Image Processing and Computer Vision

Prof. Giuseppe Lisanti

giuseppe.lisanti@unibo.it

Images...

need a mechanism to direct this light

- An imaging device gathers the light reflected by 3D objects to create a 2D representation of the scene (i.e. the image)

from 2D ot 3D

- In computer vision we basically try to invert such a process, so as to infer knowledge on the objects from one or more digital images

- Image formation and acquisition process:
  - The geometric relationship between scene points and image points
  - The radiometric relationship between the brightness of image points and the light emitted by scene points
  - The image digitization process

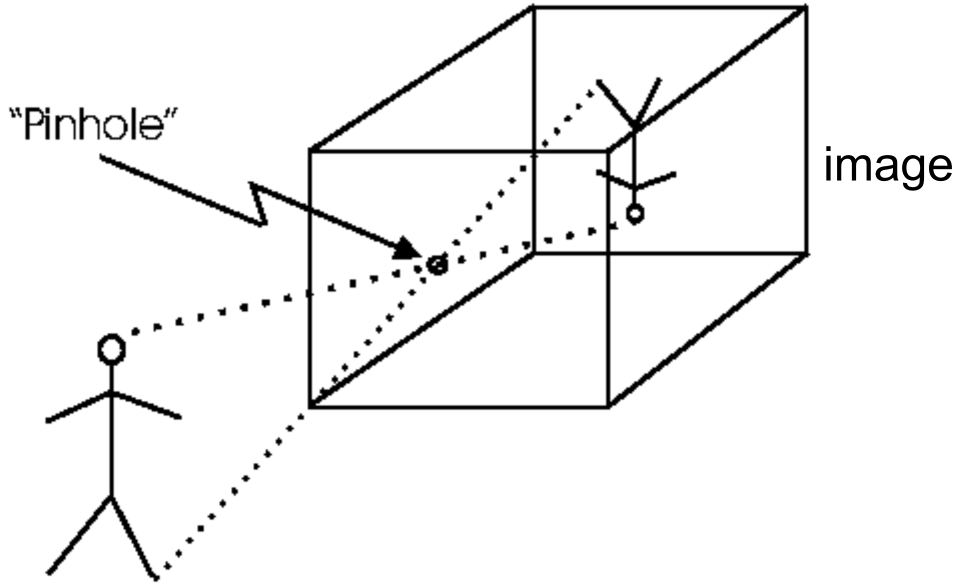
Pinhole camera model

- The “pinhole camera” is the simplest imaging device: light goes through the very small pinhole and hits the image plane

  only one light ray for each point

- Geometrically, the image is achieved by drawing straight rays from scene points through the hole up to the image plane

- Its remarkably simple geometrical model turns out to be a good approximation of the geometry of image formation in most modern imaging devices
  - however, useful images can hardly be captured by means of a pinhole camera



Perspective Projection model

- The geometric model of image formation in a pinhole camera is known as Perspective Projection

Image Plane: it is the area within the camera where the focused light from the scene converges to form an image.

M : scene point

m : corresponding image point

I : image plane

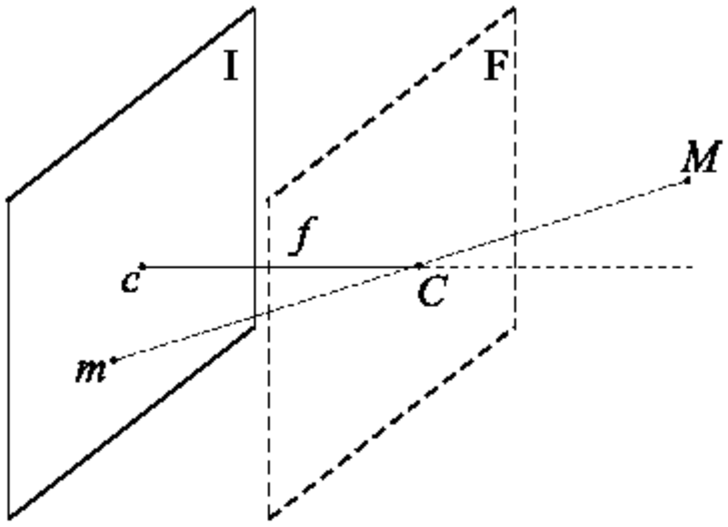
C : optical centre (pin hole)

Optical axis: line through C and orthogonal to I

c : intersection between optical axis and image plane (image centre or piercing point)

f : focal length

F : focal plane



in the Pinhole camera model, C that is the img centre is the pinhole

- We want to find a relationshitp between 3D and 2D points (between M and m)

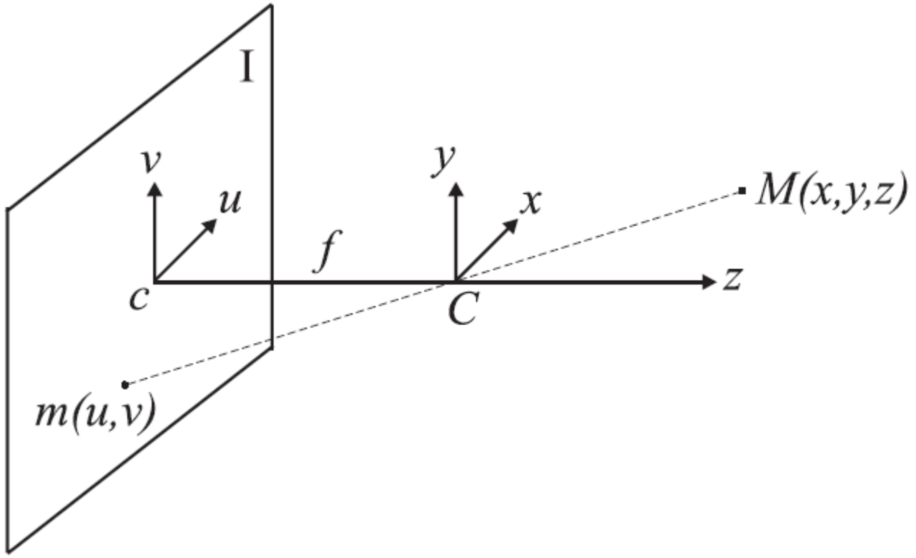
Perspective Projection

we need a coordinate system

- Given the reference frame in figure:
  - "u" is the horizontal axis in the image plane
  - "v" is the vertical axis in the image plane
  - "X" and "Y" are the respective axis in the 3D reference system

camera reference system

2D reference system

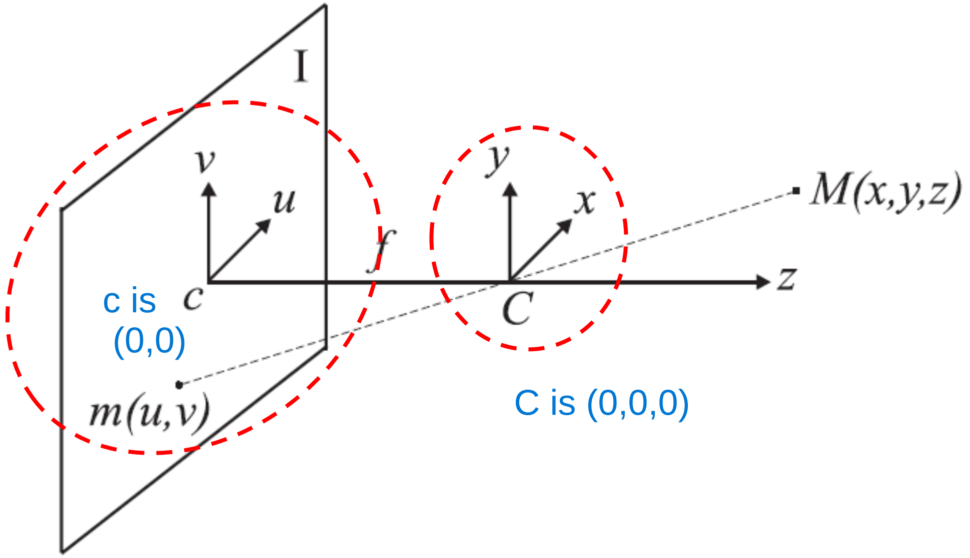


Focal length: it refers to the distance between the lens's optical center C, and the image plane when the lens is focused at infinity

"focused at infinity" means that the lens is adjusted to capture distant objects at their sharpest point without requiring any further focusing.

Perspective Projection

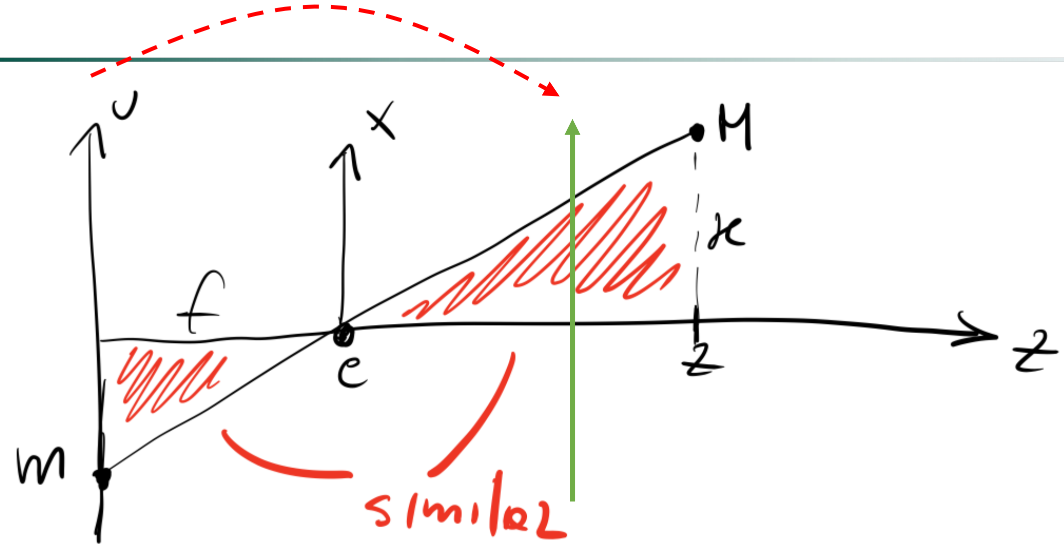
- Given the reference frame in figure:
  - "u" is the horizontal axis in the image plane
  - "v" is the vertical axis in the image plane
  - "X" and "Y" are the respective axis in the 3D reference system
    - called ***camera reference system***, because it is "attached" to the camera
  - For the perspective model these axis must be parallel
  - The equations to map scene points into their corresponding image points are defined as follows:



$$\frac{u}{x} = -\frac{f}{z} \Rightarrow u = -x\frac{f}{z} \qquad \frac{v}{y} = -\frac{f}{z} \Rightarrow v = -y\frac{f}{z}$$

