University of Bologna

Image Formation and Acquisition

Luigi Di Stefano (luigi.distefano@unibo.it)

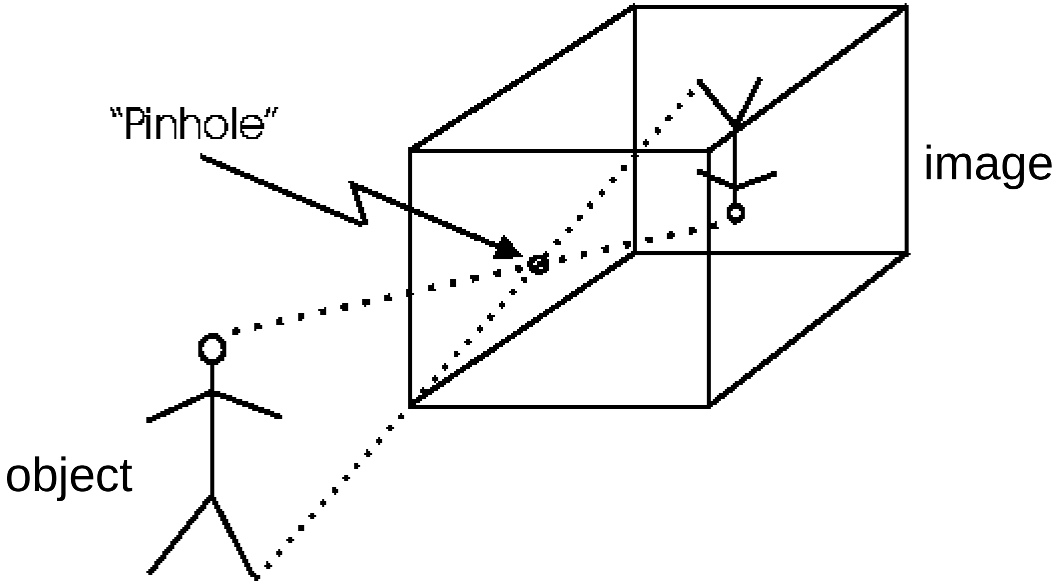
Introduction

- **An imaging device gathers the light reflected by 3D objects to create a 2D representation of the scene (i.e. the image). 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. It is therefore worth to understand the image formation and acquisition process. This requires studying:**
  - **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

The "pinhole camera" is the simplest imaging device: light goes through the very small pinhole and hits the image plane.

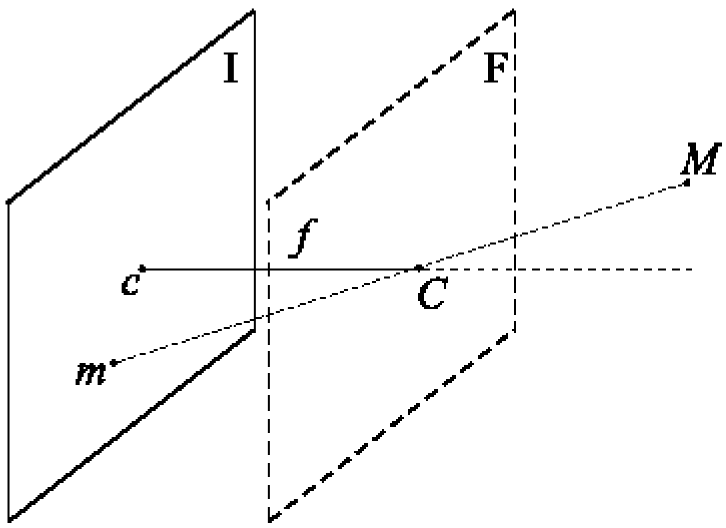
Geometrically, the image is achieved by drawing straight rays from scene points through the hole up to the image plane.



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

Perspective Projection (1)

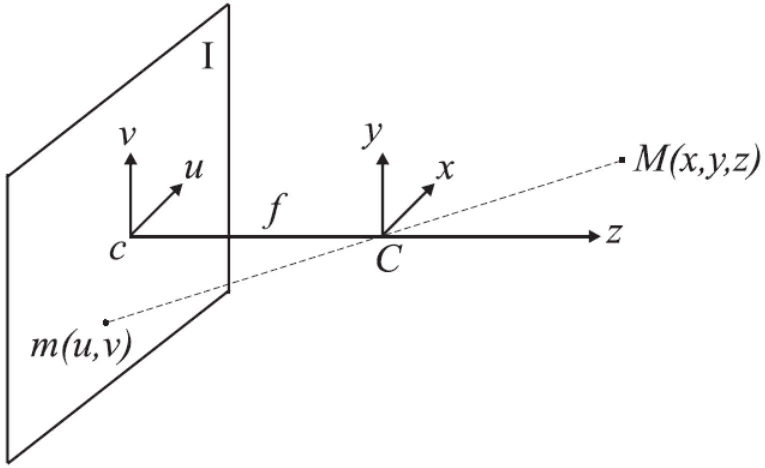
The geometric model of image formation in a pinhole camera is known as Perspective Projection.



M : scene point
m : corresponding image point
I : image plane
C : optical centre
Line through *C* and orthogonal to *I* : optical axis
c : intersection between optical axis and image plane (image centre or piercing point)
f : focal length
F : focal plane

Perspective Projection (2)

Considering the reference frames shown in the figure, the equations to map scene points into their corresponding image points are as follows:



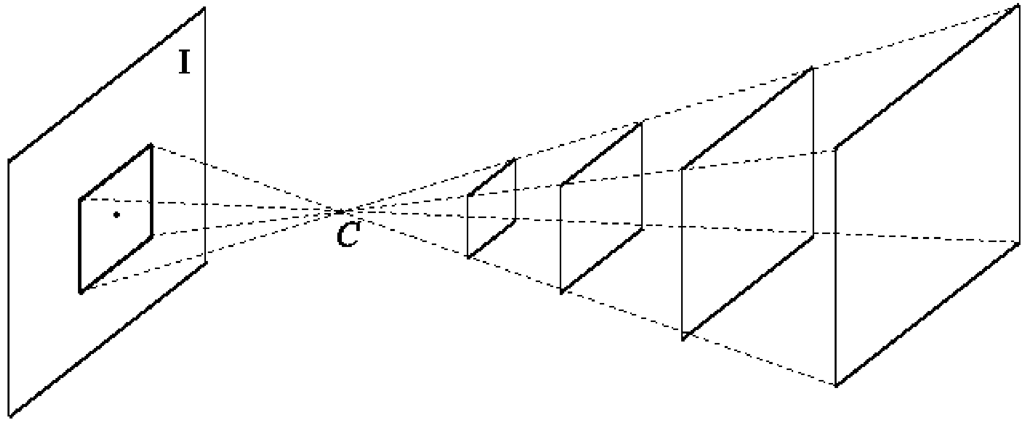
$$\frac{u}{x} = \frac{v}{y} = -\frac{f}{z} \rightarrow u = -x\frac{f}{z}; v = -y\frac{f}{z}$$

To get rid of the up-down and left right inversions, the image plane can be thought of as lying in front rather than behind the optical centre

