78/2 = 0.39 arcminute**, close to the above numbers. Each pixel must appear no larger than 0.3 arc-minute.

The generally accepted definition of **normal (20/20) visual acuity is the ability to resolve a spatial pattern whose features are separated by one minute of arc.**

If you don't want to see a pixel on a print, pixel must be smaller than the smaller detail that can be recognize.

The printer-output resolution depends on viewing distance. The normal viewing distance of a print is typically between 1.5 and 2 times the diagonal of a print.

DPI and viewing distance for a person with the nominal 6/6 vision. **The printer-output resolution depends on viewing distance.**



$$\delta = 2\arctan\left(\frac{d}{2D}\right) \quad d\ = pixel\ pitch$$

The PPI Pixels per inches, number of pixels in the picture that are contained in a single inch, and is equal to:

$$PPI = \ 25.4\ (size\ of\ inch\ in\ millimetres)\ /\ size\ of\ the\ pixel\ d$$

With this expression we obtain the minimum PPI needed in our print to not observe single pixels.

The angular diameter, angular size, is an angular distance describing how large a circle appears from a given point of view.

Resolution: lens diffraction, pixel and sensor size

Lens resolution basics

The finesse of detail in a photograph is **resolution**. This resolved detail has a particular degree of visibility, depending on **contrast**. Resolution and contrast determine image **clarity**. **Sharpness** is determined by edge definition in the resolved detail, which depends on the contrast of the edge and **acutance** is a measure of sharpness. **Resolving power** is an objective measure of resolution (in resolved line pairs or cycles per millimetres units). Resolving power and acutance aren't good measures of image quality if we take them separately.

The **modulation transfer function (MTF)** is a complex objective measure of image quality, and it combines resolution and contrast. The modulation transfer functions are mathematical expressions of the signal transmitted by a lens. The signal has two properties: frequency (spatial or temporal) and amplitude. In photographic signals, (spatial) frequency is resolution and amplitude is contrast. The more pair of lines (one black, one white) we have into a spatial unit, the higher is the signal frequency or resolution. We can think on contrast as the brightness difference between adjacent areas.

Lenses cannot keep all the details at full contrast (as present in the subject). Transmitted contrast is the degree to which black lines remain black, and white lines remain white. When contrast drops, pure black and white lines became grey, and differences between lines vanish. We can measure contrast as a percentage. Any value below 100% implies a loss, and lenses can transfer only coarse details at the maximum level of "fidelity". The finer the detail is, the larger are the contrast losses, due to aberrations and diffraction. The level of contrast provided by the system for the relevant range of detail is the key variable in determining the subjective perception of image quality.

Perceived quality

The separation between cones in the fovea is 0.0015 millimetres, which limits the maximum possible visual resolution to 20 arc-seconds. In practical terms, 60 arc-seconds can be a good practical value for the average absolute visual acuity limit. At a viewing distance of 25cm, 60 arc-sec translates to black spots on bright background of 0.07 millimetres of diameter. A spot of 0.07 mm of diameter doesn't correspond to a line of that width, but to a line pair of that width, because light spots have brightness differences from centre to borders. This implies 14 lp/mm.

A very small, isolated spot cannot excite a sufficiently large number of retinal cones, but a line does. We can perceive more detail if it is formed by lines instead of spots, which implies that we can see a line thinner than the diameter of the minimum spot size perceivable. Another factor to be considered is contrast. The eye resolution also depends on the contrast between the bright areas and the dark areas. We can also see a thinner line if it is isolated from other adjacent lines. We can see an isolated black line on bright background if the line is at least 0.001 millimetres thick (1 micron), considering an optimum (for resolution) viewing distance of 25 centimetres. This translates to 0.8 arc-seconds.

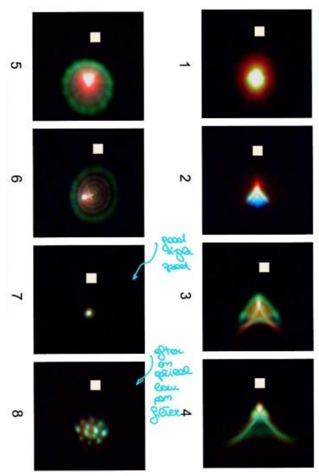
We can see an isolated 0.001 millimetres thick (1 micron) line! When we look at groups of lines the visual resolution drops. The eye can resolve two black lines on bright background separately if the distance between them (from centre to centre) is at least 0.05 millimetres (40 arc-seconds, considering **25cm of viewing distance**). This translates to **20-line pairs per millimetre**. Visual acuity figures aren't accurate for evaluating the resolving power of a digital system because it is related to the contrast. For any meaningful comparison or evaluation, it is the performance at the limit that counts.

