Cameras

A camera is a remote sensing device that can capture and store or transmit images. Light is collected and focused through an optical system on a sensitive surface (sensor) that converts intensity and frequency of the electromagnetic radiation to information, through chemical or electronic processes.

The simplest system of this kind consists of a dark room or box in which light enters only from a small hole and is focused on the opposite wall, where it can be seen by the eye or captured on a light sensitive material (i.e. photographic film). This imaging method, which dates back centuries, is called 'camera obscura' (latin for 'dark room'), and gave the name to modern cameras.

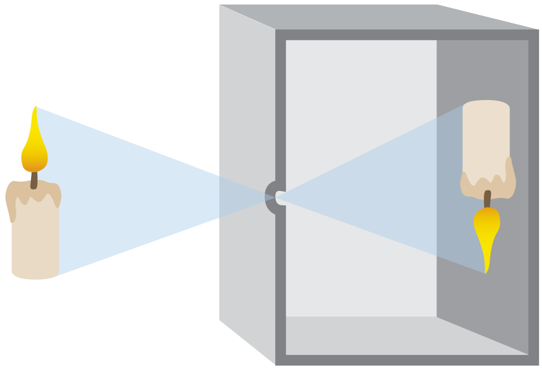
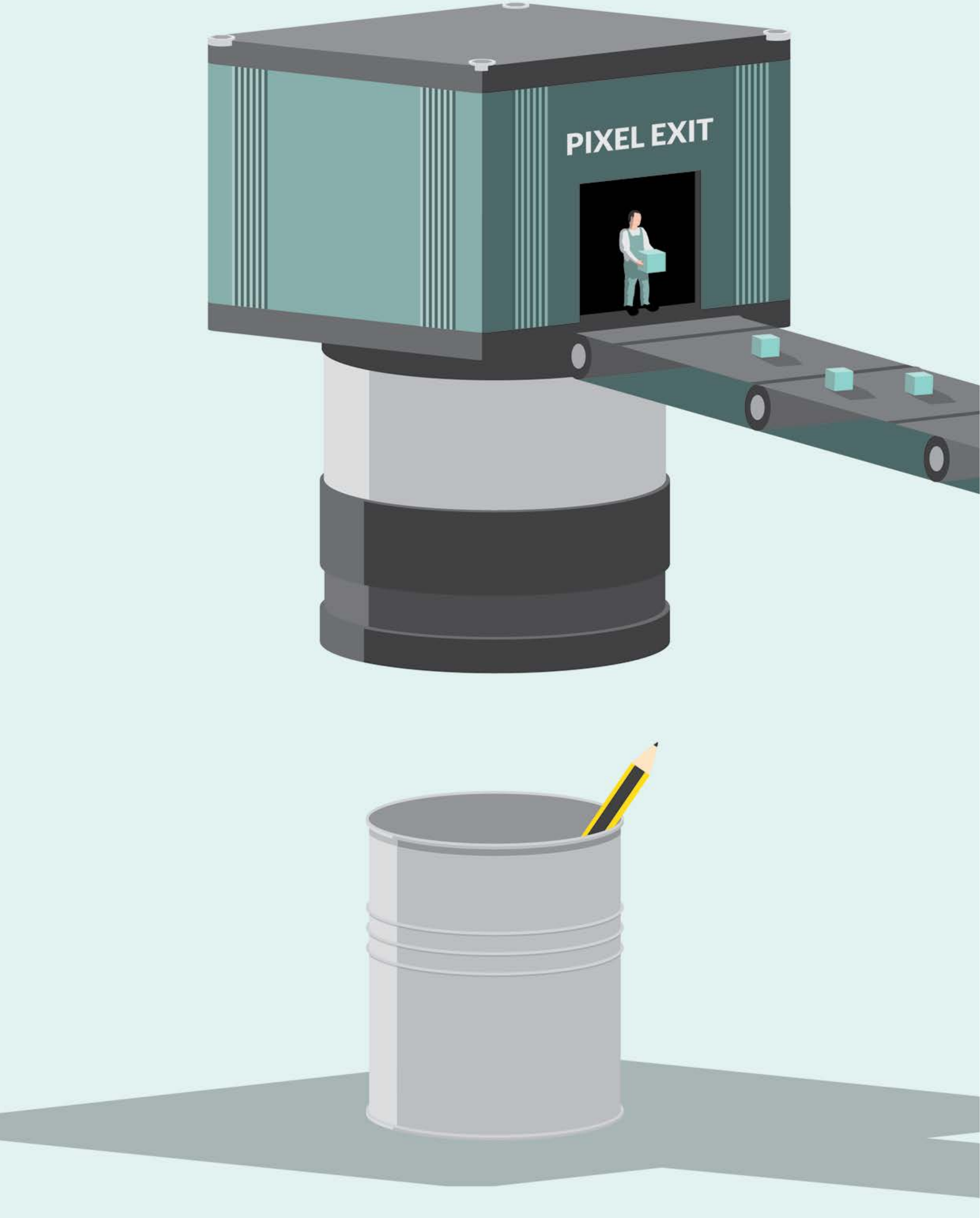
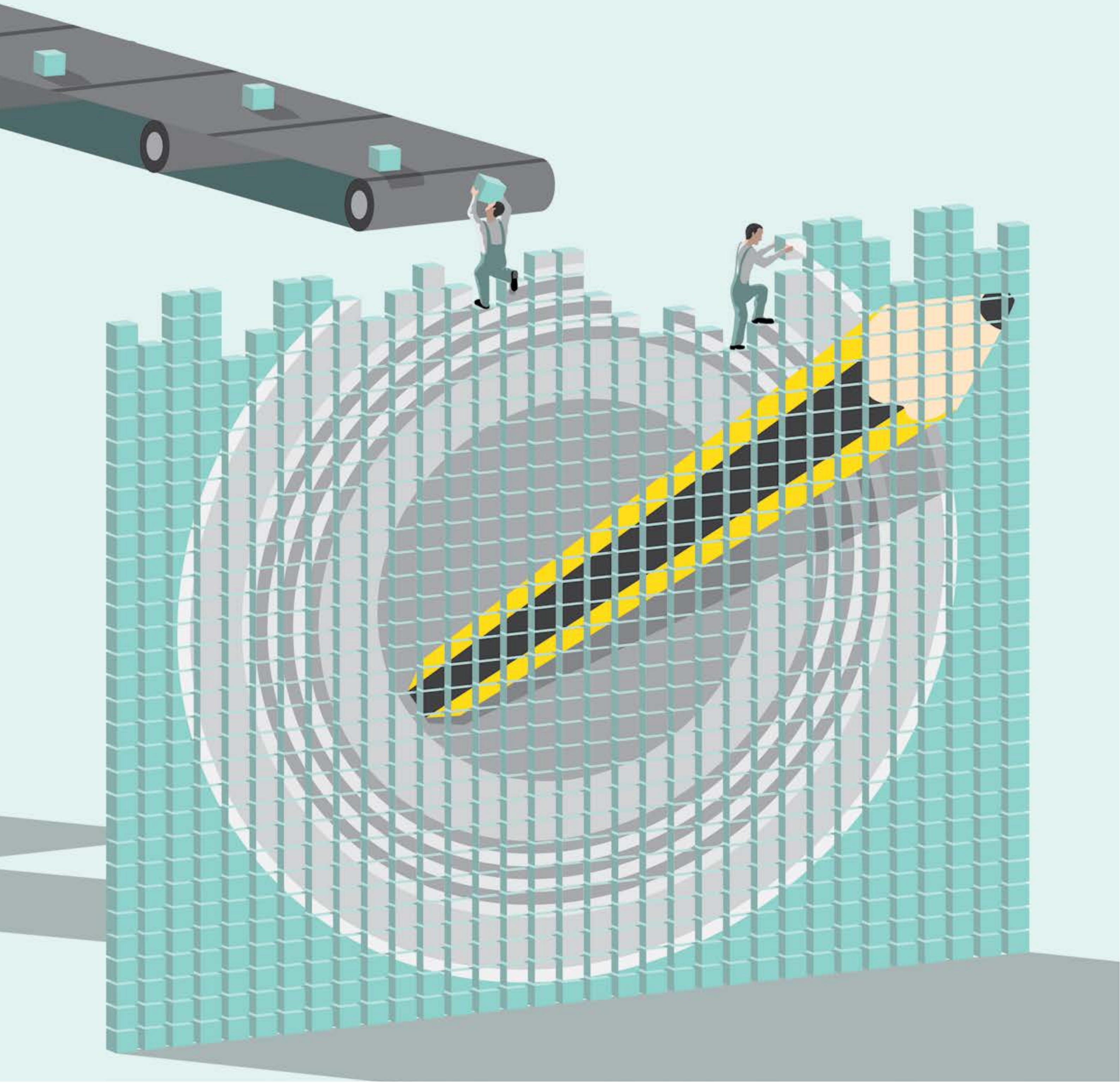