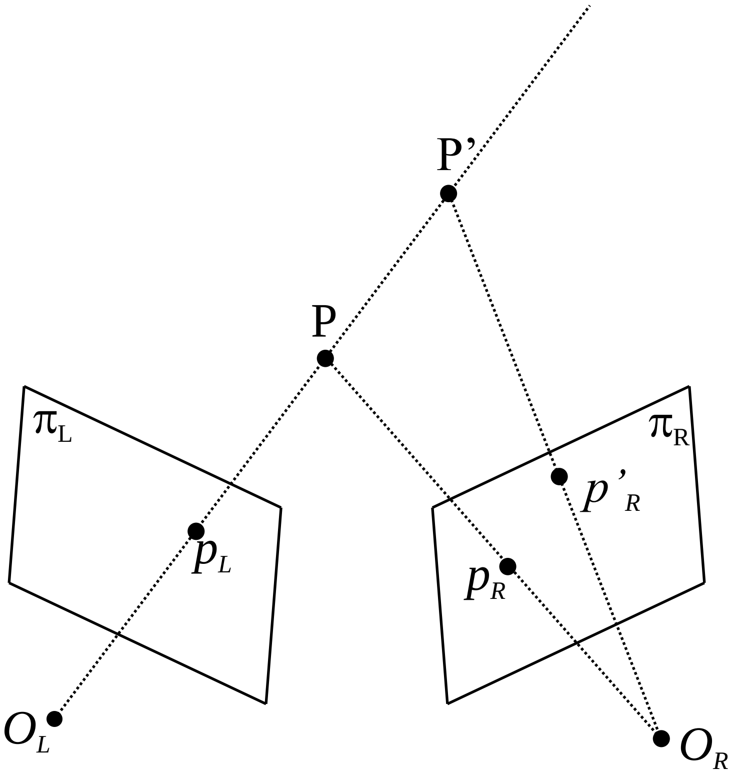
$$u = x\frac{f}{z}; \quad v = y\frac{f}{z}$$

Perspective Projection (3)

- **The image formation process deals with mapping a 3D space onto a 2D space, thus leading inevitably to loss of information.**
- **Indeed, the mapping is not a bijection: a given scene point is mapped into a unique image point, but a given image point is mapped onto a 3D line (i.e. the line through the point, *m*, and the optical centre, *C*).**
- **Thus, recovering the 3D structure of a scene from a single image is an ill-posed problem (the solution is not unique), as once we take 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).**



Stereo images allow to infer 3D



Given correspondences, 3D information can be recovered easily by triangulation

Standard stereo geometry

- **Parallel (x,y,z) axes**
- **Same focal lenght**
- **Coplanar image planes**

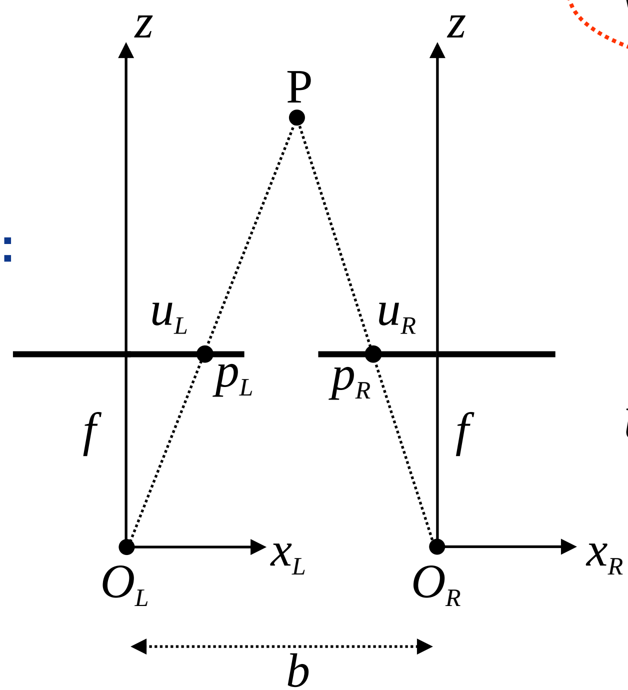
The transformation between the two reference frames is just a translation (*b*), usually horizontal:

$$\mathbf{P}_L - \mathbf{P}_R = \begin{bmatrix} \boldsymbol{b} \\ 0 \\ 0 \end{bmatrix}$$

$$x_L - x_R = b$$

$$y_L = y_R = y$$

$$z_L = z_R = z$$



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

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

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

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

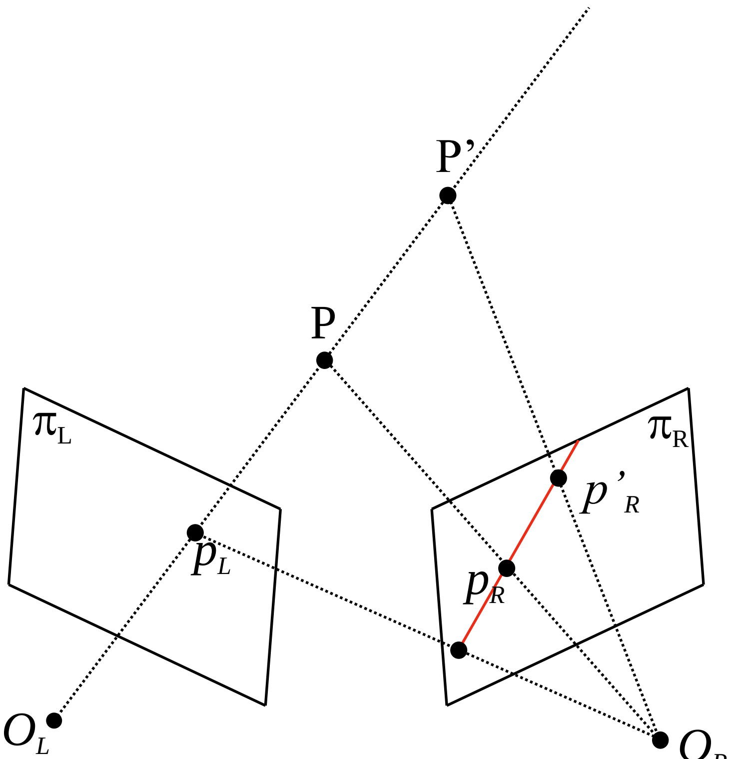
$$u_L - u_R = d$$

(disparity)

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

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

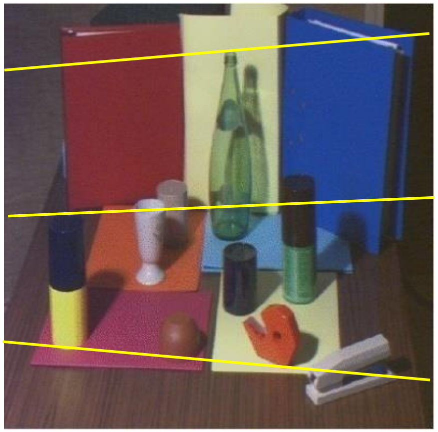
Epipolar Geometry



Epipolar line
(associated with p_L in Π_R)

The search space of the stereo correspondence problem is always 1D !