Resolution

Resolution describes how much detail a lens is capable of capturing, quantifies how close lines can be to each other and still be visibly resolved. (Line pairs per mm).

We have always to consider the resolution of the lens (related to diffraction) and also the resolution of the sensor.

Lens diffractions



Diffraction is an optical effect which limits the total resolution of digital photography — no matter how many megapixels the camera may have.

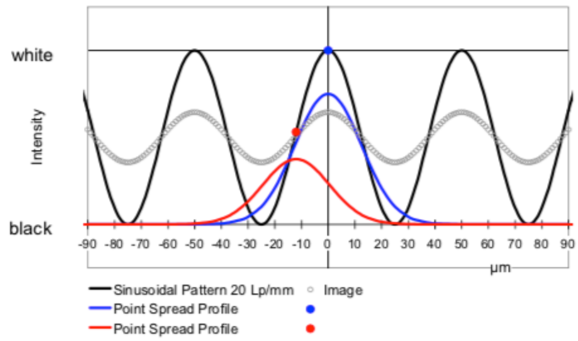
Aberrations of the lens systems, production tolerances, and ultimately the wave-like nature of light as well, are the reasons why the light originating from one point of the object is always distributed over an area around the ideal image point.

The **Point Spread Function** (**PSF**) is the response of a single source of light spot (but we cannot use it to characterize the lens because we have to convert an image to an objective value), its shape and size characterize the image quality of a lens. **Aberrations** Of the lens systems, production **tolerances**, and ultimately the **wave-like nature of light** as well, are the reasons why the light originating from one point of the object is always distributed over an area around the ideal image point.

Why is PSF not used to quantitatively describe image quality?

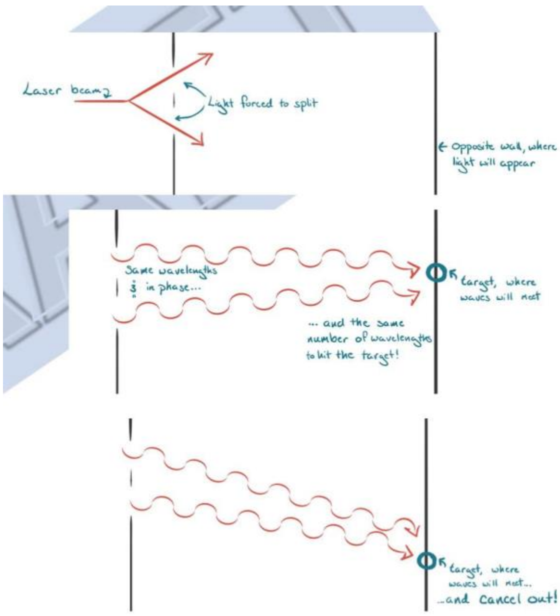
- The shape is sometimes very complicated and therefore defies a simple numerical description.
- Images are generated in the camera in a complicated way of combining the parts of a large number of single point spreads, The intensity at one point of the image is generated by a two-dimensional integration of many point spreads. It is very unusual to see single, isolated point spreads (only stars on a dark night).

Sinusoidal brightness distribution is used to examine how an object that looks as simple as possible is imaged. The sinusoidal brightness distribution is a pattern of bright and dark stripes in which the transition between bright and dark occurs gradually and continuously**.**

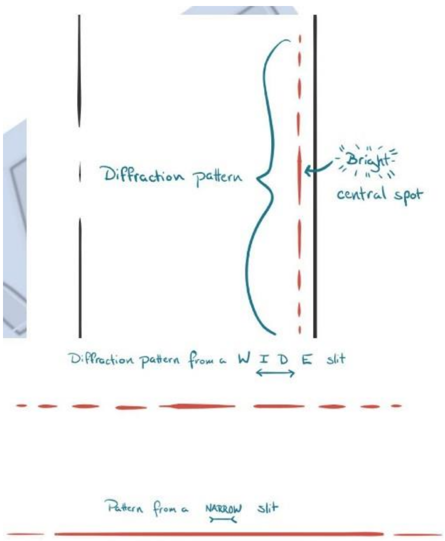


1. Double slit interference

Parallel rays of a monochromatic light of wavelength l are incident on a diffraction grating two light waves, one from the top slit and one from the bottom slit will start to interfere.