Fig. 1: Working principle of a camera obscura

Fig. 2: Camera obscura "View of Dotel de Ville, Paris, France, 2015"
Photo by Abelardo Morell ©

Camera technology has hugely improved in the last decades, since the development of Charge Coupled Device (CCD) and, more recently, of CMOS technology. Previous standard systems, such as vacuum tube cameras, have been discontinued.
The improvements in image resolution and acquisition speed obviously also improved the quality and speed of machine vision cameras.


PIXEL EXIT



Camera types

Matrix and Line scan cameras

Cameras used in machine vision applications can be divided in two groups: area scan cameras (also called matrix cameras) and line scan cameras. The first are simpler and less technically demanding, while the latter are preferred in some situations where matrix cameras are not suitable. Area scan cameras capture 2-D images using a certain number of active elements (pixels), while line scan cameras sensors are characterized by a single array of pixels.

Sensor sizes and resolution

Sensor sizes (or formats) are usually designated with an imperial fraction value – i.e. 1/2", 2/3". However, the actual dimensions of a sensor are different from the fraction value, which often causes confusion among users. This practice dates back to the 50's at the time of TV camera tubes and is still the standard these days. Also, it is always wise to check the sensor specifications, since even two sensors with the same format may have slightly different dimensions and aspect ratios. Spatial resolution is the number of active elements (pixels) contained in the sensor area: the higher the resolution, the smaller the detail we can detect on the image.
Suppose we need to inspect a 30 x 40 mm FoV, looking for 40*40 µm defects that must be viewed on at least three pixels. There can be 30*40/(0.04*0.04) = 0.75x10^6 defects. Assuming a minimum of 3 pixels are required to see a defect, we need a camera with at least 2.25 MP pixels. This gives the minimum resolution required for the sensor, although the whole system resolution (also including the lens resolution) must always be assessed. Table 1 gives a brief overview of some common sensor dimensions and resolutions. It is important to underline that sensors can have the same dimensions but different resolution, since the pixel size can vary. Although for a given sensor format smaller pixels lead to higher resolution, smaller pixels are not always ideal since they are less sensitive to light and generate higher noise; also, the lens resolution and pixel size must always be properly matched to ensure optimal system performances.

Sensor type		1/3"	1/2"	2/3"	1"	4/3"	4 K (linear)	8 K (linear)	12 K (linear)
Sensor size	(mm)	4.80 x 3.60	6.40 x 4.80	8.45 x 7.07	12.8 x 9.64	18.1 x 13.6	28.7	41	64
Pixel size	(µm)	5	5	5	5	5	7	5	5.3
Resolution	(mm)	960 x 720	1280 x 960	1690 x 1414	2560 x 1928	3620 x 2720	4000	8000	12000
Resolution	(Pixel)	0.6 M	1.2 M	2.5 M	5 M	10 M	4 K	8 K	12 K

Table 1: Examples of common sensor sizes and resolutions

Sensor types: CCD and CMOS

The most popular sensor technologies for digital cameras are CCD and CMOS.

CCD (charged-couple device) sensors consist of a complex electronic board in which photosensitive semiconductor elements convert photons (light) into electrons. The charge accumulated is proportional to the exposure time.