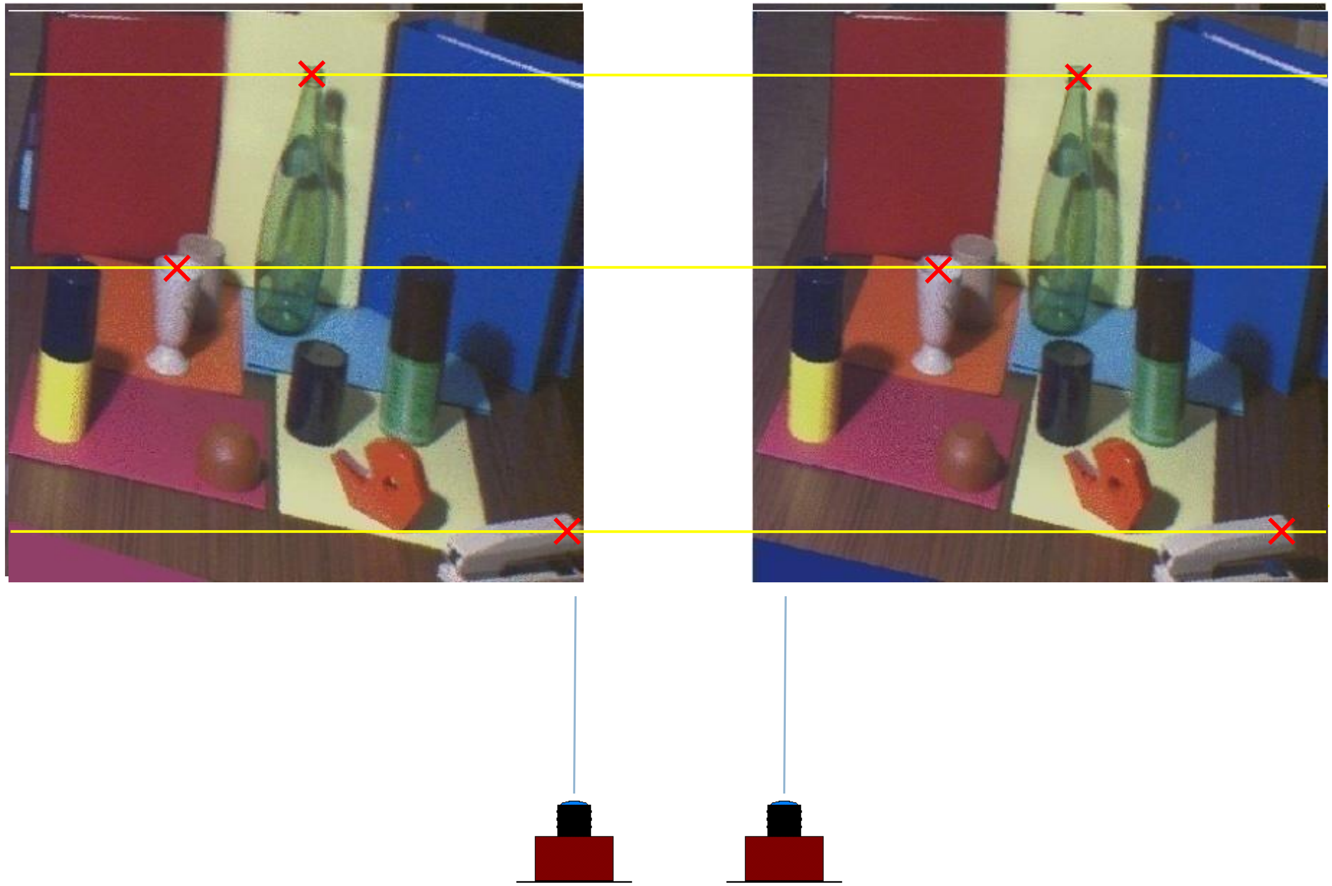
All the epipolar lines in an image meet at a point called *epipole* (i.e. the projection of the optical center of the other image)



However, searching through oblique epipolar lines is awkward !

Rectification

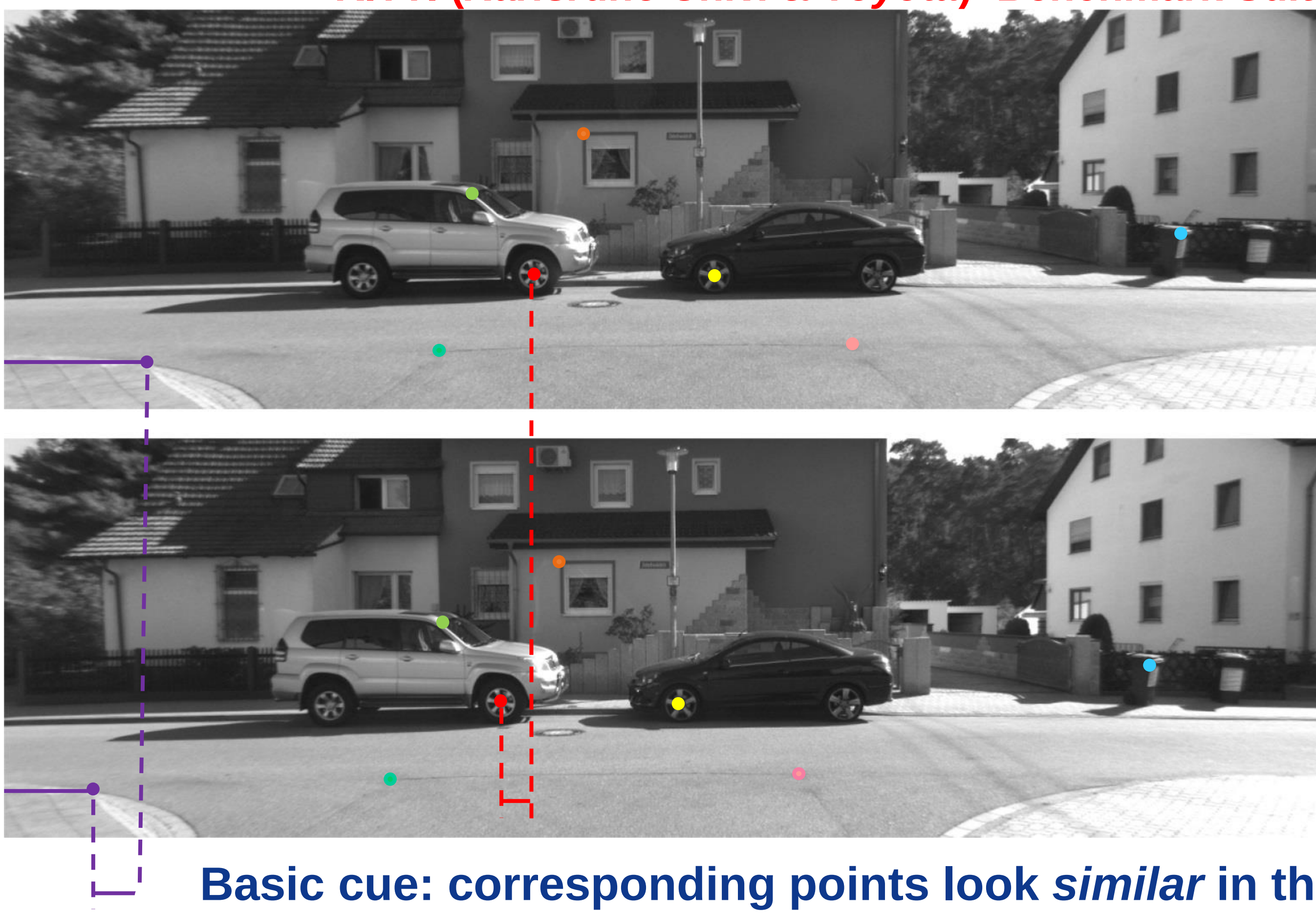
We can always warp the images as if they were acquired through a standard geometry (horizontal and collinear conjugate epipolar lines) by computing and applying to both a transformation (i.e. homography) known as *rectification*.



The Stereo Correspondence Problem

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*.