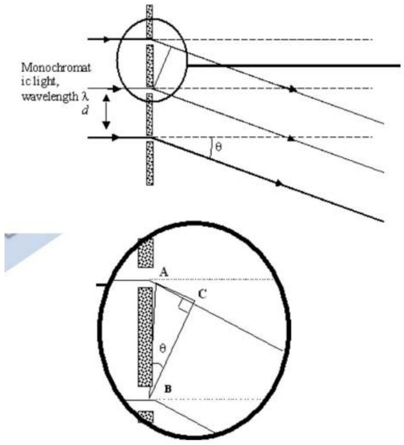


- **Constructive interference**, the difference in wavelengths is an integer number of whole wavelengths.
- **Destructive interference**, the difference in wavelengths is an integer number of whole wavelengths plus a half wavelength.



Moving away from the centre, the two waves' **pathlengths** will get more and more different, until they have the same plus a half wavelength (one of the waves will hit the wall with a crest, the other hits with a trough) resulting in a dark spot there. In the centre if have always the maximum bright spot.

Derivation of the Diffraction Grating Formula



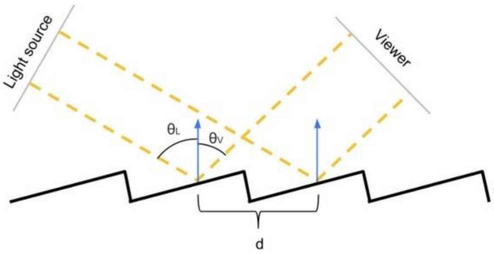
Diffraction grating with slit separation d.

Constructive interference: Light from A must be in phase with light from B, and this can only happen when the path difference is a whole number of complete wavelengths

$$d \ sin\theta = n\lambda$$

Where d is the distance between the two apertures; n goes to 1 to infinity and λ is the wavelength of light. For n=1 we have the first diffraction maximum.

Iridescent reflections

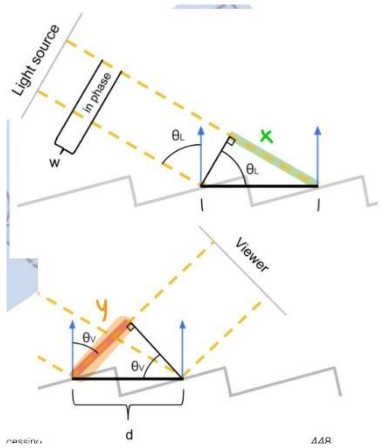


If a material presents a microscopic slit or a bump, this will an incoming planar wave to scatter in all direction. If slits or bumps are arranged in a regular pattern, a new wavefront is generated for each one of them. All those wavefronts will interfere with each other, causing certain wavelengths (colours) to appear prominently.

Let's consider a material which features imperfections that repeat at a known distance *d*. The angle between the incident light rays and the surface normal is θ_L (angle of the source light). Let's also imagine that the viewer is oriented in such a way that they receive all the reflected rays with angle θ_V (angle between the surface of CD and the human eye)

the scattering pattern itself repeats every *d* nanometre.

Rays are guaranteed to be in phase until the first one hit the surface. The second ray travels an extra distance X (green) before hitting the surface.



$$X = dsin\theta_L$$

The first ray travels the extra distance Y towards the viewer before the second hit:

$$Y = dsin\theta_V$$

If the difference of the rays path is zero or an integer multiple of the wavelength, the rays are in phase. Mathematically, those two rays are in phase if they satisfy the following condition:

$$dsin\theta_L - \ dsin\theta_V = n * w$$

$$sin\theta_L - sin\theta_V = \frac{n * w}{d}$$

All the incoming wavelengths w which are integer multiples of: $d(sin\theta_L - sin\theta_V)$ will interfere constructively and appear strongly in the final reflection.

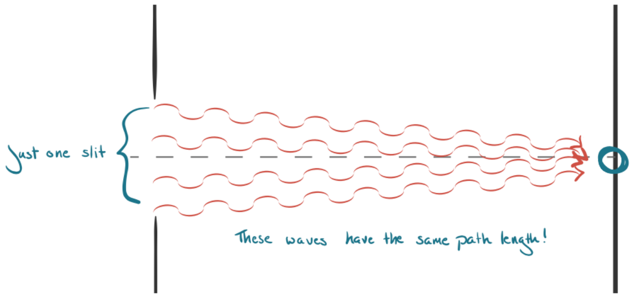
2. Single slit interference

A slit (wider than a wavelength) illuminated by light diffracts the light into a series of circular waves. These can be explained by assuming that the slit behaves as thought it has a larger number of point sources spaced evenly across the width of the slit.