Perspective Projection

obtained this thanks to the concept of traingle similarity



$$\frac{u}{x} = -\frac{f}{z} \Longrightarrow u = -x\frac{f}{z}$$

$$\frac{v}{y} = -\frac{f}{z} \Longrightarrow v = -y\frac{f}{z}$$

$$\Longrightarrow \frac{u}{x} = \frac{v}{y} = -\frac{f}{z}$$

- The minus means the axis get inverted as happens in Pinhole camera model
- How do we get rid of the up-down and left-right inversions?
  - Change of sign => the image plane can be thought of as lying in front rather than behind the optical centre (in real world we do not get flipped images)

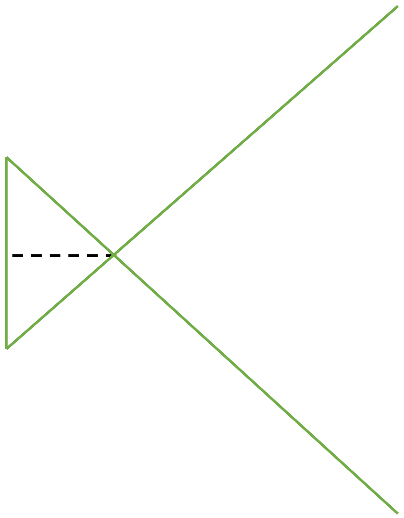
$$u = x\frac{f}{z}; \quad v = y\frac{f}{z}$$

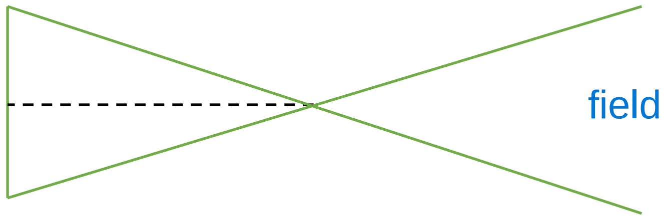
Perspective Projection

- Image coordinates are a scaled version of scene coordinates (function of depth)

$$u = x\frac{f}{z}; \quad v = y\frac{f}{z}$$

- How do they scale?
  - The farther the point the smaller the coordinates (object distant from the camera) if z increases
  - The larger the focal length the bigger the object in the image (and viceversa) if f larger





field of view is compressed

Perspective Projection

- Image coordinates are a scaled version of scene coordinates (function of depth)

$$u = x\frac{f}{z}; \quad v = y\frac{f}{z}$$

- How do they scale?
  - The farther the point the smaller the coordinates (object distant from the camera)
  - The larger the focal length the bigger the object in the image (and viceversa)

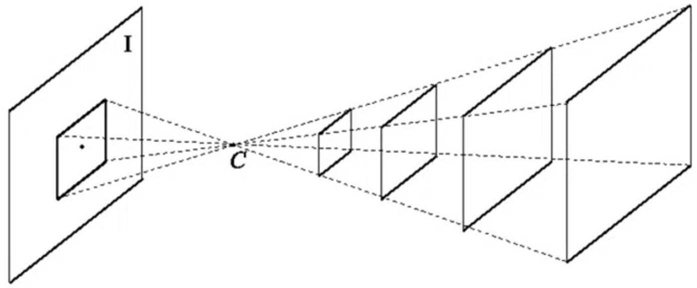
  also depending on the dimension of 3D object

- We scale the world inversely with respect to the depth

  very important

- The image formation process deals with mapping a 3D space onto a 2D space => loss of information

  we project a higher dimensional space into a lower one. therefore, loss of information is inevitable



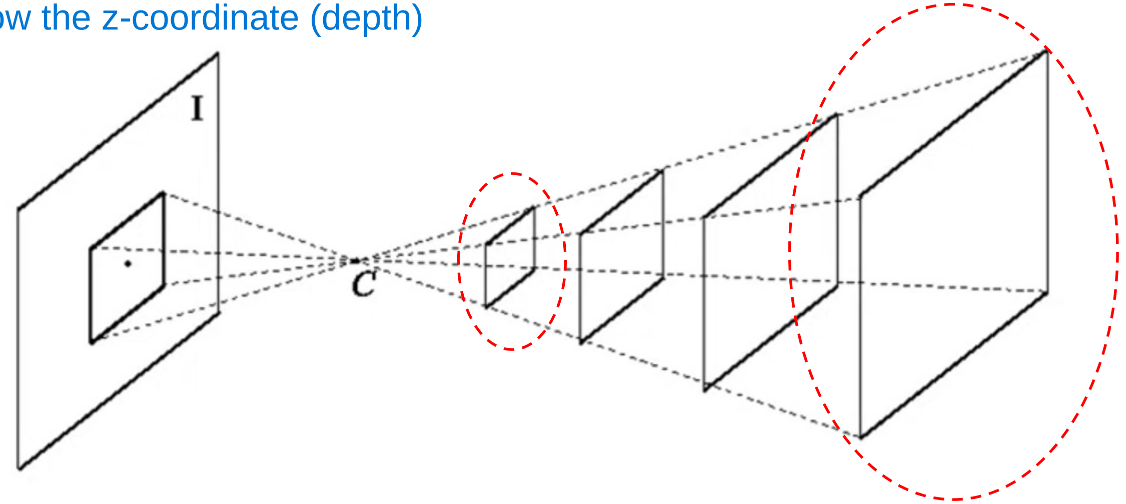
Perspective Projection

- The mapping is not a bijection: a given scene point is mapped into a unique image point

that is not true viceversa

  - a given image point is mapped onto a 3D line (i.e. the line through the point, m, and the optical centre, C).

because we do not know the z-coordinate (depth)



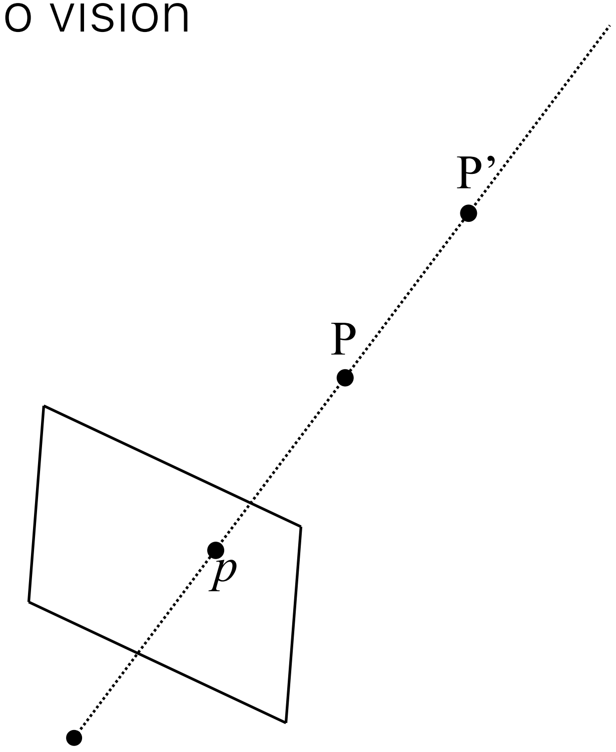
- Recovering the 3D structure of a scene from a single image is an ill-posed problem (the solution is not unique)
  - For an image point we can only state that its corresponding scene point lays on a line but cannot disambiguate a specific 3D point along such a line (i.e. we know nothing about the distance to the camera).

- How can we solve this?

Stereo images allow to infer 3D

how to retrieve the z-coordinate

- Solution: use multiple images (at least two) => stereo vision



Camera center is behind,
coordinate not flipped

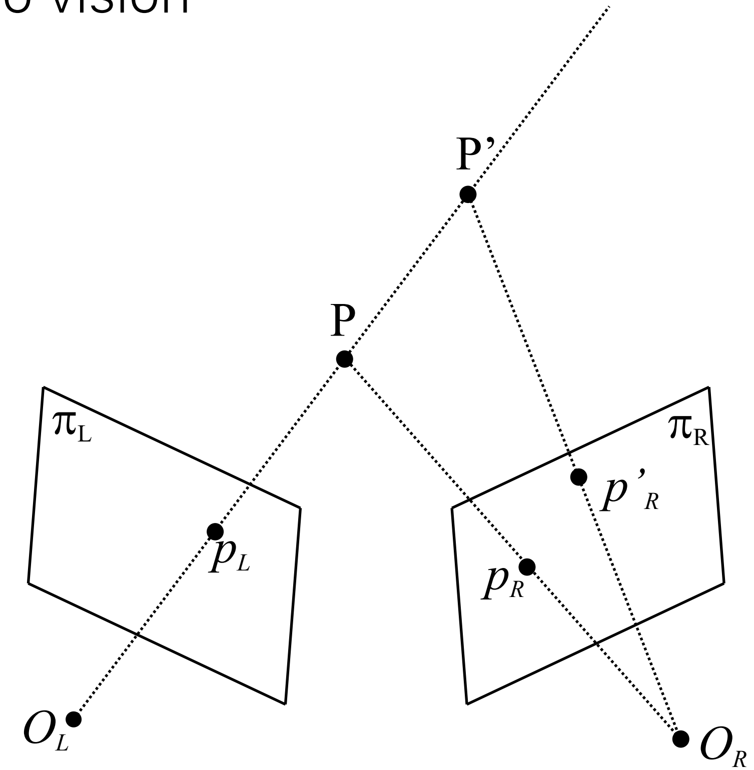
Stereo images allow to infer 3D

- Solution: use multiple images (at least two) => stereo vision
- The human visual system is a stereo vision system