Light is collected in a potential well and is then released and read out in different ways (cf. Fig. 3). All architectures basically shift the information to a register, sometimes passing through a passive area for storage.

The charge is then amplified to a voltage signal that can be read and quantified.

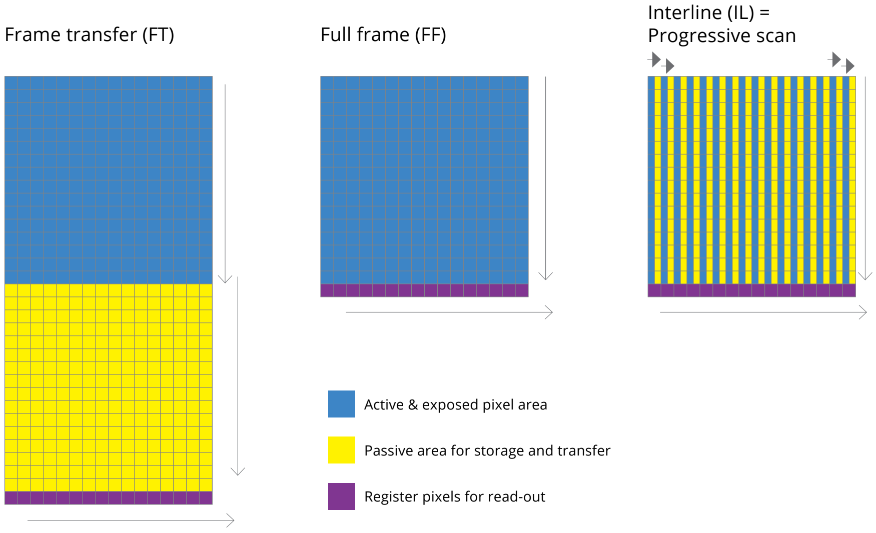


Fig. 3: CCD architectures

CMOS (complementary metal-oxide semiconductor) sensors are conceptually different from CCD sensors, since the readout can be done pixel by pixel rather than in sequential mode. In fact, signal is amplified at each pixel position, allowing you to achieve much higher frame rates and to define custom regions of interest (ROIs) for the readout.
CMOS and CCD sensors were invented around the same time and, although historically CCD technology was regarded as superior, in recent years CMOS sensors have caught up in terms of performance.

Global and rolling shutter (CMOS). In rolling shutter CMOS sensors, the acquisition is progressive from the upper to the last row of pixels, with up to 1/frame rate time difference between the first and the last row.
Once the readout is complete, the progressive acquisition process can start again. If the object is moving, the time difference between pixels is clearly visible on the image, resulting in distorted objects (see Fig. 4). Global shutter is the acquisition method in which all pixels are activated simultaneously, thus avoiding this issue.

Fig. 4: Rolling shutter effect

Sensor and camera features

Sensor characteristics

Pixel defects can be of three kinds: hot, warm and dead pixels. Hot pixels are elements that always saturate (give maximum signal, e.g. full white) whichever the light intensity is. Dead pixels behave the opposite, always giving zero (black) signal. Warm pixels produce random signal. These kinds of defects are independent of the intensity and exposure time, so they can be easily removed – e.g. by digitally substituting them with the average value of the surrounding pixels.

Noise. There are several types of noise that can affect the actual pixel readout. They can be caused by either geometric, physical and electronic factors, and they can be randomly distributed as well as constant. Some of them are presented below:

- Shot noise is a consequence of the discrete nature of light. When light intensity is very low - as it is considering the small surface of a single pixel – the relative fluctuation of the number of photons in time will be significant, in the same way as the heads or tails probability is significantly far from 50% when tossing a coin just a few times. This fluctuation is the shot noise.
- Dark current noise is caused by the electrons that can be randomly produced by thermal effect. The number of thermal electrons, as well as the related noise, grows with temperature and exposure time.
- Quantization noise is related to the conversion of the continuous value of the original (analog) voltage value to the discrete value of the processed (digital) voltage.
- Gain noise is caused by the difference in behavior of different pixels (in terms of sensitivity and gain). This is an example of 'constant noise' that can be measured and eliminated.