KITTI (*Karlsruhe Univ. & Toyota*) Benchmark Suite



Basic cue: corresponding points look *similar* in the two images

Some properties of Perspective Projection

- **The farther objects are from the camera, the smaller they appear in the image. As a matter of example, 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:**

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

The relationship turns out 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.

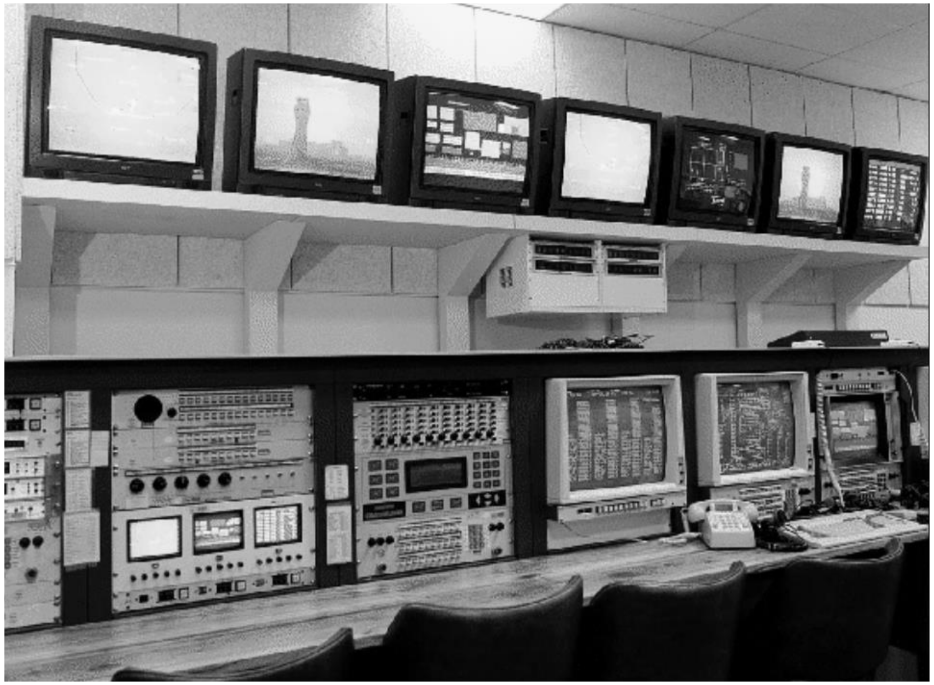
- **Perspective projection maps 3D lines into image lines.**

- **Ratios of lengths are not preserved (unless the scene is planar and parallel to the image plane).**

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

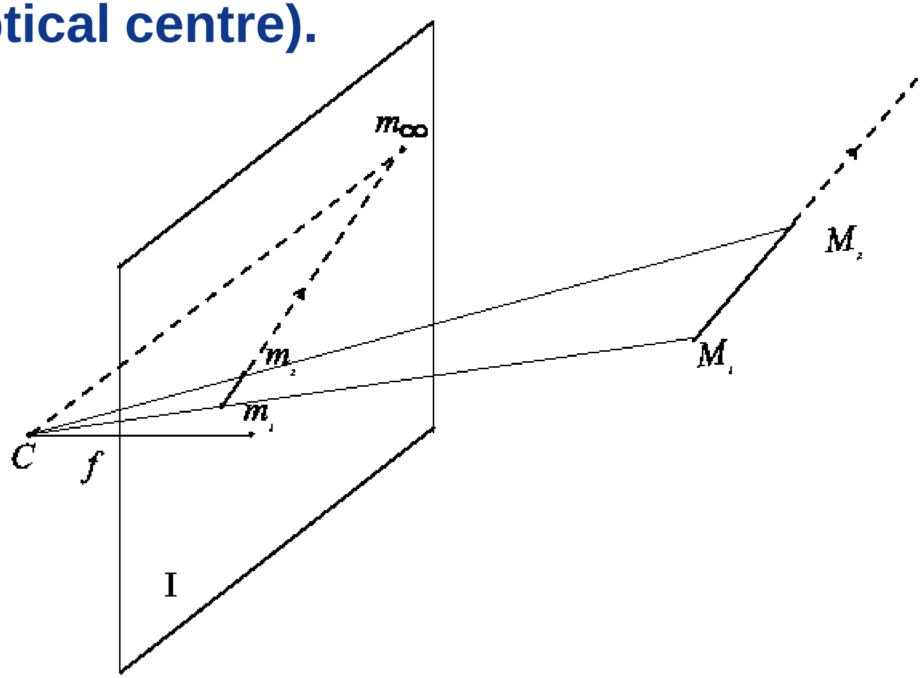
Vanishing Points (1)

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



Vanishing Points (2)

The vanishing point of a 3D line is the *image* of the *point at infinity* of the line (i.e. the image of the point on the line which is infinitely distant from the optical centre).



As such, it can be determined by the intersection between the image plane and the line parallel to the given one and passing through the optical centre.

Accordingly, all parallel 3D lines will share the same vanishing point, i.e. they "meet" at their vanishing point in the image, but in the "special case" of such a point being at infinity (i.e. the 3D lines are parallel to the image plane).

Vanishing Points (3)

Let's now find the vanishing point of the line:

$$M = M_0 + \lambda D = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + \lambda \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

where M_0 is a point on the line, D the direction cosines vector.

We first project a generic point of the line:

$$m = \begin{bmatrix} u \\ v \end{bmatrix}, \qquad u = f\frac{x_0 + \lambda a}{z_0 + \lambda c}, \quad v = f\frac{y_0 + \lambda b}{z_0 + \lambda c}$$

Then, to get the vanishing point we consider the infinitely distant point along the line:

$$m_\infty = \begin{bmatrix} u_\infty \\ v_\infty \end{bmatrix}, \qquad u_\infty = \lim_{\lambda \to \infty} u = f\frac{a}{c}, \quad v_\infty = \lim_{\lambda \to \infty} v = f\frac{b}{c}$$

As expected, the vanishing point depends on the orientation of the line only, not on its position, and whenever the line is parallel to the image plane (*c=0*) it goes to infinity. It can also been shown easily that in such a case the image of the line has the same orientation as the 3D line.

Orientation of parallel lines from vanishing points

Knowledge of the vanishing point of a sheaf of parallel lines (and of the focal length) allows for determining the unknown orientation of the lines.

$$\begin{cases} u_\infty = f\frac{a}{c} \\ v_\infty = f\frac{b}{c} \end{cases}$$

$$c^2\left(u_\infty^2 + v_\infty^2\right) = f^2\left(1 - c^2\right)$$

$$c = \frac{f}{\sqrt{u_\infty^2 + v_\infty^2 + f^2}}$$

$$a = \frac{u_\infty}{\sqrt{u_\infty^2+v_\infty^2+f^2}}$$

$$b = \frac{v_\infty}{\sqrt{u_\infty^2+v_\infty^2+f^2}}$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \frac{1}{\sqrt{u_\infty^2 + v_\infty^2 + f^2}} \begin{bmatrix} u_\infty \\ v_\infty \\ f \end{bmatrix}$$

Steering a robot by vanishing points

A mobile robot can be driven through indoor hallways by tracking the dominant vanishing point and steering the robot so as to keep it at the centre of the image.

http://www.youtube.com/watch?v=nb0VpSYtJ_Y

See also : http://www.roborealm.com/help/Vanishing_Point.php

Camera orientation from vanishing points

Computer Vision Laboratory

Knowledge of the vanishing points of two orthogonal directions allows for determining camera orientation wrt to a scene plane.