- Stereo images allow to infer 3D
- Given correspondences, 3D information can be recovered easily by triangulation

the correspondences are given by an algorithm that tells us that pl and pr are the same points from the same object in the scene

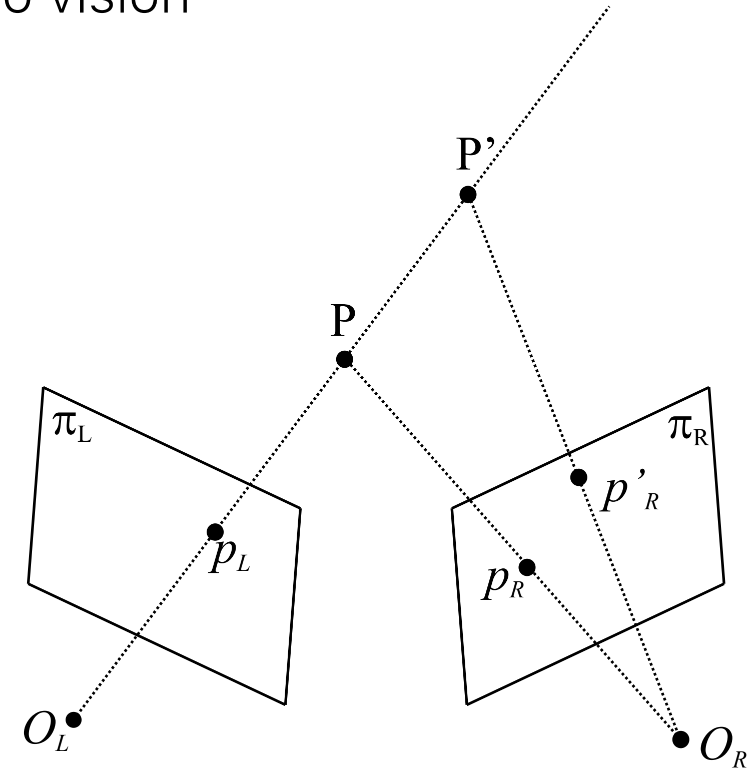
then, through triangulation, I can retrieve P



Stereo images allow to infer 3D

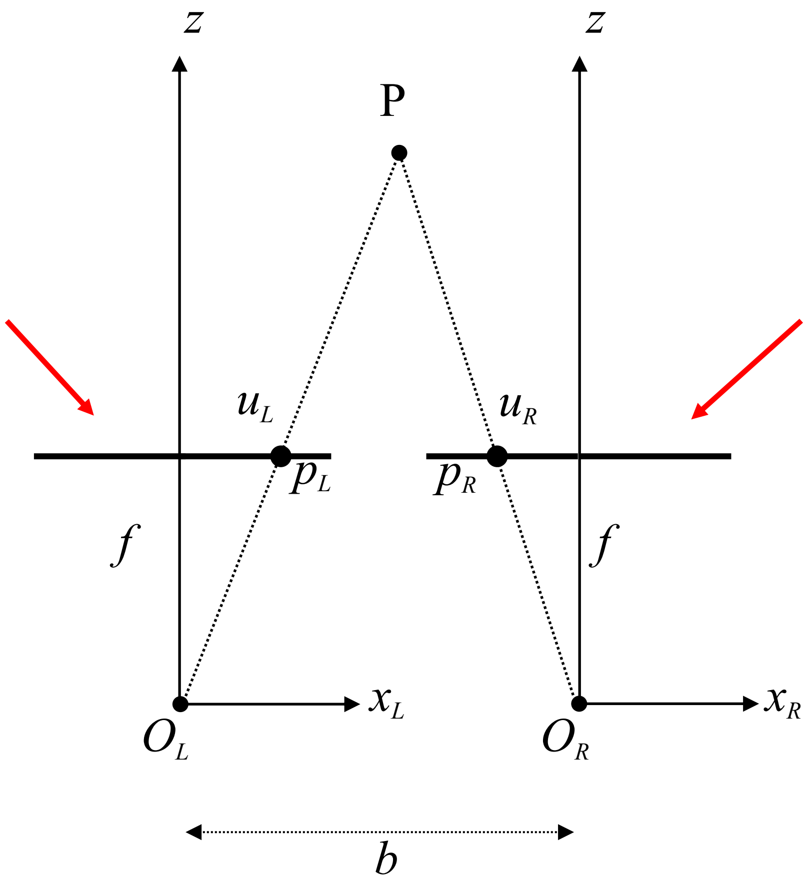
- Solution: use multiple images (at least two) => stereo vision
- The human visual system is a stereo vision system

- Stereo images allow to infer 3D
- Given correspondences, 3D information can be recovered easily by triangulation



Standard stereo geometry

- Assumptions:
    - Parallel (x, y, z) axes
    - Same focal length => coplanar image planes
- The transformation between the two reference frames is just a translation ($\boldsymbol{b}$), usually horizontal
- You need to **sense** two images at the very same moment
- You can put the two cameras "as you want" but they must observe the same object



$$P_L - P_R = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}$$

$$x_L - x_R = b$$

$$y_L = y_R = y$$

$$z_L = z_R = z$$

Standard stereo geometry

- The two cameras are displaced at a given quantity ***b*** called baseline

- Disparity: difference between the horizontal coordinates in the left and right images (horizontal displacement)

- Inverse relation: the larger the disparity the smaller depth, and vice versa
  - If a point has a large disparity it is close to the camera…

$$v_L = v_R = y \cdot f/z$$

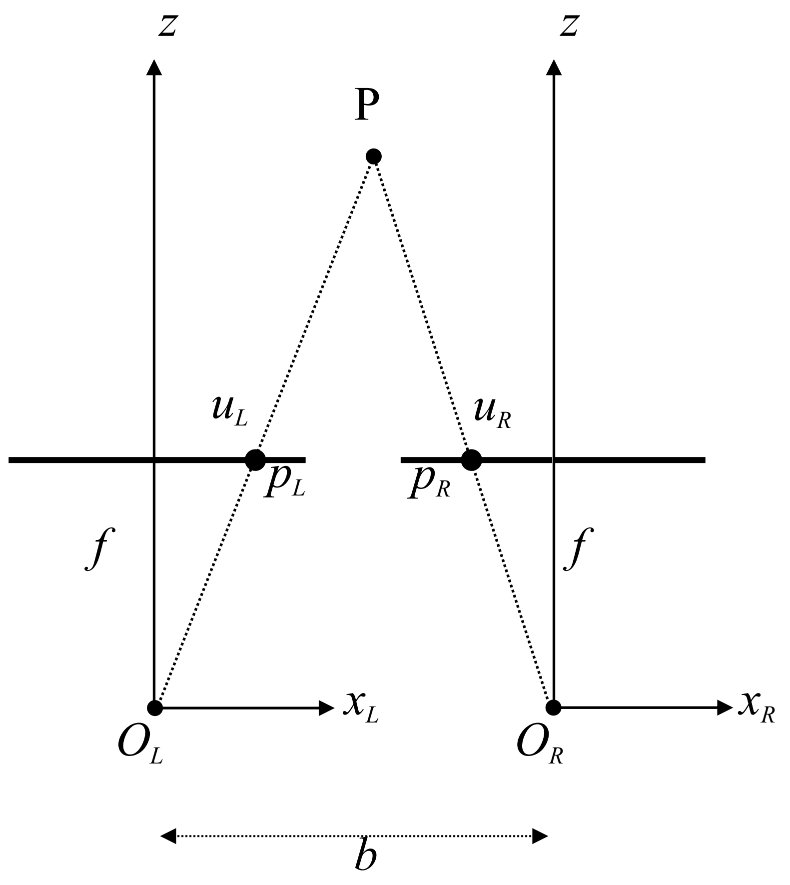
$$u_L = x_L \cdot f/z$$
$$u_R = x_R \cdot f/z$$

$$u_L - u_R = b \cdot f/z$$

$$u_L - u_R = d$$
(disparity)

$$d = b \cdot f/z \;\Rightarrow\; z = b \cdot f/d$$

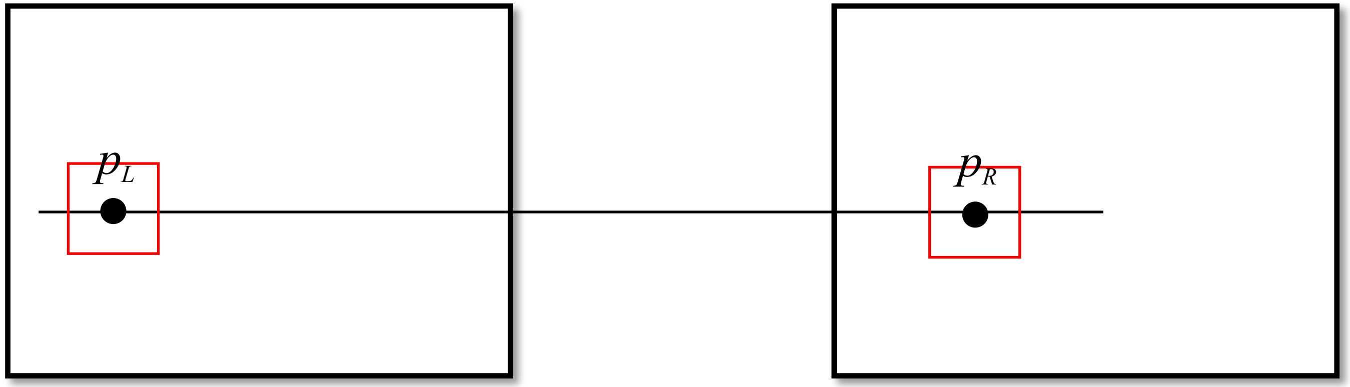
Fundamental relationship in stereo vision