Sensitivity is a parameter that quantifies how the sensor responds to light. Sensitivity is strictly connected to quantum efficiency, that is the fraction of photons effectively converted into electrons.

Dynamic range is the ratio between the maximum and minimum signal that is acquired by the sensor. At the upper limit, pixels appear to be white for every higher value of intensity (saturation), while pixels appear black at the lower limit and below.
The dynamic range is usually expressed by the logarithm of the min-max ratio, either in base-10 (decibel) or base-2 (doublings or stops), as shown below. Human eyes, for example, can distinguish objects both under starlight and on a bright sunny day, corresponding to a 90 dB difference in intensity. This range, though, cannot be used simultaneously, since the eye needs time to adjust to different light conditions.
A good quality LCD has a dynamic range of around 1000:1, and some of the latest CMOS sensors have measured dynamic ranges of about 23 000:1 (reported as 14.5 stops).

Factor	Decibels	Stops
1	0	0
2	3.01	1
3.16	5	1.66
4	6.02	2
10	10	3.32
32	15.1	5
100	20	6.64
1024	30.1	10
10 000	50	13.3
1 000 000	60	19.9
1 073 741 824	90.3	30
10 000 000 000	100	33.2

Table 2: Dynamic range D, Decibels (10 log D) and Stops (log2 D)

SNR (signal-to-noise ratio) considers the presence of noise, so that the theoretical lowest grey value as defined by the dynamic range is often impossible to achieve. SNR is the ratio between the maximum signal and the overall noise, measured in dB. The maximum value for SNR is limited by shot noise (that depends on the physical nature of light, and is this inevitable) and can be approximated as

$$SNR_{max} = \mathit{sqrt}\ [\ \textit{maximum saturation capacity in electrons of a single pixel}\ ]$$

SRN gives a limit on the grey levels that are meaningful in the conversion between the analog signal (continuous) and the digital one (discrete). For example, if the maximum SNR is 50 dB, a good choice is a 8 bit sensor, in which the 256 grey levels corresponds to 48 dB. Using a sensor with higher grey levels would mean registering a certain degree of pure noise.

Spectral sensitivity is the parameter describing how efficiently light intensity is registered at different wavelengths. Human eyes have three different kinds of photoreceptors that differ in sensitivity to visible wavelengths, so that the overall sensitivity curve is the combination of all three. Machine vision systems, usually based on CCD or CMOS cameras, detect light from 350 to 900 nm, with the peak zone being between 400 and 650 nm. Different kinds of sensor can also cover the UV spectrum or, on the opposite side, near infrared light, before going to drastically different technology for far wavelengths such as SWIR or LWIR.

EMVA Standard 1288

The different parameters that describe the characteristics and quality of a sensor are gathered and coherently described in the EMVA standard 1288. This standard illustrates the fundamental parameters that must be given to fully describe the real behavior of a sensor, together with the well-defined measurement methods to get these parameters.

1288 EMVA Standard Compliant

The standard parameters are:

- Sensitivity, linearity of signal versus light intensity and noise
- Dark current (temperature dependence: optional)
- Sensor non-uniformity and defect pixels
- Spectral sensitivity (optional)

	Sensitivity, linearity and noise	Dark current	Sensor non-uniformity and defect pixel	Spectral sensitivity
Measuring procedure	Test measuring amount of light at increasing exposure time, from closed shutter to saturation. Quantity of light is measured (e.g. photometer)	Measured from dark images taken at increasing exposure times. Since dark current is temperature dependent, behavior at different T can be given	A number of images are taken without light (to see hot pixels) and at 50% saturation. Parameters of spatial distortion are calculated using Fourier algorithms	Images taken at different wavelengths
Result	Quantum efficiency (photons converted over total incoming photons ratio in %)	Signal registered in absence of light, in electrons per second	Dark and bright signal non-uniformity	Spectral sensitivity curve
	Temporal dark noise, in electrons (e-)		Dark and bright spectrograms and (logarithmic) histograms	
	Absolute sensitivity threshold (minimum number of photons to generate a signal)			
	Dynamic range, in stops			
	SNR, in stops			
	Saturation capacity (maximum number of electrons at saturation)			