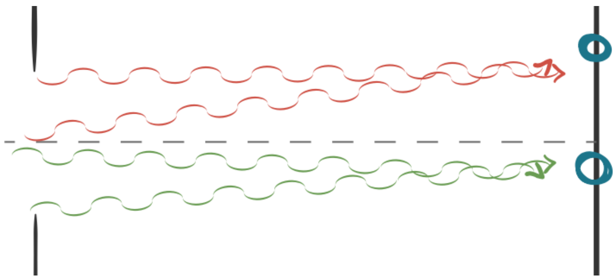
The Hughens-Fresnel principles said that: each portion of the slit acts as a source of waves.

So, even one single aperture can produce interferences. The points at the edges of the slit have an equal distance from the centre of the slit (the same path length), interfere constructively.

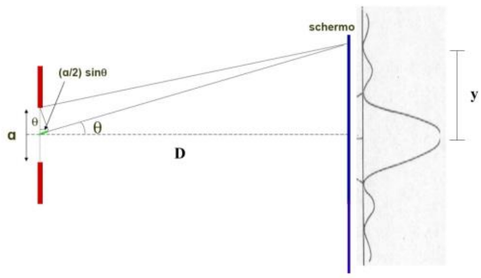
Two point that are further in, but still the same distance from the middle of the slit, they will also have equal path lengths to the centre point on the wall. They will also interfere constructively with one another. So, in the wall I can see always additive interferences.



Considering now, a single aperture but split: two points, one at the top edge and one just below the centre line, each pair will have the same difference in path length. (not the absolute path length, just the same difference in path length). The point on the wall where one pair has a half wavelength difference in path length, there will be a lot of destructive interference.



Now we evaluate the interference just for the first half of the aperture: To simplify consider light of a single wavelength. If the incident light is coherent, these sources all have the same phase. Divergent rays travel different distances, some move out of phase and begin to interfere with each other. This interference produces a diffraction pattern with peak intensities where the amplitude of the light waves adds, and less light where they subtract.



The light from a source located at the top edge of the slit interferes destructively with a source located at the middle of the slit, when the path difference between them is equal to $\frac{\lambda}{2}$. The source just below the top of the slit will interfere destructively with the source located just below the middle of the slit at the same angle. The minimum intensity occurs with a difference given by:

$$\frac{a}{2}\ sin\theta = \frac{\lambda}{2}\ (where\ a\ is\ the\ dimension\ of\ the\ aperture)$$

A similar argument can be used to show that if we imagine the slit to be divided into four, six, eight parts, etc., minima are obtained at angles θn given by:

$$a \sin \theta_n = n\lambda$$

Where n is an integer other than zero.

The intensity profile can be calculated using the Fraunhofer diffraction equation as:

$$I(\theta) = I_0 \ \text{sinc}^2\left(\frac{d\pi}{\lambda}\sin\theta\right)$$

3. Circular Aperture: The Airy Disk

The far-field diffraction of a plane wave incident on a circular aperture is often referred to as the Airy Disk (any source of light produces it). The variation in intensity with angle is given by:

Where a is the radius of the circular aperture, k is equal to $2\pi/\lambda$ and J_1 is a Bessel function.

$$I(\theta) = I_0 \left(\frac{2J_1(ka\sin\theta)}{ka\sin\theta}\right)^2$$

The angle at which the first minimum occurs is given by: $\theta = 1.22\frac{\lambda}{d}$ where θ is in radians, λ is the wavelength of the light, and d is the diameter of the aperture.

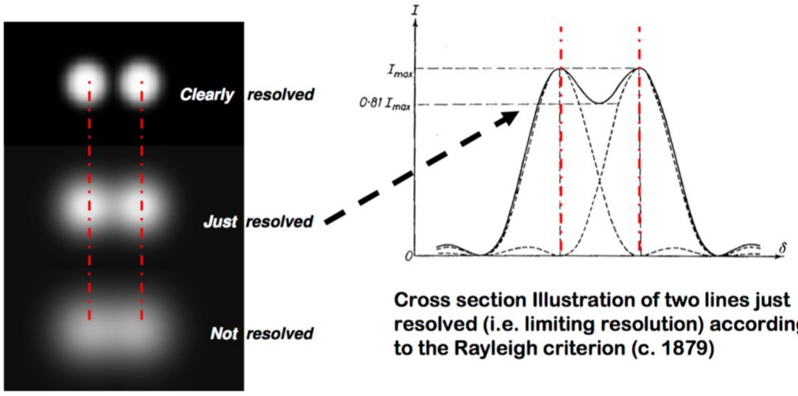
The smaller the aperture, the larger the spot size at given distance, and the greater the divergence of the diffracted beams:

So, changing the size of the frontal lens, we obtain a different value of the Airy Disk. Usually, the intensity of the signal outside the centre, could be too slow than the noise level, so our sensor sees only the centre part. We can change the opening of our lens with diaphragm (diaframma).