Standard stereo geometry

$v_L = v_R = y \cdot f/z$

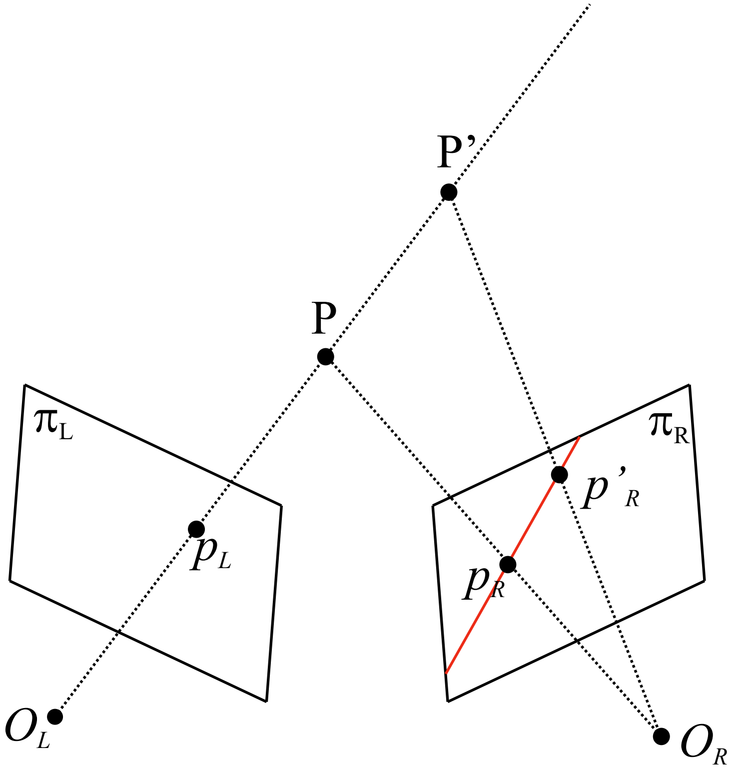
- Since we are given just two 2D images there is no info about the correspondence between two points in the two images
  - Given p_L , I want to find p_R
- We can search along horizontal lines if the two images are aligned, as said before
  - Stereo matching



  - For example you search for same color/pattern along the same line (better: use a window around the points across the line, block matching) search for intensity value on a small area around the pl and pr

Epipolar Geometry

- What if the two camera are no longer aligned? Do we need to search through the whole image?
  - We can project the line related to point P_L in the **right plane** and search across that line
    - The search space of the stereo correspondence problem is always 1D !
  - Issue: this projection can be computed only if the relation / transformation between the two cameras is known (relative mapping between the two cameras)
    - A roto-translation and the focal lenghts



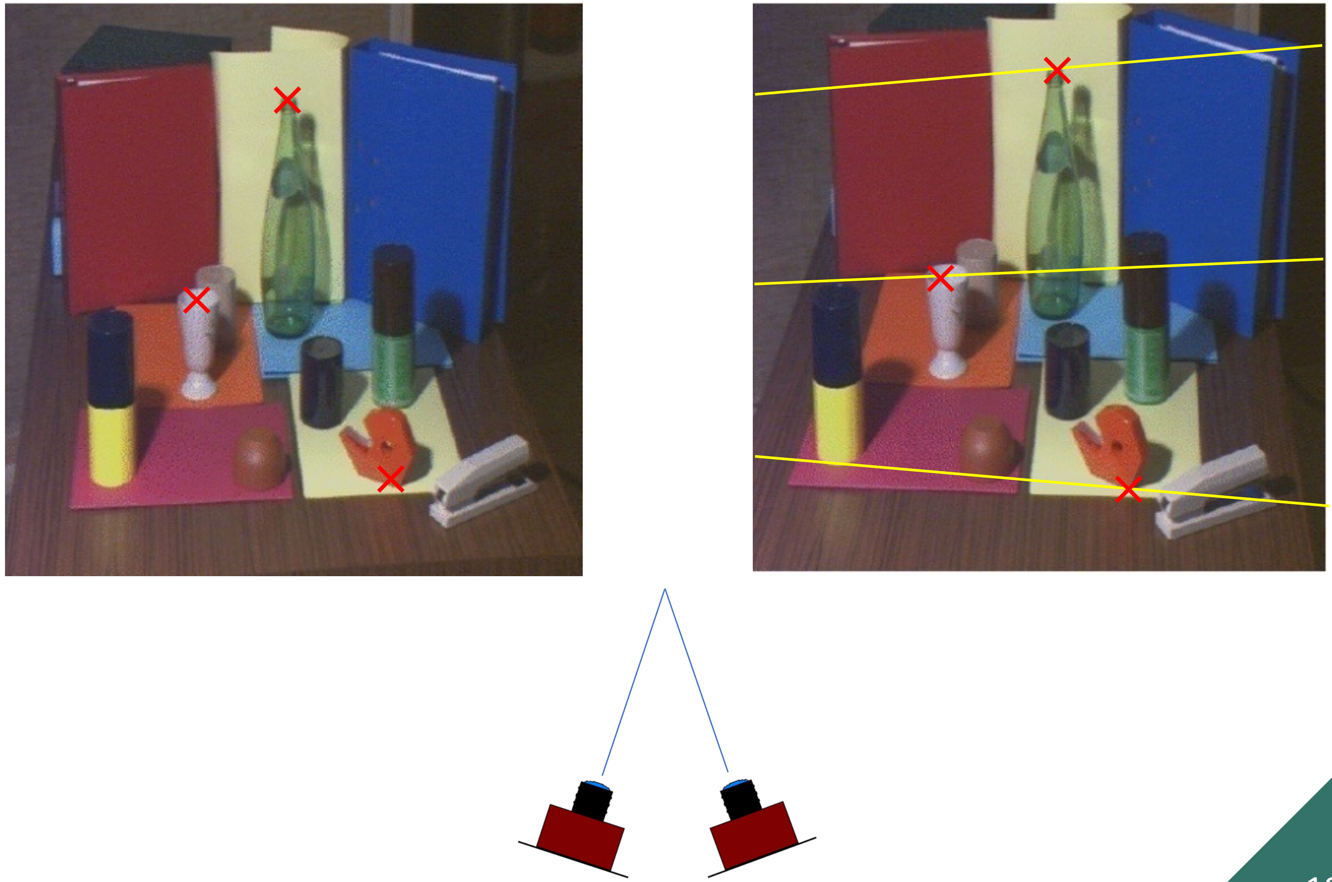
Epipolar line
(associated with p_L in π_R)

Epipolar Geometry

- It is almost impossible to build a stereo rig which is perfectly aligned horizontally

- Searching through oblique epipolar lines is awkward!
  - It is also computationally less efficient

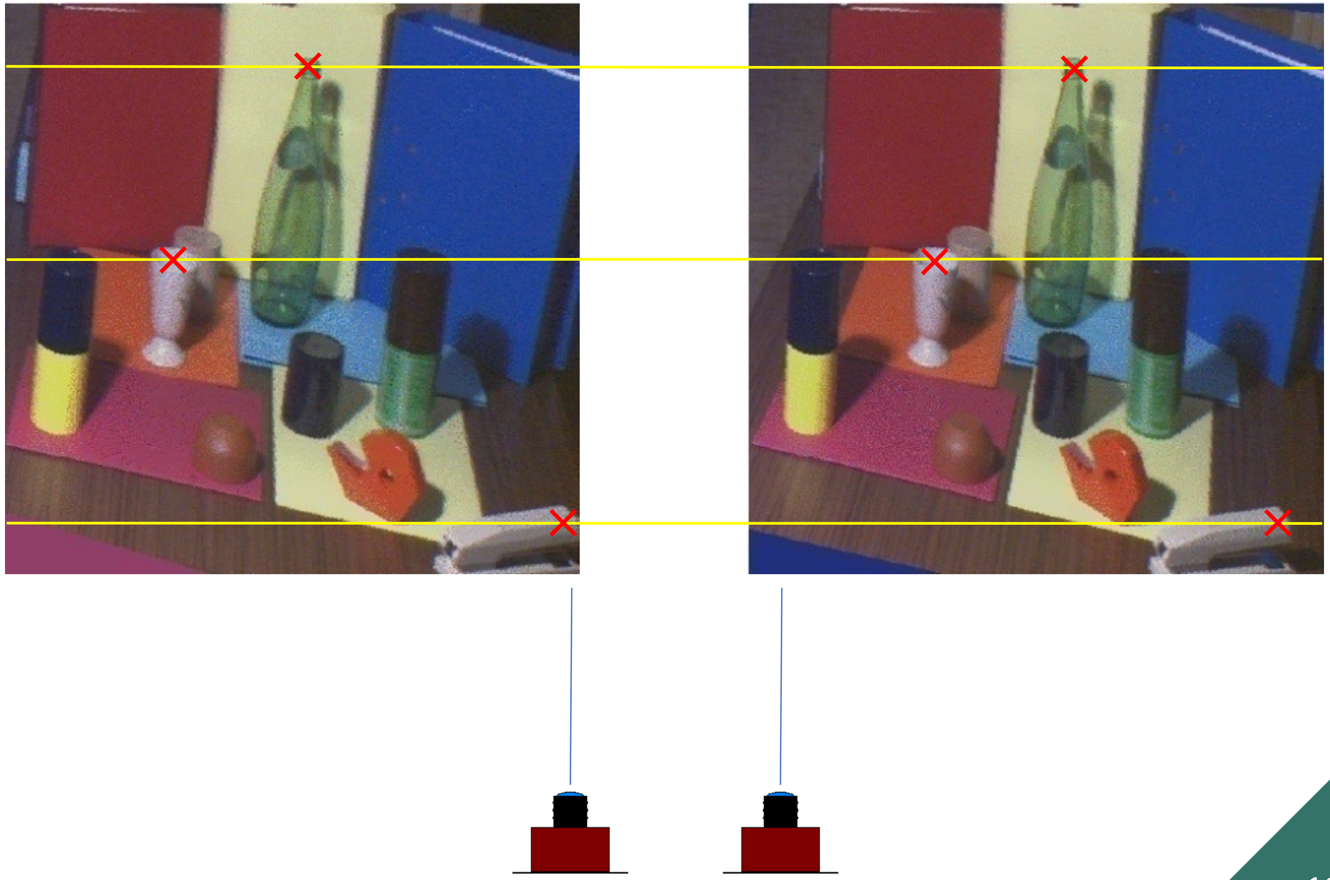
- What can we do?



Rectification

- What people do in practice is to convert epipolar geometry to standard geometry (Rectification / Warping)

- Warp the images as if they were acquired through a standard geometry (horizontal and collinear conjugate epipolar lines)
    - Compute and apply to both images a transformation (i.e. homography) known as **rectification**



Stereo Correspondence

- Given a point in one image (e.g. L) find that in the other image (R) which is the projection of the same 3D point. Such image points are called corresponding points.