From the vanishing point of the horizontal lines on the façade we get the unit vector, $\mathbf{i}_c$, which defines the orientation of such lines in the camera reference frame. Likewise, from the vanishing point of the vertical lines on the façade we get $\mathbf{j}_c$, the vector product between the two providing $\mathbf{k}_c$:

$$\mathbf{k}_c = \mathbf{i}_c \times \mathbf{j}_c$$

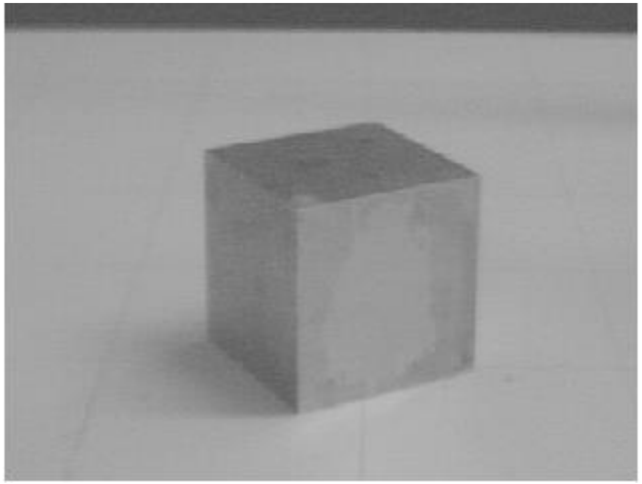
The three found unit vectors define the orientation of the camera wrt the considered scene plane, i.e. through the rotation matrix from the reference frame attached to the plane to the camera reference frame (and viceversa, by matrix transposition)

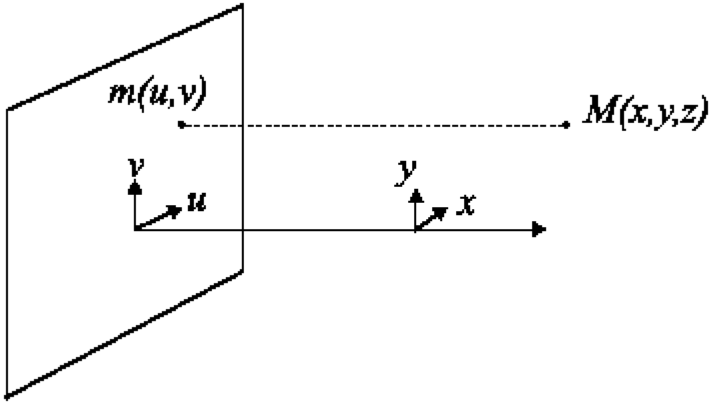
$$\mathbf{R}_{pc} = \begin{bmatrix} \mathbf{i}_c & \mathbf{j}_c & \mathbf{k}_c \end{bmatrix}$$

$$\mathbf{R}_{cp} = \mathbf{R}_{pc}^T$$

Weak Perspective (1)

Perspective effects may be not so evident, this occurring whenever the framed subject is thin compared to the distance from the camera.





In such cases, perspective projection can be approximated by a *scaled* orthographic projection (weak perspective):

$$\begin{cases} u = x \\ v = y \end{cases} \rightarrow \begin{cases} u = sx \\ v = sy \end{cases}$$

ortographic projection

scaling

Weak Perspective (2)

Let's denote as $\left[z_0 - \Delta z, z_0 + \Delta z\right]$ the range of distances of the framed subject. If Δz is small compared to z_0 we obtain:

$$\frac{f}{z_0 + \Delta z} \approx \frac{f}{z_0 - \Delta z} \approx \frac{f}{z_0}$$

$$u \approx \frac{f}{z_0} x, \qquad v \approx \frac{f}{z_0} y$$

Which means that the perspective scaling factor $\left(f/z\right)$ is approximately constant $\left(f/z_0\right)$ for all points of the framed subject.

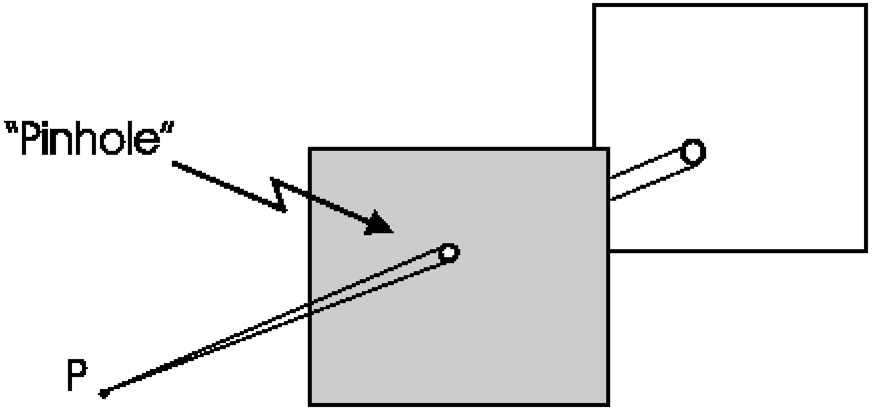
The above equations define a *scaled* orthographic projection :

$$u = sx, \qquad v = sy$$

ortographic projection + scaling

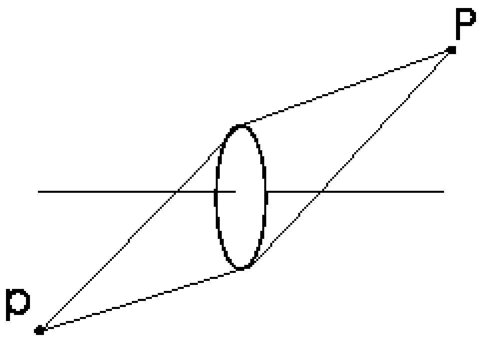
Using Lenses

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).



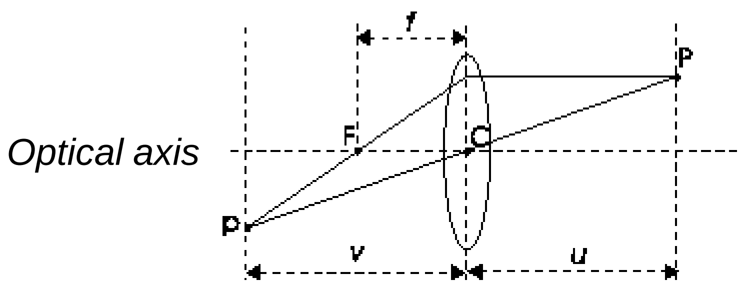
The drawback is that such a small aperture allows gathering a very limited amount of light. Thus, getting sufficiently bright images mandates very long exposure times. As a result, only static scenes can be acquired by a pinhole device to avoid motion blur.

Therefore, cameras rely on lenses to gather more light from a scene point and focus it on a single image point. This enables much smaller exposure times, as required e.g. to avoid motion blur in dynamic scenes. However, the DOF is no longer infinite, for only points across a limited range of distances can be simultaneously on focus in a given image.