The Airy disk describes the best focused spot of light that a perfect lens with a circular aperture can make. Because **any detector** (eye, film, digital) used to observe the diffraction pattern can **have** an **intensity threshold for detection**, **the full diffraction pattern may not be apparent.** In astronomy it may be that none of the rings are apparent, in which case the star image appears as a disk (central maximum only) rather than as a full diffraction pattern. Furthermore, **fainter stars will appear as smaller disks than brighter stars**, **because less of their central maximum reaches the threshold of detection**. While in theory all-stars or other "point sources" of a given wavelength and seen through a given aperture have the same Airy disk radius differing only in intensity, the appearance is that fainter sources appear as smaller disks, and brighter sources appear as larger disks.

The Airy disk describes the best focused spot of light that a perfect lens (without any kind of aberrations) with a circular aperture can make. The width of the airy disk is used to define the theoretical maximum resolution for an optical system (defined as the diameter of the first dark circle). Even if one were able to make a perfect lens, there is still a limit to the resolution of an image created by such a lens. **An optical system in which the resolution is no longer limited by imperfections in the lenses but only by diffraction is said to be diffraction limited.**

Airy Disk and Resolution



Cross section Illustration of two lines just resolved (i.e. limiting resolution) according to the Rayleigh criterion (c. 1879)

$$\frac{\text{peak-to-peak difference}}{\text{peak-to-peak sum}} \equiv \text{modulation} = \frac{1.0 - 0.81}{1.0 + 0.81} = 0.10$$

The diameter d of the Airy disk is measured from the centre of the first black ring, across the bright centre portion, to the centre of the same black ring on the other side.

We can distinguish two near point respecting the Rayleigh Criterion: two-point sources can be considered to be resolvable if the separation of the two images is at least the radius of the Airy disk, i.e. if the first minimum of one coincides with the maximum of the other. 0.10 is the minimum condition for distinguish 2 airy Disk.

We can calculate the distance between the two-object projected as the result of two airy disk: $x = 1.22\frac{\lambda f}{d}$ where X is the separations of the images of the two objects on the film and f is the focal length.

The ratio f/d is the f-number of the lens.

We can also say that $\frac{x}{f} = 1.22\frac{\lambda}{d} = \theta$

Diffraction thus sets a fundamental resolution limit that is independent of the number of megapixels, or the size of the film format. It depends only on the f-number of the lens, and on the wavelength.
The larger the aperture of the lens, and the smaller the wavelength, the finer the resolution of an imaging system. This is why telescopes have very large lenses or mirrors. The telescope brings the plane-parallel waves incident upon the aperture into focus, forming an image of the star in the focal plane. It does this by inducing a phase change on the wavefront which varies across the telescope aperture. However, **since the aperture does not cover the entire wavefront radiated by the star, but only a very small portion of it, a diffraction pattern is produced.** The bright central spot contains 84% of the light, and the first ring contains less than 2%.

Pixel size

The size of the airy disk is primarily useful in the context of pixel size.

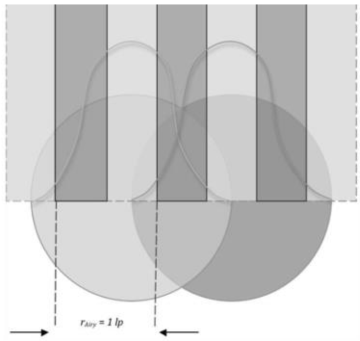
Large pixels – those which are bigger than the Airy disk – do not show diffraction at the same apertures that a small-pixel camera would. In smaller pixel size the same Airy Disk covers more pixels so the diffraction is more evident because is spread over more pixels. (but the size of the Airy disk depends on the colour of the source light that we are considering).

At **low f-stop** values, the width of airy disk is typically much smaller than the width of a camera pixel: pixel size determines the camera's maximum resolution.

At **high f-stop** values, the width of airy disk can surpass pixels size: camera's resolution is said to be "**diffraction limited**" (pixel size is smaller than the airy disk), the resolution of the camera is determined only by the width of the airy disk, and this resolution will always be less than the pixel resolution. An airy disk can have a diameter of about 2-3 pixels before diffraction limits resolution (assuming an otherwise perfect lens).