Corresponding points look similar in the two images

Points farther away have a smaller disparity, while close points have a larger disparity

thumb example

KITTI (Karlsruhe Univ. & Toyota) Benchmark Suite

Properties of Perspective Projection

- The image of a 3D line segment of length L lying in a plane parallel to the image plane at distance z from the optical centre will exhibit a length given by:

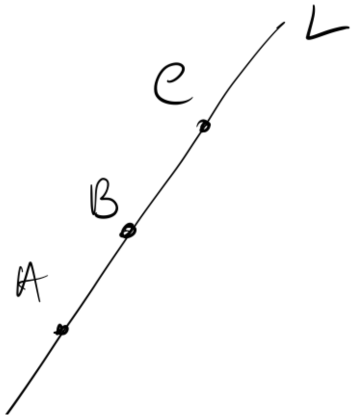
$$l = L\frac{f}{z}$$

- This relationship is more complicated for an arbitrarily oriented 3D segment, as its position and orientation need to be accounted for as well

- Nonetheless, for given position and orientation, length always shrinks alongside distance

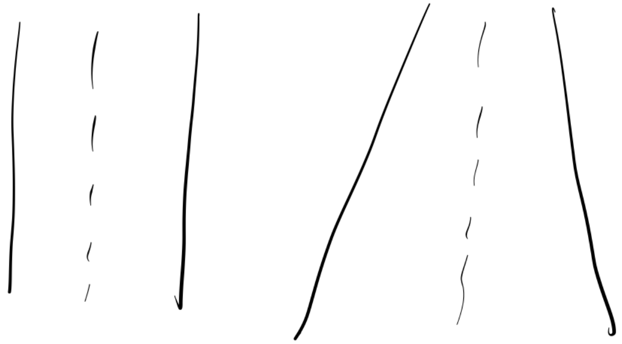
Properties of Perspective Projection

- Perspective projection maps 3D lines into image lines
  - Ratios of lengths are not preserved (unless the scene is planar and parallel to the image plane).



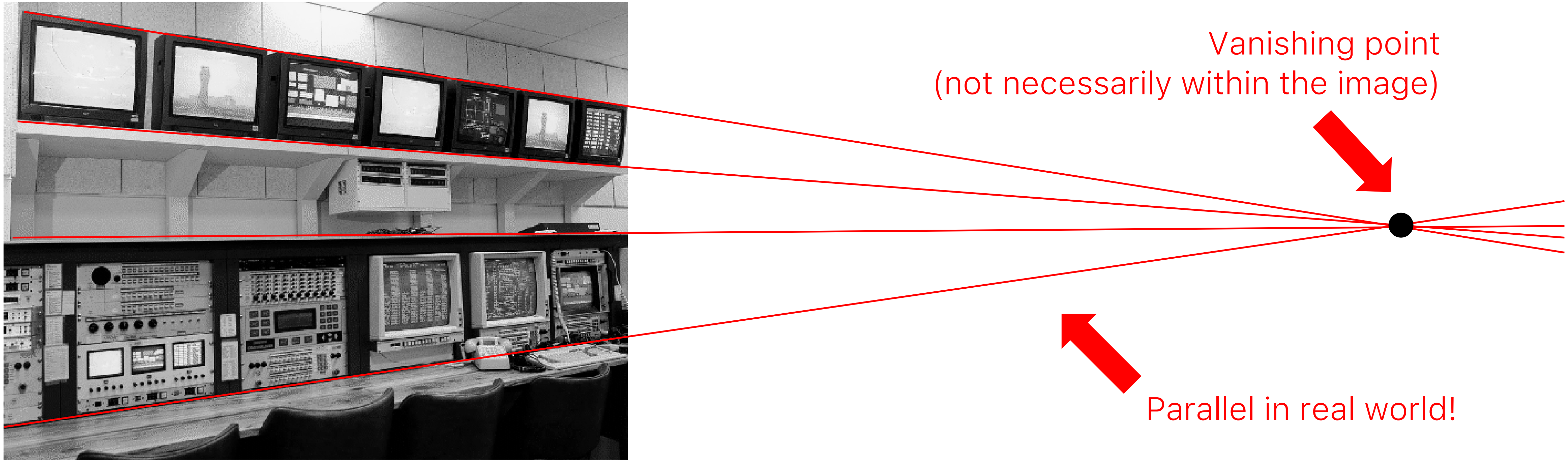
$$\frac{AB}{BC}$$

  - Parallelism between 3D lines is not preserved (except for lines parallel to the image plane)



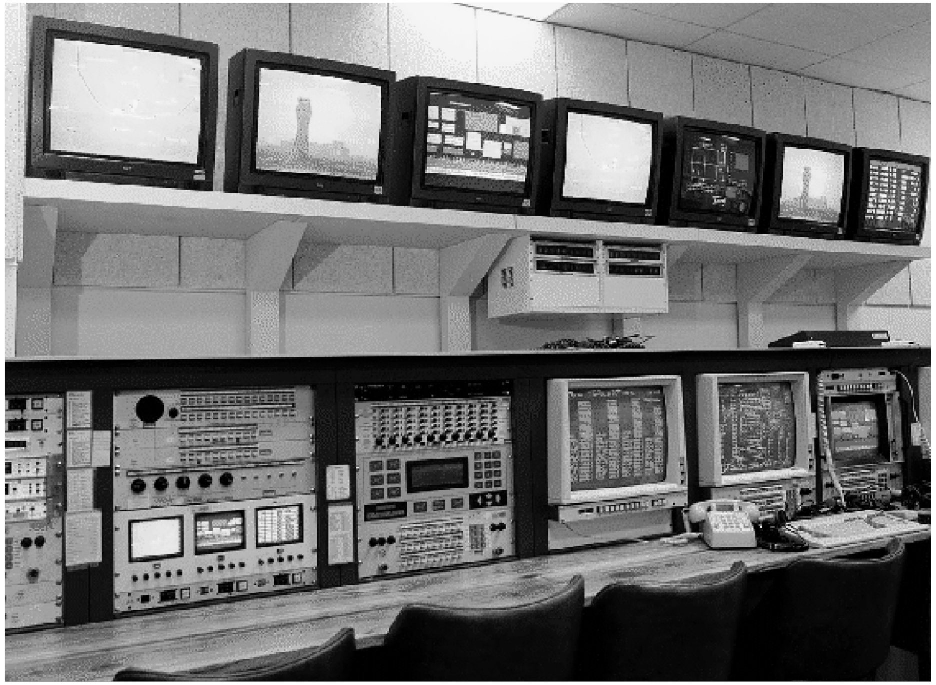
Vanishing point

- The images of parallel 3D lines intersect at a point



Vanishing point

- The images of parallel 3D lines intersect at a point, which is referred to as vanishing point.

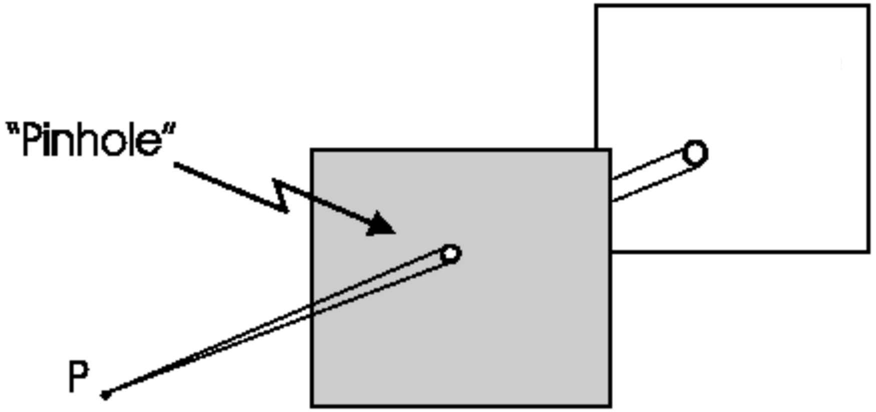


- If the lines are parallel to the image plane they meet at infinity

Depth of Field (DOF)

Depth of field (DOF) refers to the range of distance in a photograph or a scene that appears acceptably sharp and in focus. The depth of field is influenced by various factors such as the lens aperture, focal length, and v, the focusing dist.

- A scene point is on focus when all its light rays gathered by the camera hit the image plane at the same point



- In a pinhole device this happens to all scene points because of the very small size of the hole, so that the camera features an infinite Depth of Field (DOF)

in the pinhole model all the scene points, even at different distances, are on focus because of the small size of the hole. This means that we will have one ray light for each scene point

- The drawback is that such a small aperture allows gathering a very limited amount of light
- If a point is projected onto a circle instead of a point (bigger pinhole) the image is not sharp (not on focus) => you want to map all scene points to all image points   more ray lights for each scene point
- Getting sufficiently bright images requires very long exposure times
  - If we cannot gather through aperture we have to gather (integrate) through time…
  - Long time => moving object? => only static scenes can be acquired by a pinhole device to avoid motion blur

how to gather more light?
- increase aperture
- increase exposure time

Lenses

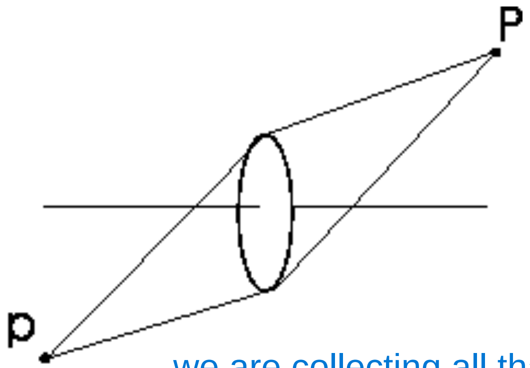
one drawback: the image point is at a fixed distance

- Use lenses to gather more light from a scene point and focus it on a single image point

  we can also fix a position thus avoiding motion blur

- This enables much smaller exposure times

  => avoid motion blur in dynamic scenes



we are collecting all that light
and concentrating onto one single point
(focus)

when we increase the aperture, in lenses, we gather more light, but as a trade-off, we no longer have an infinite depth of field