Camera Parameters

Exposure time is the amount of time in which light is allowed to reach the sensor. The higher this value, the higher the quantity of light represented on the resulting image.
Increasing the exposure time is the first and easiest solution when light is not enough but it is not free from issues: first, noise always increases with the exposure time; also, blur effects can appear when dealing with moving objects. In fact, if the exposure time is too high, the object will be impressed on a number of different pixels, causing the well-known ‘motion blur’ effect (see Fig. 5).
On the opposite side, too long exposure times can lead to overexposure – namely, when a number of pixels reach maximum capacity and thus appear to be white, even if the light intensity on each pixel is actually different

Fig. 5: Motion blur effect

Frame rate. This is the frequency at which a complete image is captured by the sensor, usually expressed in frames per second (fps). It is clear that the frame rate must be adjusted to the application: a line inspecting 1000 bottles per minute must be able to take images with a minimum frame rate of 1000/60 = 17 fps.

Triggering. Most cameras give the possibility to control the beginning of the acquisition process, adjusting it to the application. A typical triggering system is one in which light is activated together with the image acquisition after receiving an input from an external device (e.g. position sensor).
This technique is essential when taking images of moving objects, in order to ensure that the features of interest are in the field of view of the imaging system.

Gain in a digital camera represents the relationship between the number of electrons acquired and the analog-to-digital units (ADUs) generated, i.e. the image signal. Increasing the gain means increasing the ratio between ADUs and electrons acquired, resulting in an apparent higher brightness of the image. Obviously, this process increases the image noise as well, so that the overall SNR will be unchanged.

Binning is the camera feature that combines the readout of adjacent pixels on the sensor, usually in rows/columns, more often in 2 x 2 or 4 x 4 squares (see Fig. 6).
Although resolution obviously decreases, there are a number of other features improving. For example, with 2x2 binning, resolution is halved, but sensitivity and dynamic range are increased by a factor of 4 (since the capacitiec of each potential well are summed), readout time is halved (frame rate doubled) and noise is quartered.

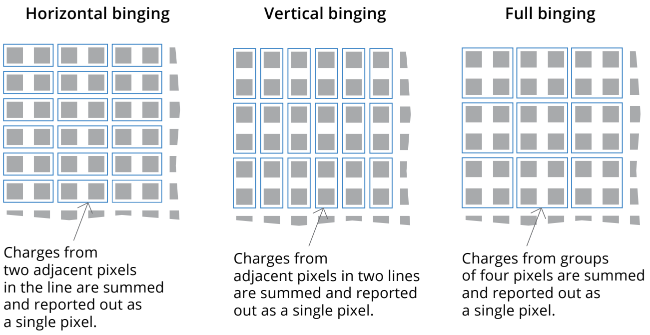


Fig. 6: Sensor binning

Digital camera interfaces

Camera Link

The Automated Imaging Association (AIA) standard, commonly known as Camera Link, is a standard for high-speed transmission of digital video. AIA standard defines cable, connector and camera functionality between camera and frame grabber.
Speed. Camera Link offers very high performance in terms of speed. It usually has different bandwidth configurations available, e.g. 255 MB/s, 510 MB/s and 680 MB/s. The bandwidth determines the ratio between image resolution and frame rate: a typical basic-configuration camera can acquire 1 Mpixel image at 50 frames/s or more; a full-configuration camera can acquire 4 Mpixel at more than 100 frames/s. Camera Link HS is the newer standard that can reach 300 MB/s on a single line, and up to 6 GB/s on 20 lines.
Costs. Camera Link offers medium- to high-performance acquisition, thus usually requiring more expensive cameras. Also, this standard requires a frame grabber in order to manage the hefty data load, not needed with other standards.
Cables. Camera Link standard defines a maximum length of 10 m for the cables; one cable is needed for basic configuration, where two are needed for full configuration cameras.
Power over cable. Camera Link offers a PoCL module (Power over Camera Link) that provides power to the camera. Also, several grabbers work with this feature.
CPU usage. Since Camera Link uses frame grabbers, which transfer images to a computer as stand-alone modules, this standard does not consume a lot of the system CPU.