Circle of confusion diameter limit (CoC limit or CoC criterion): the largest blur spot that will still be perceived by the human eye as a point (it depends on Visual acuity, Viewing conditions, Enlargement from the original image to the final image).

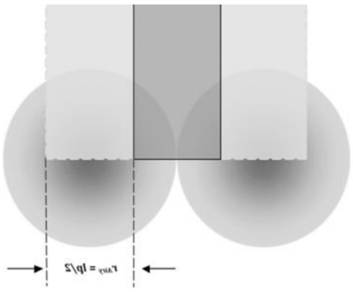
Rayleigh Criteria: two Airy discs are separated such that the first minima of one fall on the maximum of the other, thus the discs are separated by their radius. To resolve these two discs, it needs a minimum of three pixels, one centred on each peak and one centred on the dip between them. To do this, **the pixels can be no larger than a quarter of the diameter of the disc.**

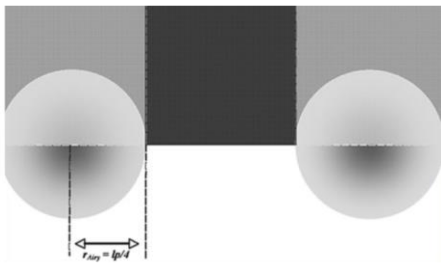


Increase the distance, increase the contrast and decrease the distance x to distinguish two Airy Disk.

Resolution of the system

Due to the regular arrangement of pixels digital sensors are bandwidth-limited detectors. The Nyquist-Shannon theorem says that in order to get an accurate reproduction of a continuous signal with a particular frequency the sampling frequency must be at least the double of that number. **It needs at least 2 samples per cycle, and this means two pixels per line pair.**





When the frequency is higher, the two signals are overlapped, the contrast is lower and also the distances between two-line decreases; instead with lower frequency the contrast is higher.

The Diffraction Limit is the theoretical maximum resolving power of the lens given in line pairs per millimetres $MAX\ resolution\ =\ 1/(wavelength\ *\ f-number)$

maximum resolution for a diffraction-limited (aberration free) lens at different apertures and for different levels of contrast for a diffraction-limited lens and green-yellow light.

Small format sensors may be lens-limited in terms of resolution (but low signal to noise ratios) =>more sensor size, higher space resolution but also more noise. Alternative way for more detail is more capture surface, this is, a larger format, but aberrations are harder to control.

The best way to increase the quality is using a bigger sensors size but he needs to use larger lens and so, larger focal length that introduces other bad effect that are difficult to eliminate if we don’t have a good and expensive lens.

Photography

Larger apertures are better? Common lenses are also quite soft when used at the largest aperture available. Camera lens typically have an optimal aperture in between the largest and smallest settings; **Are smaller pixels somehow worse?** Not necessarily. **The camera with the smaller pixels will render the photo with fewer artifacts (aliasing). Smaller pixels can yield a higher resolution if**

using a larger aperture is possible. On the other hand, when other factors such as noise and dynamic range are considered, the "small vs. large" pixels debate becomes more complicate.

Resolution and contrast

Resolution and contrast are equally important.

Resolution describes how much detail a lens is capable of capturing — and not the quality of the detail that is captured, it is impossible to quantify lens performance by merely looking at resolution (details in an image). **A good lens must be able to resolve enough detail, while having a reasonable amount of contrast to distinguish between those details.**

Sharpness contributes to our perception of the quality of a digital image and is determined by **edge definition** in the resolved detail (**edge contrast**). "**Acutance**" describes a **subjective perception of sharpness** that is related to the edge contrast of an image. Acutance **is related to the amplitude of the derivative of brightness with respect to space.**

Due to the nature of the human visual system, an image with higher acutance appears sharper even though an increase in acutance does not increase real resolution.

1) Low acutance and high resolution. The lens was able to resolve plenty of detail, but the transition between the edges is not sudden, which makes the image appear a little soft.
2) Low resolution, but high acutance, an excessive amount of sharpening was applied to make the image look sharp. Some features are grossly over-exaggerated
3) The sharpest and the most detailed. Our perception of sharpness is highly dependent on both resolution and acutance

SHARPENING: exaggerates the light and dark edges of the transition; making the bright side of an edge a little bit brighter, and the darker side a little bit darker it improves the **micro contrast** and the edge steepness, The subjective impression of sharpness is significantly improved, without significantly increasing the resolution of detail.

Sharpness is always subjective. One image could appear sharp to one person, while appearing blurry / soft to another (never-ending debates on why one lens is better than another).