- DOF is no longer infinite => only points across a limited range of distances can be simultaneously on focus in a given image

  the DOF is limited, then only the scene points within a certain range of distances from the lense will appear on focus

- Cameras often feature complex optical systems, comprising multiple lenses

- We will consider the approximate model known as thin lens equation: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

Lenses

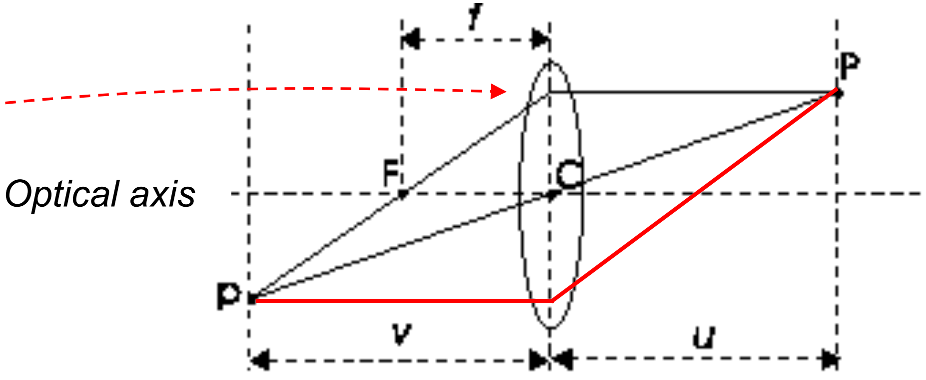
the complete model

all the rays reflected by P go through the lenses

all these rays get deflected towards one single point on the image plane

- To graphically determine the position of a focused image point we can leverage on two properties of thin lenses:
  1. Rays parallel to the optical axis are deflected to pass through *F*
  2. Rays through *C* are undeflected

they continue on the straight line represented by the optical axis



$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

P : scene point
p : corresponding focused image point
u : distance from P to the lens
v : distance from p to the lens
f : focal length (parameter of the lens)
C : centre of the lens
F : focal point (or focus) of the lens

meaning that all image point in the image plane are on focus, which means that the scene points are within the DOF

- If the image is on focus, the image formation process obeys to the perspective projection model:
  - the centre of the lens is the optical centre
  - the distance *v* acts as the effective focal length of the projection v = f

f is a parameter of the lens

u and v have to vary together

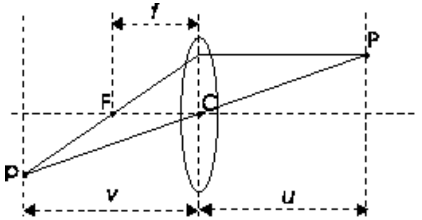
- By fixing the position of the image plane… and fixing the DOF, we can have only some elements on focus

Lenses

The focusing plane is the specific distance at which the lens is focused, and it corresponds to the point in the scene that is in focus

an object is on focus if it is positioned at or very close to the focusing plane, as long as it falls within the DOF, the depth of field.

in other words, an object is on focus when u corresponds to the focusing distance, which is the distance between the center of the lens and the focusing plane



- On one hand: choosing the distance of the image plane determines the distance at which scene points appear on focus in the image

fixing v, the distance from p to the lense

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \rightarrow u = \frac{vf}{v-f}$$

specified

- On the other hand: to acquire scene points at a certain distance we must set the position of the image plane accordingly:

fixing u, the distance from P to the lense

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \rightarrow v = \frac{uf}{u-f}$$

u and v are kind of inversely proportional

- Given the chosen position of the image plane, scene points both in front and behind the focusing plane will result out-of-focus, thereby appearing in the image as circles, known as *Circles of Confusion* or *Blur Circles*, rather than points
- The advantage of lenses is to have a small exposure time for capturing moving objects but we pay in terms of depth of field

Lenses

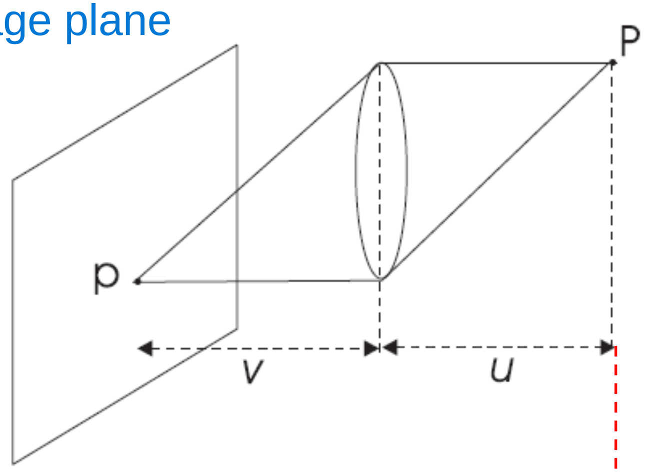
The size of these circles of confusion is influenced by factors such as the aperture size, the focal length of the lens, and the distance of the object from the focusing plane. A larger aperture (smaller f-number) results in a shallower depth of field, and the circles of confusion are more pronounced. Conversely, a smaller aperture (larger f-number) increases the depth of field, reducing the size of the circles of confusion.

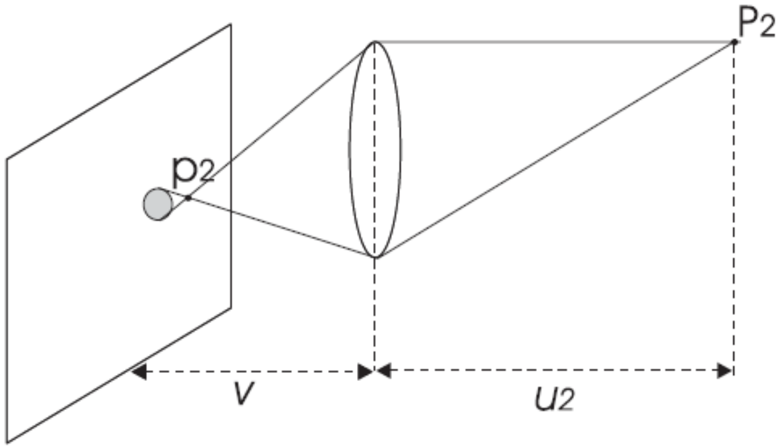
all the reflected raylights gets concentrated on the same point, p, in the image plane

P belongs to the focusing scene plane

p is sharp (on focus)

because u corresponds with the focusing distance

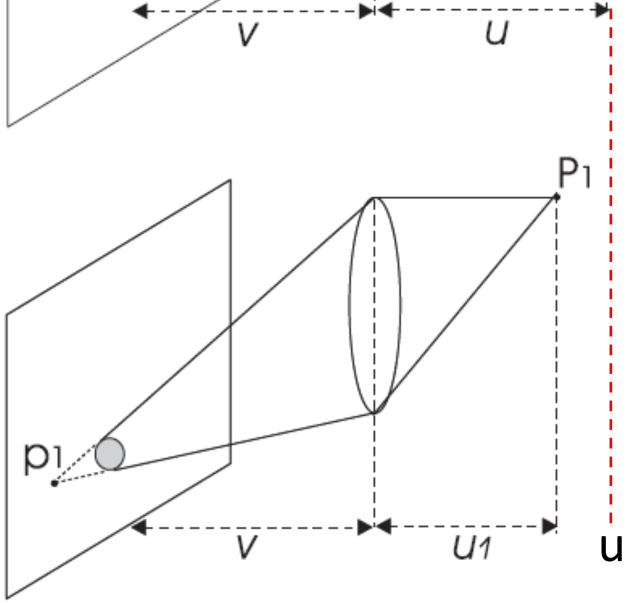




p_2 is in front of the image plane

on the other hand, here P2 is at a distance greater than the focusing distance; therefore, it is not on focus

here P1 does not lay in the focusing plane; in fact, it is placed at a distance smaller than the focusing distance; as a consequence of that, P is not on focus



p_1 is behind the image plane

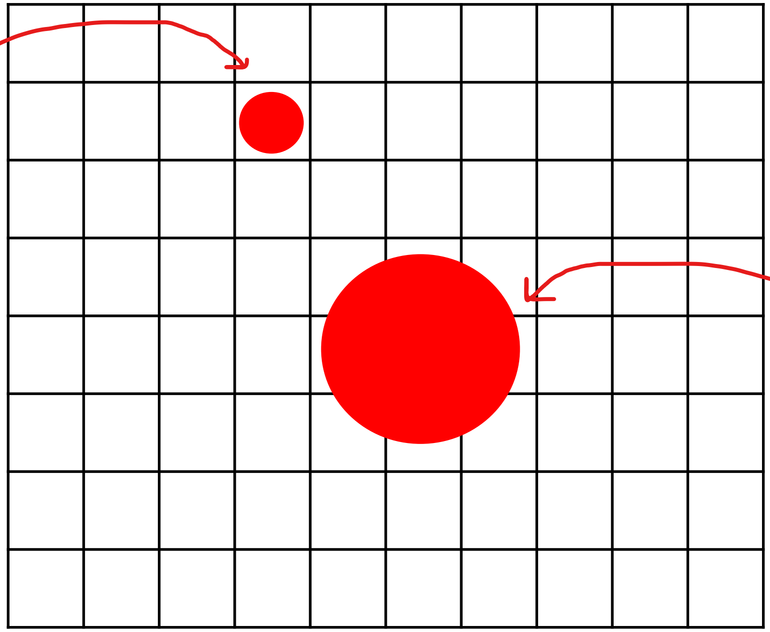
remember: u is the distance between the center of the lens and the point/object

Diaphragm

in our model the DOF is given by sensors. A sensor is a matrix of photoreceptors. Each pixel in this matrix has a specific dimension. If the circle is going to be small enough to still be concentrated on the same photoreceptro, then it will be on focus

- In theory, when imaging a scene through a thin lens, only the points at a certain distance can be on focus, all the others appear blurred into circles
  - However, as long as such circles are smaller than the size of the photosensing elements, the image will still look on-focus (i.e. the light is collected by a single pixel of the camera sensor)

if the circle of confusion is smaller than the pixel, I'll get an focus image



if the circle of confusion gets a lot bigger (the point is far away from the camera or too close to the camera, namely before/after the focusing distance) the image is going to be blurred