Thin Lens Equation

- **Cameras often feature complex optical systems, comprising multiple lenses. Yet, we will consider here the approximate model known as *thin lens equation*:**



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

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

- **To graphically determine the position of a focused image point we can leverage on the following two properties of thin lenses:**
  1. **Rays parallel to the optical axis are deflected to pass through *F*.**
  2. **Rays through *C* are undeflected.**
- **It is worth pointing out that, if the image is on focus, the image formation process obeys to the perspective projection model, with the centre of the lens being the optical centre and the distance *v* acting as the *effective* focal length of the projection (a different concept than the focal length of the lens !)**

Circles of Confusion (1)

- **Due to the thin lens equation, choosing the distance of the image plane determines the distance at which scene points appear on focus in the image:**

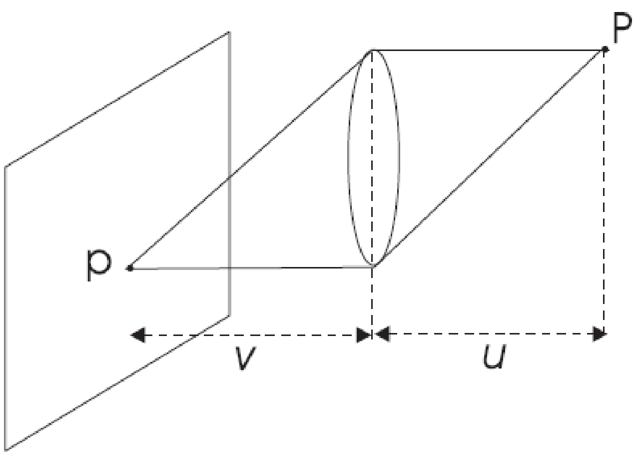
$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \rightarrow u = \frac{vf}{v-f}$$

- **Likewise, to acquire scene points at a certain distance we must set the position of the image plane accordingly:**

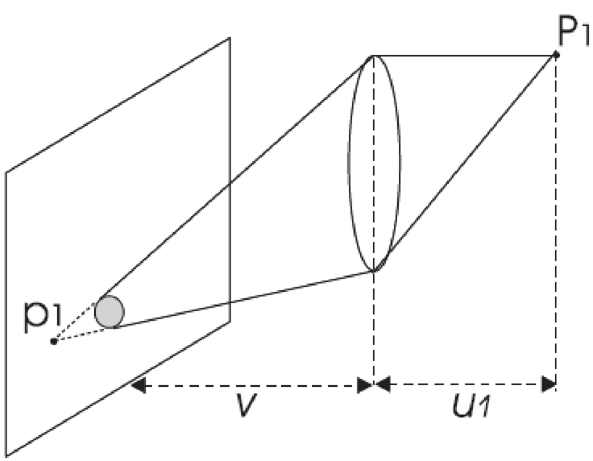
$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \rightarrow v = \frac{uf}{u-f}$$

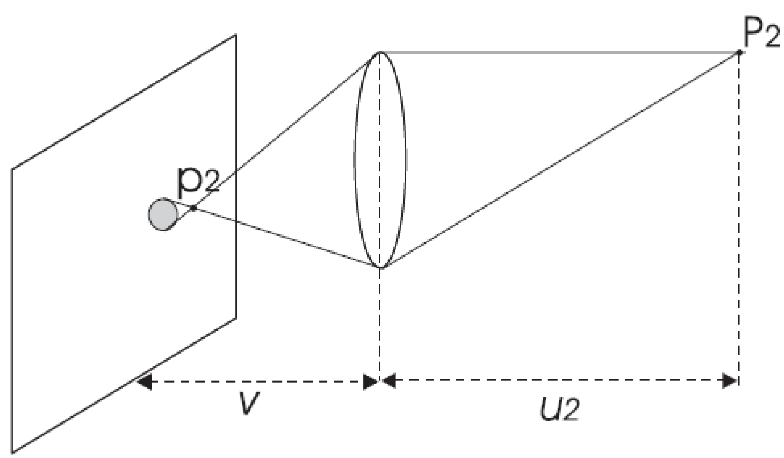
- **Given the chosen position of the image plane, scene points both in front and behind the focusing plane will result out-of-focus, thereby appearing in the image as circles, known as Circles of Confusion or Blur Circles, rather than points.**

Circles of Confusion (2)



P belongs to the focusing scene plane
P_1 lies closer to the lens than P ($u_1<u$)
P_2 is farther away to the lens than P ($u_2>u$)





Diaphragm, DOF and F-number

- **In theory, when imaging a scene through a thin lens, only the points at a certain distance can be on focus, all the others appearing 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. 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.**

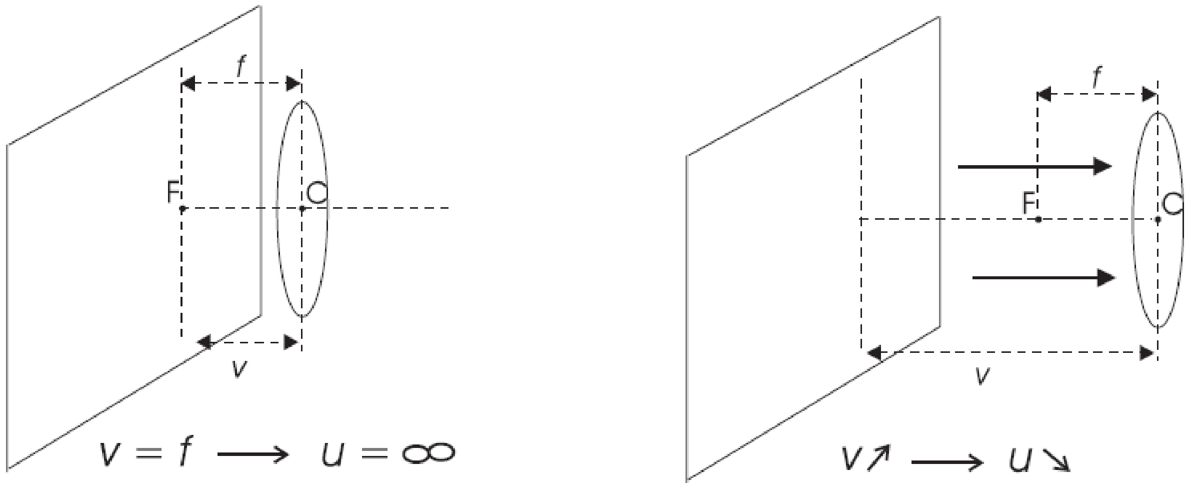
- **Cameras often deploy an adjustable diaphragm (iris) to control the amount of light gathered through the *effective aperture* of the lens. The smaller the diaphragm aperture is, the larger turns out the DOF as a result of the smaller size of blur circles.**

- **The *F-number* is the ratio of the focal length to the effective aperture of the lens $\left(f/d\right)$. F-number discrete units (also known as *stops*) are usually reported on the diaphragm (e.g. 1.4, 2, 2.8, 4, 5.6, 8, 11, 16…) to allow the user to adjust the effective aperture. The higher the chosen *stop*, the smaller is the diaphragm aperture and thus the larger is the actual DOF.**

Focusing Mechanism

- **To focus on objects at diverse distances, another mechanism allow the lens (or lens subsystem) to translate along the optical axis with respect to the –fixed- position of the image plane (in nowadays cameras, a solid state sensor mounted on a PCB).**