Resolving power is an objective measure of resolution, and acutance is a (subjective) measure of sharpness Manufacturers need objective methods to measure lens performance in controlled lab environments without relying on human perception.

The modulation Transfer Function

We have to define an objective method to measure the sharpness, so we use the modulation transfer function.

In optics, the difference between bright and dark is referred to as "**contrast**":

$$Contrast = \frac{Maximum - Minimum}{Maximum + Minimum}$$

Comparing the contrast of the image with the contrast of the object by dividing these two figures by each other, we get a simple figure that provides a statement about the imaging properties of the lens: **the modulation transfer.**

A Modulation Transfer Function (MTF) quantifies how well a subject's regional brightness variations are preserved when they pass through a camera lens. A very coarse pattern with large separations between bright and dark stripes could be imaged well by a lens with a relatively large point spread function. If we decrease the separation between the stripes (the separation between bright and dark approaches the size of the point spread), then a lot of light from the bright zone is radiated into the darker zones of the pattern and the image contrast becomes noticeably lower. **How the lens reproduces a stripe pattern of various degrees of fineness, determines the modulation transfer for every one of these patterns.**

Plotting MT as a function of the fineness of the stripe pattern we obtain the modulation transfer function plot (MTF). **The number of line pairs (black, white) in the pattern per millimetres is the spatial frequency, lp/mm.**

- MTF of 1.0 represents perfect contrast preservation
- MTF of 0 represent line pairs that can no longer be distinguished at all.
- An MTF of 9% (0.09) is implied in the definition of the Rayleigh diffraction limit

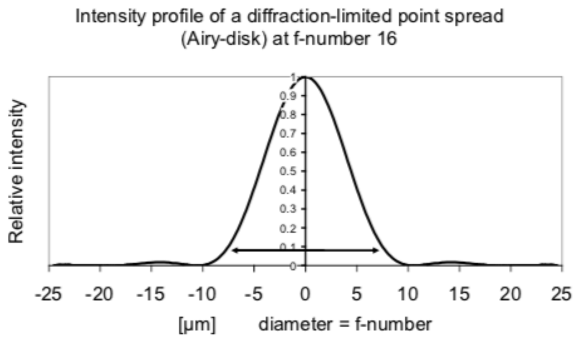
A diffraction-limited lens has an almost perfectly straight MTF curve which decreases in proportion to the spatial frequency.

The zero MTF value is reached at the so-called limit frequency, which is determined by the f-number and the wavelength of the light.

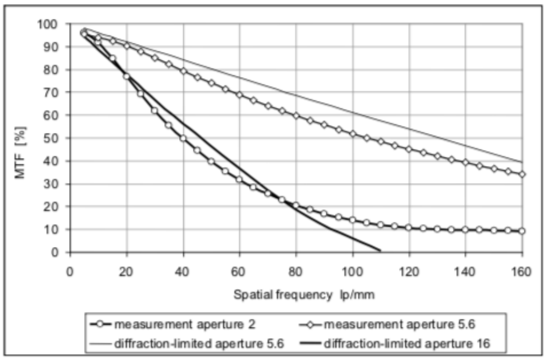
$$\boldsymbol{limit\ frequency = 1/(f - number * wavelength)}$$

Where: $f - number = lens\ focal\ /\ lens\ diameter$

A **limit frequency estimate** for medium wavelengths of visible light (0.000555 mm) is: **The width of the point spread in µm corresponds to the f-number, and the limit frequency is about 1500 divided by the f-number.**

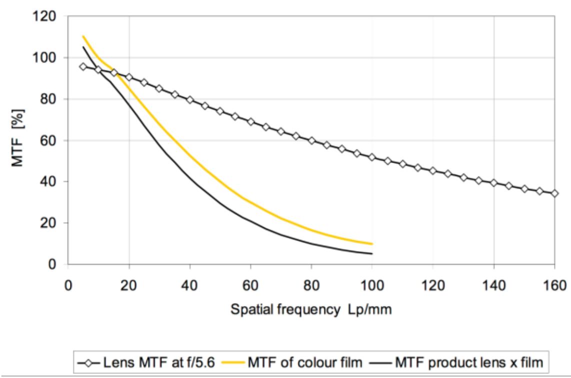


A diffraction-limited lens has an almost perfectly straight MTF curve which decreases in proportion to the spatial frequency. For real lenses with residual aberrations the MTF values initially decrease quickly and then very slowly approach the zero line (limit frequency for this lens).



The spatial frequency at which the MTF value reaches zero or falls below a low threshold (e.g. 10%) is the **resolving power of the lens** in air. The measurement can be very imprecise (F-number 2 in the image). For this reason alone, the resolving power in air is therefore not a suitable quality criterion for lenses!