Diaphragm

- In theory, when imaging a scene through a thin lens, only the points at a certain distance can be on focus, all the others appear blurred into circles
  - However, as long as such circles are smaller than the size of the photosensing elements, the image will still look on-focus (i.e. the light is collected by a single pixel of the camera CCD)
  - The range of distances across which the image appears on focus - due to blur circles being small enough - determines the DOF (Depth of Field) of the imaging apparatus

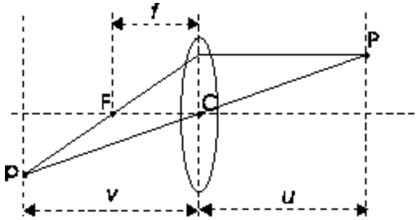
- Cameras often deploy an adjustable diaphragm (iris) to control the amount of light gathered through the *effective aperture* of the lens
  - Reduce aperture => less light => smaller blur circle
  - More aperture => more light => larger blur circle

trade-off

- We close the diaphragm => increase depth of field => not enough light => increase exposure time => moving object => motion blur => still scenes

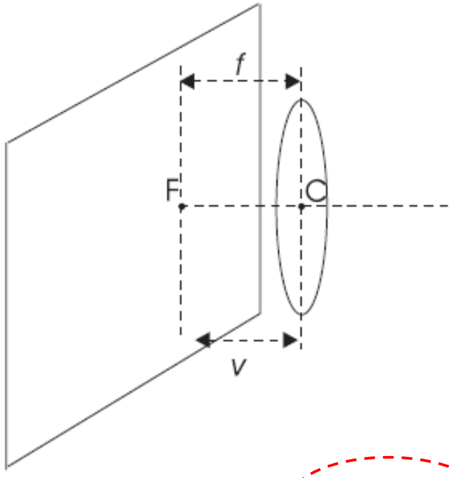
another trade-off

Focusing Mechanism

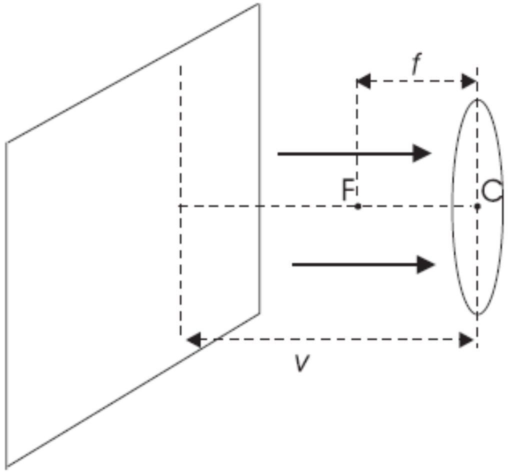


- To focus on objects at diverse distances:
  - mechanism that allows the lens (or lens subsystem) to translate along the optical axis with respect to the *fixed* position of the image plane

$$\cancel{\frac{1}{u}} + \frac{1}{v} = \frac{1}{f}$$



$v = f \longrightarrow u = \infty$

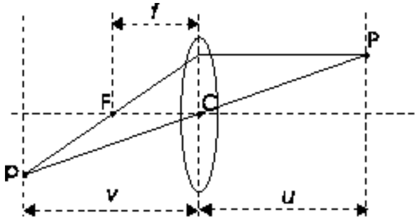


- At one end position ($v=f$) the camera is focused at infinity

all points at infinity are on focus

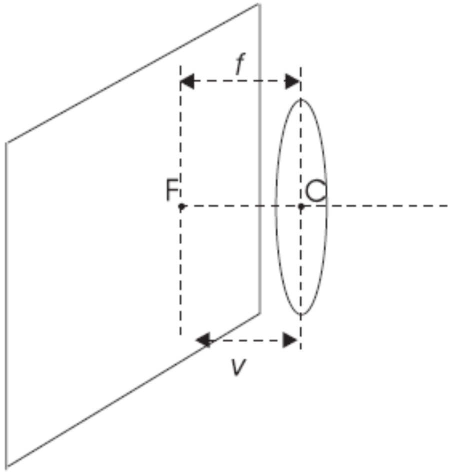
two end positions

Focusing Mechanism

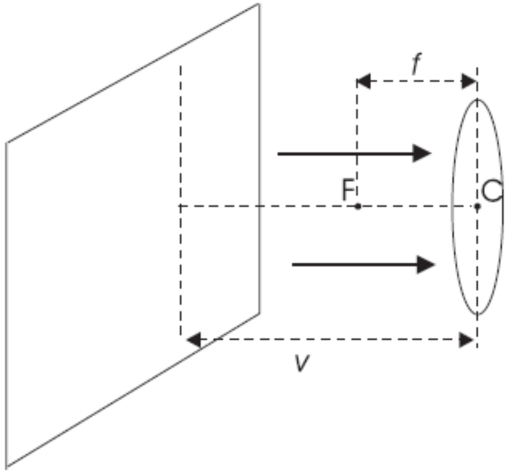


- To focus on objects at diverse distances:
  - mechanism that allows the lens (or lens subsystem) to translate along the optical axis with respect to the *fixed* position of the image plane

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$$



$v = f \longrightarrow u = \infty$



$v \nearrow \longrightarrow u \searrow$

as long as we increase "v" we decrease "u"

- At <u>one end position</u> (v=f) the camera is focused at infinity
- The mechanism allows the lens to be translated farther away from the image plane up to a certain maximum value (the <u>second end position</u>), which determines the minimum focusing distance

traslate the lens until the image can be on focus, until the boundary constituted by the foc. distance

Image Digitization

geometric optic part is done
now let's start another part

- Generally speaking, the image plane of a camera consists of a planar sensor which converts the irradiance at any point into an electric quantity (e.g. a voltage)
  - Transduction from light to an electric quantity => "records" the light coming from a scene point

image on computers are not continouos

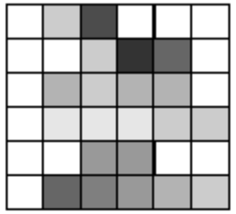
- Ho do we discretise it?

we sample due to the finite nature of sensors, which are matrices of photoreceptors: we have only a finite number of pixels, then it's a preliminar sample -> this is called spatial sampling

Continuous Image

Sampling

1° step



Sampled according to a two dimensional grid

we quantize because we convert the light to a value understandable by computers

Quantization

2° step

255	204	77	255	255	255
255	255	204	51	102	255
255	178	204	178	178	255
255	230	230	230	204	204
255	255	153	153	255	255
255	102	128	153	178	204

Sampled and Quantized Image

- Such a continuous "electric" image is sampled and quantized to end up with a digital image suitable to visualization and processing by a computer
  - The continuous voltages in the sampled image are gonna be quantized into a fixed number of levels

Image Digitization

--> we use 1 byte to represent the intensity values because it's a good trade-off between precision and storage efficiency

hypothize we are working with gray-scale images for now
we will have only a single value per pixel

is due to sensor, which is usually a 2D array

- **Sampling** – The planar continuous image is sampled along both the horizontal and vertical directions to pick up a 2D array (matrix) of N×M samples known as pixels:

this is the standard representation of images

$$I(x,y) \implies \begin{bmatrix} I(0,0) & I(0,1) & \dots & I(0,M-1) \\ \vdots & & & \\ \vdots & & & \\ I(N-1,0) & I(N-1,1) & \dots & I(N-1,M-1) \end{bmatrix}$$

(0, 0) on the top-left is a standard de-facto

- **Quantization** – The continuous range of values associated with pixels is quantized into $l = 2^m$ discrete levels known as gray-levels

  we have to choose the number of irradiance levels, l -> it depends on how many bits we have at disposal

  - m is the number of bits used to represent a pixel, with the memory occupancy (in bits) of a gray-scale image given by: $B = N \times M \times m$
    - m=8 in gray-scale digital images, so that, e.g., a VGA format (480×640) image requires 300 Kbytes for storage while a 1mpx image requires 1 Mbytes.

      rows columns

      VGA format is a sensor 480x640 pixels

      we use 1 byte to quantize voltages, then we have 2^8 = 256 possible intensity values
    - colour digital images are instead typically represented within computers using 3 bytes per pixels (one byte for each of the RGB channels). for each channel we will have a matrix of pixels and for each of them we will have 8 bits = 1 byte, thus 24 bits

- The more the better => more pixels + more bits per pixel => higher quality image

Digitization vs. Image Quality

- The more bits we spend for its representation, the higher the quality of the digital image (we get a closer approximation to the ideal continuous image)
- This applies to both sampling as well as quantization parameters

8-bit quantization

512×512, l=256
(original)

2^8 levels allowed
for the quantization

less pixels, less quality

64×64, l=256
(coarser sampling)

we only use 4 bits for quantization

we have
less shades
of gray to
represent
the whole
spectrum

less smooth

512×512, l=16
(coarser quantization)

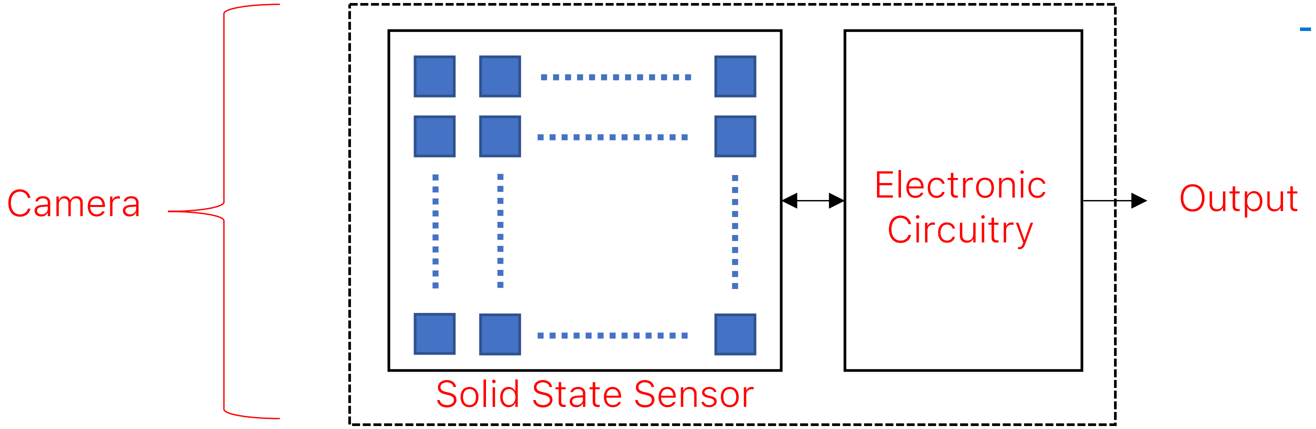
Camera sensors

I need electronic circuits to convert and process these values acquired from the photoreceptors, and also converting them from analog to digital values

each of them will register a different irradiance value

- The sensor is a 2D array of photodetectors (photogates or photodiodes)
  - During exposure time, each detector converts the incident light into a proportional electric charge (i.e. photons to electrons)
  - The companion circuitry reads-out the charge to generate the output signal, which can be either *digital* or analog
  - For digital: the camera includes also the necessary ADC circuitry

electronic circuits have two tasks:
- processing
- converting



- Nowadays, there is never a continuous image in practice => the image is sensed directly as a sampled signal