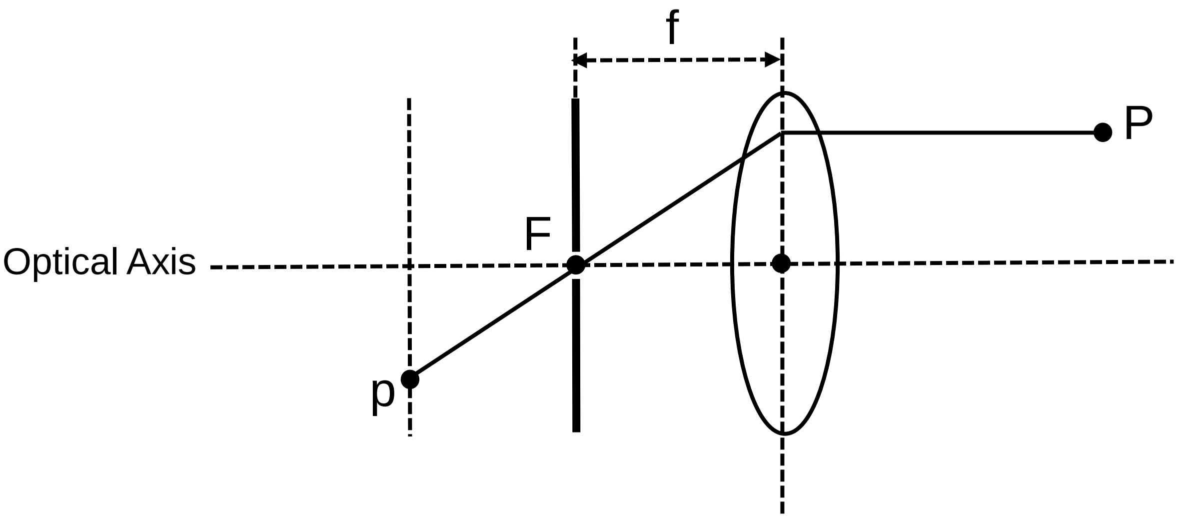
$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$$



- **At one end position (*v=f*) the camera is focused at infinity, then the mechanism allow the lens to be translated farther away from the image plane up to a certain maximum value (the second end position), which determines the minimum focusing distance.**

Telecentric Lens

- **By placing a diaphragm with a small hole at the focal point of the lens we can block all light rays but those parallel to the optical axis, thereby realizing an ortographic projection. This kind of optical device is referred to as *telecentric lens*. The size of objects to be imaged with telecentric lenses cannot be larger than the lens itself.**



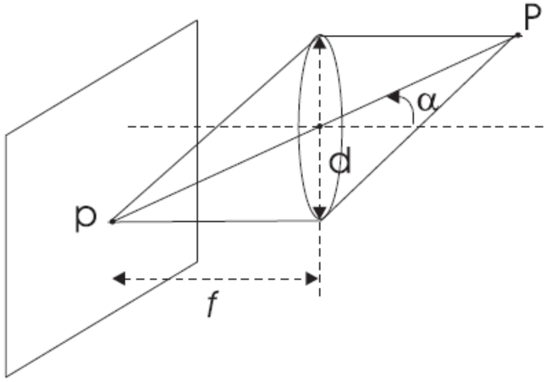
- **Although bulky and expensive, telecentric lenses do not exhibit perspective distortion and thus are often used in machine vision to analyse with high accuracy 3D objects that might appear at different distances or off-axis.**

Fundamental Radiometric Relation

- **The *Irradiance, E,* of a point on a surface is the amount of light incident on that point (formally measured as a power per unit area).**
- **The *Radiance*, *L*, of a surface point in any direction is the amount of light emitted by the point in that direction (formally measured as power per unit area in the direction of radiation per unit solid angle).**

- **The fundamental radiometric relation shows that the *Irradiance* of an image point is proportional to the *Radiance* of the corresponding scene point along the viewing direction:**

$$E(p) = L(P) \cdot \frac{\pi}{4} \cdot \left(\frac{d}{f}\right)^2 \cdot \cos^4 \alpha$$



Thus, an imager senses the light emitted by scene points, the brightness of image points being proportional to the light emitted along the direction of projection by their corresponding scene points.

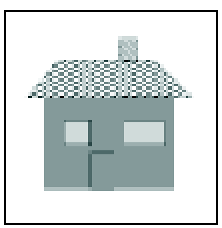
Reflectance of a Surface

- The *Radiance* of a surface point along the viewing direction depends on the power and position of the light sources as well as on the reflectance properties of the material.
- Such reflectance properties are usually described by a complex function called Bi-Directional Reflectance Function (BDRF). Accordingly, it is possible to determine the amount of light emitted in a certain direction given the amount of light received from the sources.
- Two simplified and opposite reflectance models are given by the Lambertian (diffusely reflecting) and specular surface. The former appears equally bright from any viewing direction (i.e. the incoming light is spread equally across all viewing directions): $L(P) = \rho(P) \cdot E(P), with\ \rho(P): albedo\ (visible\ light)$
- Conversely, a specular surface reflects the light coming from any direction in only one direction, so as the incident and reflected rays are coplanar and form the same angle wrt the normal at the surface point.
- Real surfaces typically show a mixed behavior between diffusion and specular reflection, and as such are usually described by more sophisticated reflectance models.

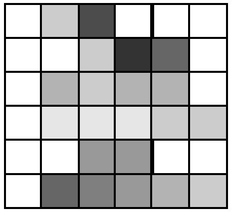
Image Digitization

- **First, we take an abstract and general view of the image digitization process. Then, we will discuss some practical aspects related to image digitization in modern cameras.**

- **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).**
- **Afterwards, such a continuous "electric" image is sampled and quantized to end up with a digital image suitable to visualization and processing by a computer.**



Sampling



Quantization

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

Continuous Image

Sampled Image

Sampled and Quantized Image

Sampling and Quantization

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

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

- ***Quantization*** **– The continuous range of values associated with pixels is quantized into** $l = 2^m$ **discrete levels known as *gray-levels*. Thus, *m* is the number of bits used to represent a pixel, with the memory occupancy (in bits) of a gray-scale image given by:**

$$B = N \times M \times m$$

- **Usually, *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. Colour digital images are instead typically represented within computers using 3 bytes per pixels (one byte for each of the RGB channels).**

Digitization vs. Image Quality (1)

- **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. We show here a few examples of the impact on quality of digitization parameters using the Lenna image (**http://www.cs.cmu.edu/~chuck/lennapg/lenna.shtml**)**

512×512, l=256 (original)

64×64, l=256 (coarser sampling)

Loss of details and jagged contours due to aliasing !

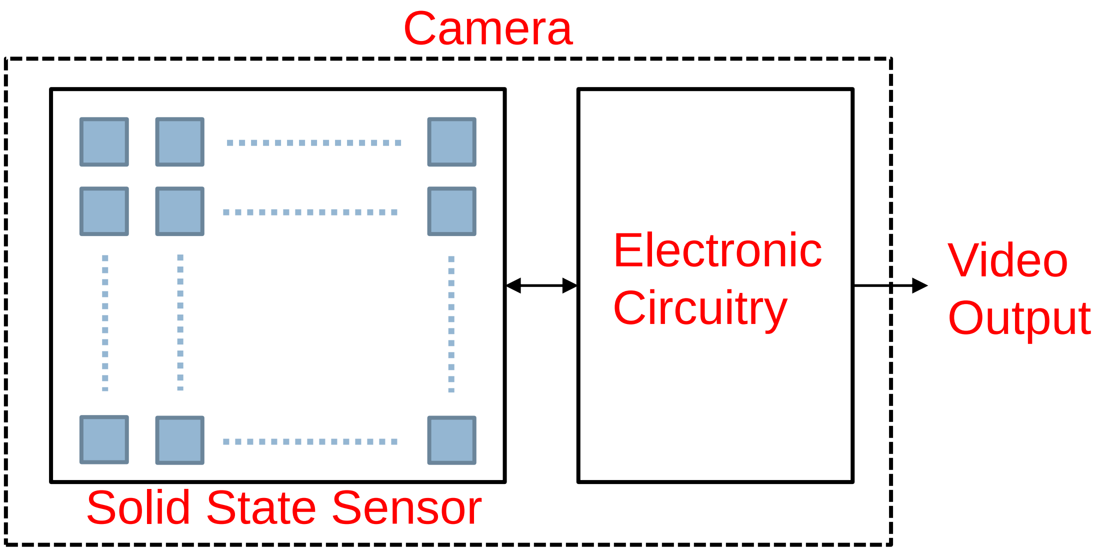
Digitization vs. Image Quality (2)