All image components, (human eye. lens, sensors etc), have their own imaging properties, each of which can also be described by a transfer function (advantage). **In frequency domain the response of an entire system can be calculated by multiplying the responses of each component:**



MTF and quality of real pictures:

Fine periodic patterns represent only a small fraction of the subject properties which our eye uses to recognize imaging quality. Most important are really the edges, the borders between two areas with different brightness.

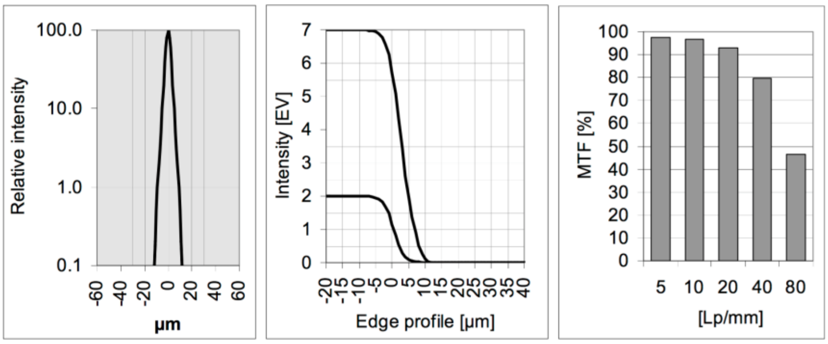
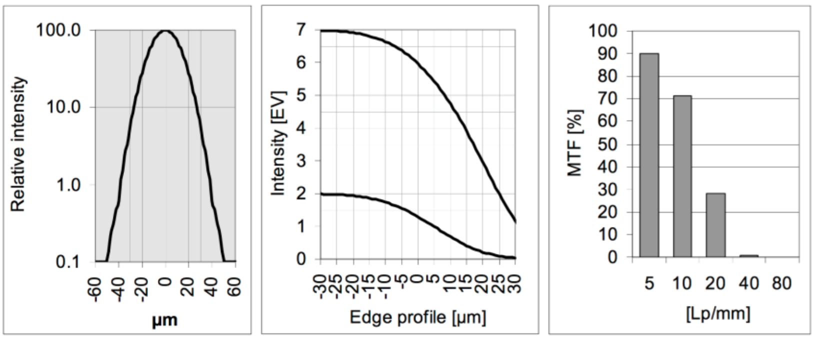
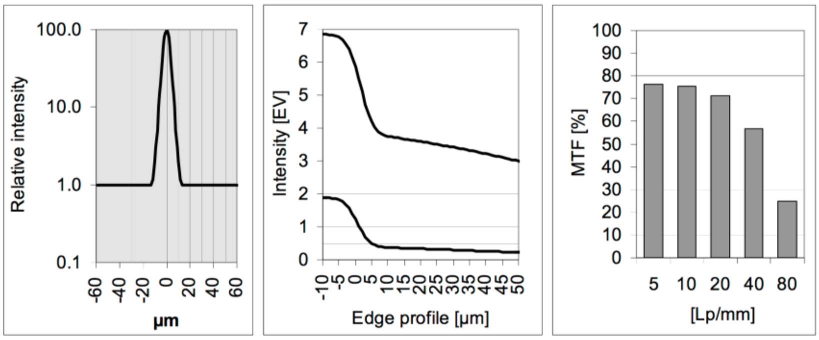


Figure 1: example of a very good imaging performance

Point Spread narrow: the image of the edge is sharp => For a lens with such imaging performance, the image quality achieved is usually limited by the sensor or by other factors such as focusing accuracy, camera movements etc.



Point Spread large: the edges are no very sharp, (edge profile is flat, the transition from maximum brightness to black is large). A deep black is nevertheless achieved, the contrast between the ends of the above scale is therefore high.



Point Spread narrow but surrounded by a weak halo. We start form 1.0 not from zero this is due for the bias in this way we never reach the zero. The edge definition is high, but a broad bright fringe

Figure 2: good quality, low micro contrast, high degree of veiling glare

stretches into the dark zone, the lens exhibits flare. The contrast near the edge is low. Edges with low to medium contrast are reproduced with the same degree of sharpness, Fine structures with a lot of contrast appear a little bit flat and can show flare.

With poor micro contrast (middle) the right peak in particular is broadened towards the left. The separation of the two peaks on the grey scale is the same as with the good picture on the left. With the high level of veiling glare, the lower peak of the histogram is shifted upwards because the black is lit up by the veiling glare in the whole area.