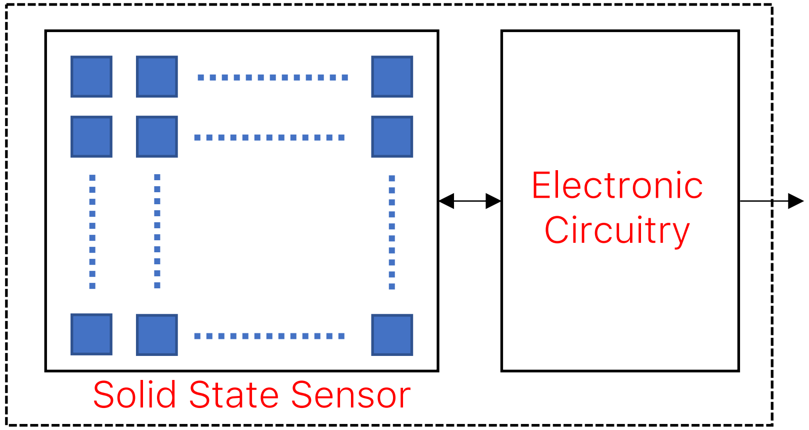
Camera sensors

CMOS used for webcams, phones, ...

the main difference:
- CCD has a unique electronic circuit which will do all the job of proc. & conversion
- CMOS has a circuit of that type for each photoreceptor, then each pixel has its own elect. circuit for processing and conversion

this allows to work only on a part of the image, in order to speed-up the processing

- Today, the two main sensor technologies are:
  - CCD (Charge Coupled Devices);
  - CMOS (Complementary Metal Oxide Semiconductor)
- Unlike CCD, CMOS technology allows the electronic circuitry to be integrated within the same chip as the sensor ("one chip camera")
  - This provides more compactness, less power consumption and often lower system cost.
- Unlike CCD, CMOS sensors allow an arbitrary window to be read-out without having to receive the full image
  - This can be useful to inspect or track at a higher speed a small Region Of Interest (ROI) within the image
- CCD technology typically provides higher Signal-to-Noise Ratio (SNR) , higher Dynamic Range (DR) and better uniformity

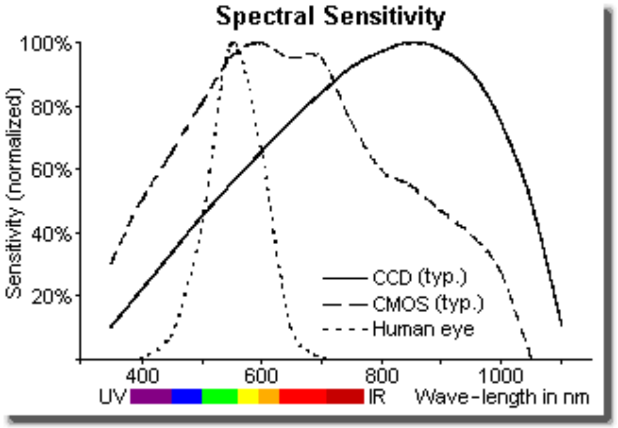


Colour Sensors

these two are very different for which regard the processing of light they receive

infact, CCD and CMOS have a different sensitivity towards the SPECTRUM of light

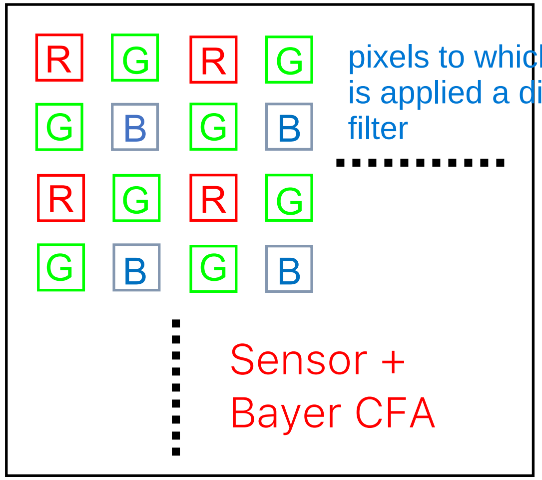
CCD and CMOS are built to acquire elements that we (humans) cannot see (like infrareds)



- CCD/CMOS sensors are sensitive to light ranging from near-ultraviolet (200 nm) through the visible spectrum (380-780 nm) up to the near infrared (1100 nm). The sensed intensity at a pixel results from integration over the range of wavelengths of the spectral distribution of the incoming light multiplied by the spectral response function of the sensor ==> CCD/CMOS sensor cannot sense colour.

for sure, none of them can acquire colors, because every photoreceptor (whichever type) will register one single electrical value

- To create a colour sensor, an array of optical filters (Colour Filter Array) is placed in front of the photodetectors, so as to render each pixel sensitive to a specific range of wavelengths. only way to register colors: applying filters
  - In the most common, Bayer CFA, green filters are twice as much as red and blue ones to mimic the higher sensitivity of the human eye in the green range.
  - To obtain an RGB triplet at each pixel, missing samples are interpolated from neighbouring pixels (demosaicking). However, the true resolution of the sensor is smaller due to the green channel being subsampled by a factor of 2, the blue and red ones by 4.
    To obtain a full-color RGB triplet at each pixel, missing color information needs to be interpolated. Demosaicking is the process of estimating the missing color values by looking at the values of neighboring pixels.



pixels to which is applied a different filter

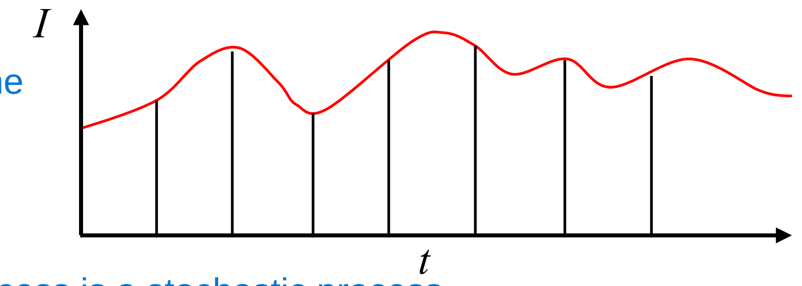
one drawback: subsampling lead us to loosing resolution, because we are using the same amount of pixel to register colors

- A "full resolution" – though more expensive - colour sensor can be achieved by deploying an optical prism to split the incoming light beam into 3 RGB beams sent to 3 distinct sensors equipped with corresponding filters.

Signal to Noise Ratio

there are some parameters that tells us how good our sensors are

we will never observe the same intensity value at different time-steps, for different reasons



the acquisition process is a stochastic process

- *Signal-to-Noise Ratio (SNR)* – The intensity measured at a pixel under perfectly static conditions varies due to the presence of random noise (i.e. a pixel value is not deterministic but rather a random variable).

how our sensors is good acquiring light even though the presence of noise

- The main noise sources are:
  - *Photon Shot Noise* – The time between photon arrivals at a pixel is governed by a Poisson statistic and thus the number of photons collected during exposure time is not constant.
  - *Electronic Circuitry Noise* – It is generated by the electronics which reads-out the charge and amplifies the resulting voltage signal.
  - *Quantization Noise* – related to the final ADC conversion (in digital cameras).
  - *Dark Current Noise* – a random amount of charge due to thermal excitement is observed at each pixel even though the sensor is not exposed to light.

- The SNR can be thought of as quantifying the strength of the "true" signal with respect to the unwanted fluctuations induced by noise (i.e. the higher the better). It should be measured according to standard procedures and it is usually expressed either in *decibels* or *bits*:

$$\mathrm{SNR}_{dB} = 20 \cdot \log_{10}(\mathrm{SNR}); \quad \mathrm{SNR}_{bit} = \log_2(\mathrm{SNR})$$

Dynamic Range

when you acquire an image, this goes through a quantization step;
↳ this tells you how many levels of intensities you can represent in your scene
↳ but if I have dark regions (due to shadows), and very bright ones (due to sunlight) you're gonna to be able to observe details only in the dark or the bright regions, coz I dont have enough levels of intensies (up to 256)

Dynamic Range allows us to cope with this difference of intensity, giving us the tools to represent details in both the regions

- Dynamic Range (DR) – If the sensed amount of light is too small, the "true" signal cannot be distinguished from noise
  - given E_{min} - the minimum detectable irradiation (which depends on what?) it's a function of noise
  - given E_{max} - the saturation irradiation (i.e. the amount of light that would fill up the capacity of a photodetector).
  - The DR of a sensor is defined as $DR = \frac{E_{max}}{E_{min}}$; and, like the SNR, it is often specified in decibels or bits.
- As it is the case of the SNR, also for the DR the higher is the better. Indeed, the higher the DR the better is the ability of the sensor to simultaneously capture in one image both the dark and bright structures of the scene.
- An active research field in image processing deals with creating High Dynamic Range (HDR) images by combining together a sequence of images of the same subject taken under different exposure times (see e.g. http://www.hdrsoft.com/index.html). combine 3 images to see details of both dark and bright regions