CoaXPress

CoaXPress is the second standard, developed after Camera Link. It basically consists in power, data and control for the device sent through a coaxial cable.

Speed. A single cable can transmit up to 781.25 MB/s from the device to the frame grabber and 20 Mbit/s of control data from the frame grabber to the remote device, that is 5-6 times the GigE bandwidth. Some models can run also at half speed (390.625 MB/s). At present, up to 4 cables can be connected in parallel to the frame grabber, reaching a maximum bandwidth of approx. 1800 MB/s.

Costs. In the simplest case, CoaXPress uses a single coaxial line to transmit data, and coaxial cables are a simple and low-cost solution. On the other hand, a frame grabber is needed, i.e. an additional card must be installed, resulting in an additional cost on the system.

Cables. Maximum cable length is 40 m at full bandwidth, or 100 m at half bandwidth.

Power over cable. Voltage supply provided goes up to 13 W at 24 V, that is enough for many cameras.

CPU usage. CoaXPress, just like Camera Link, uses frame grabbers, which transfer images to computer as stand-alone modules, i.e. this standard is very light on consuming the system CPU.

GiG-E

Gig-E Vision is a camera bus technology that standardizes the Gigabit Ethernet, adding a 'plug and play' behavior (such as device discovery) to the latter. For its relatively high bandwidth, long cable length and diffused usage it is a good solution for industrial applications.

Speed. Gigabit Ethernet has a theoretical maximum bandwidth of 125 MB/s, that goes down to 100 MB/s when considering practical limitations. This bandwidth is comparable to FireWire standard and is second only to Camera Link.

Costs. System cost of GigE vision is moderate; cabling is cheap and it doesn't require a frame grabber.

Cables. Cabling length is the keystone of GigE standard, going up to 100 m. This is the only digital solution comparable to analog visioning in terms of cable length, and this feature has helped GigE Vision to replace analog e.g. in monitoring applications.

Power over cable. Power over Ethernet (PoE) is often available on GigE cameras. Nevertheless, some Ethernet cards cannot supply enough power, so that powered switch, hub, or a PoE injector must be used.

CPU usage. CPU loads of a GigE system can be different depending on drivers used. Filtered drivers are more generic and easer to create and use, but operate on data packets at high level, affecting the system CPU. Optimized drivers are specifically written for a dedicated network interface card, that working at lower lever affects poorly the system CPU load.

USB 3.0

The USB (Universal Serial Bus) 3.0 standard is the second revision of USB standard, developed for computer communication. Building on USB 2.0 standard, it provides a higher bandwidth and up to 4.5 W of power.

Speed. While USB 2.0 goes up to 60 MB/s, USB 3.0 speed can reach 400 MB/s, similar to the Camera Link standard used in medium configuration.

Costs. USB cameras are usually low cost; also, no frame grabber is required. For this reason, USB is the cheaper camera bus in the market.

Cables. Passive USB 3.0 cable has a maximum length of about 7 meters, and active USB 3.0 cable can reach up to 50 m with repeaters.

Power over cable. USB 3.0 offers power up to 4.5 W that allows to get rid of a separate power cable.

CPU usage. USB 3.0 Vision permits image transfer directly into PC memory, without CPU usage.

GenIcam Standard

The GenICam standard (GENeric Interface for CAMeras) is meant to provide a generic software interface for all cameras, independently from cameras hardware. Some of the new technology standard, anyway, are based on GenICam (es. Camera Link HS, CoaXPress, USB3 Vision).

GenICam standard purpose is to provide a 'plug and play' feature for every image system. In consists in three modules that help solving main tasks in machine vision filed in a generic way:

- GenApi: using a description file (XML), camera configuration and access-control is possible
- Standard Feature Naming Convention (SFNC): these are recommended names for common features in cameras to reach the goal of interoperability
- GenTL: describes the transport layer interface for enumerating cameras, grabbing images and transporting them to the user interface