512×512, l=256
(original)

512×512, l=16
(coarser quantization)

Loss of smooth gray-scale transitions
(false contouring or posterization)

Digitization in Practice



Hence, there is never a continuous image in practice, for the image is sensed directly as a sampled signal. In analog cameras though, the native sampling taking place at the sensor is lost in the generation of the analog output, which is then sampled and quantized by a dedicated circuitry within the computer known as *analog* frame grabber. As a result, the pixels in a digital image coming from an analog camera do not correspond to those sensed by the photodetectors.

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). Then, the companion circuitry reads-out the charge to generate the output signal, which can be either ***digital*** or analog. In the former case, the camera includes also the necessary ADC circuitry, while nowadays the latter type of cameras are mostly manufactured for the sake of legacy systems.

Today, the two main sensor technologies are CCD (Charge Coupled Devices) and CMOS (Complementary Metal Oxide Semiconductor)

Camera Parameters (1)

We provide here an intuitive explanation of some of the main camera parameters. Formal definitions and measurement procedures can be found in the EMVA Standard 1288.

- **<u>Signal-to-Noise Ratio (SNR)</u> – 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). The main noise sources are as follows.**
  - ***Photon Shot Noise* – The time between photon arrivals at a pixel is governed by a Poisson statistics 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.**

Camera Parameters (2)

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*:

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

- **<u>Dynamic Range (DR)</u> – If the sensed amount of light is too small, the "true" signal cannot be distinguished from noise: let's call E_{min} the *minimum detectable irradiation*. On the other hand, the charge stored at each pixel cannot exceed a certain quantity: let's call 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**

$$\mathrm{DR} = \frac{\mathrm{E}_{\max}}{\mathrm{E}_{\min}};$$

and, like the SNR, often specified in *decibels* or *bits*.

Camera Parameters (3)

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 matter taken under different exposure times (see e.g. http://www.hdrsoft.com/index.html).

- **<u>Sensitivity (Responsivity)</u> – deals with the amount of signal that the sensor can deliver per unit of input optical energy.**

- **<u>Uniformity (spatial or pattern noise)</u> – due to manufacturing tolerances both the response to light and the amount of dark noise vary across pixels.**

CCD vs CMOS

- **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 SNR, higher DR and better uniformity.**

Typical application scenarios

CMOS:
- **Cellular Phones**
- **Handheld Devices**
- **Webcams**
- **Video-surveillance**
- **Barcode Scanners**
- **……**

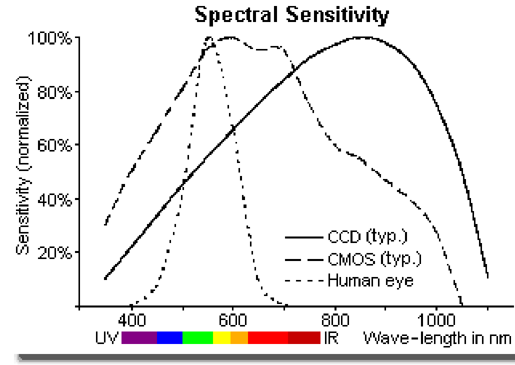
CCD:
- **Digital Photography**
- **TV Broadcasting**
- **High Performance Industrial Imaging**
- **Scientific and Medical Imaging**
- **…….**

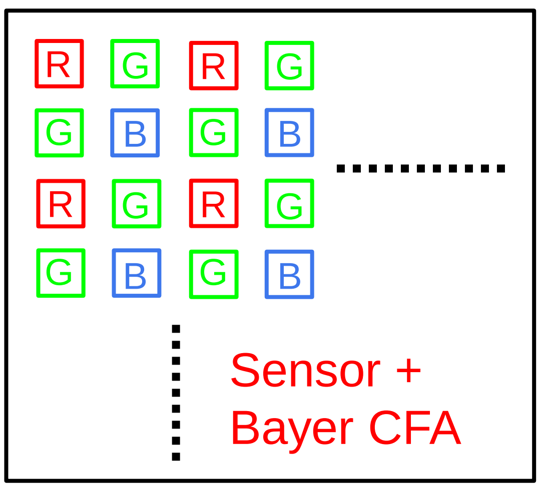
High volumes, space/power constraints

High-end imaging applications

Colour Sensors

- **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. As such CCD/CMOS sensor cannot sense colour.**



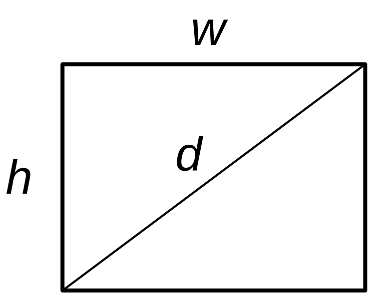


- **To create a colour sensor, an array of optical filters (Colour Filter Array) is placed in front of the photodetectors, so as to render each pixel sensitive to a specific range of wavelengths. In the most common Bayer CFA green filters are twice as much as red and blue ones to mimic the higher sensitivity 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.**

- **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.**

Sensor Sizes

- **CCD/CMOS sensors come in different sizes, which are specified in inches for the sake of legacy wrt old cameras based on cathode ray tubes. In such old cameras the size (in inch) was the outer diameter of the tube, the effective image plane size being roughly 2/3 of the diameter. Nowadays, the size of the diagonal of a solid state sensor is roughly 2/3 of its size. The table below reports some typical sensor sizes, together with the corresponding (square) pixel sizes for a VGA format (*w* × *h* = 640 × 480) sensor. In case of a higher resolution format, pixel sizes shrink proportionally (e.g. by a factor of 2 for a 1280 × 960 sensor).**



$$\frac{w}{h} = \frac{4}{3}$$

Size (inch)	Width (mm)	Height (mm)	Diagonal (mm)	VGA Pixel Size (µm)
1	12.8	9.6	16	20
2/3	8.8	6,6	11	13.8
1/2	6.4	4.8	8	10
1/3	4.8	3.6	6	7.5
1/4	3.2	2.4	4	5

Main References

1) V. S. Nalwa, “A Guided Tour of Computer Vision”, Addison-Wesley Publishing Company, 1993.

2) R. Haralick, L. Shapiro “Computer and Robot Vision – Vol. II, Addison-Wesley Publishing Company, 1993.

3) B. Horn, “Robot Vision”, The MIT Press, 1986.

4) C. Steger, M. Ulrich, C. Wiedemann, “Machine Vision Algorithms and Applications”, Wiley-VCH, Weinheim, 2008.

5) “Standard for Characterization of Image Sensors and Cameras”, EMVA Standard 1288, European Machine Vision Association, Rel. 3.0, November 2010.

6) D. Litwiller, “CCD vs. CMOS: Facts and Fiction”, Photonics Spectra, Jan. 2001.