# Perspective projection model

J. A. Quiroga
aq@fis.ucm.es

# Perspective projection model

## Another imaging model?

the paraxial image formation generates a non-linear relation between object and image coordinates. If a point $M = (x, y, z)^T$ measured from the lens principal planes is imaged, the image point coordinates $M' = (x', y', z')^T$ are given by

$$z' = \frac{fz}{f+z}; x' = \frac{z'}{z}x; y' = \frac{z'}{z}y$$

$$(x', y', z') = \left(\frac{fx}{f+z}, \frac{fy}{f+z}, \frac{fz}{f+z}\right)^T$$

This imaging model is non-linear and it is not useful in metrology and computer vision . Here we are going to discuss a new imaging model, <u>the perspective projection</u>, for which the relation between the coordinates of object and image points is linear and we can easily perform the next tasks:

- Measure incident angles
- Measure the spatial position of a calibrated object
- Measure the spatial orientation of a plane
- Formulate the triangulation problem
- Formulate easily the imaging process of a complex 3D scene

# Perspective projection model

## Imaging as a perspective projection-the thin lens

We are going to change the sign criteria, so that the positive distances are measured from the lens center $C$ and the image plane is located at a negative distance. In this model the image plane is referred as the focal plane (no relation with the paraxial focal point). With positive lenses the image is inverted, and for this reason the $(x, y)$ axes of the focal plane are inverted. The intersection of the optical axis with the focal plane is the principal point $O$. We are going to call the distance from the image plane to the lens "focal distance" $f_d$ (no relation with the lens focal). The image plane is located at $z = -f_d$ in the lens reference system.

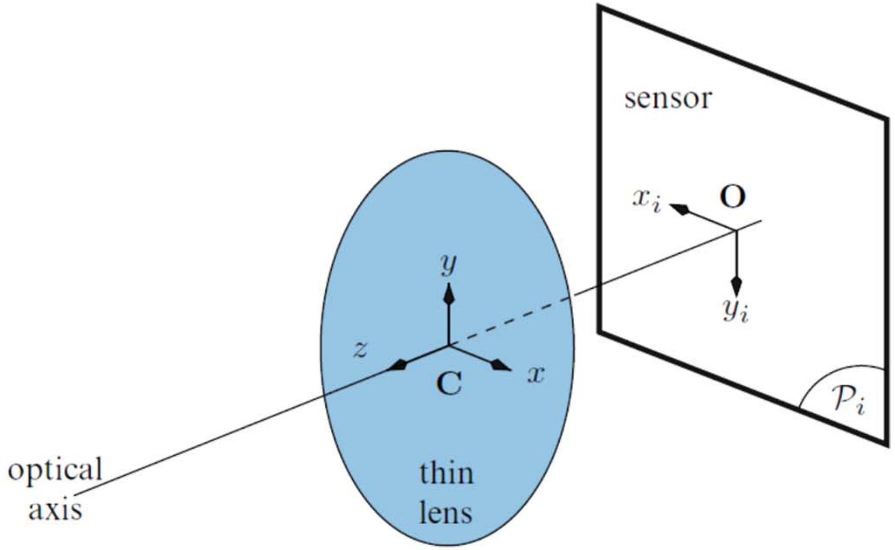



**Fig. 2.1** Thin lens with image/sensor plane and associated coordinate frames

Schreier, M. A. *Image Correlation for Shape, Motion and Deformation Measurements*.

# Perspective projection model

## Imaging as a perspective projection-the thin lens

Lets denote the spatial position of a object point as $M = (X, Y, Z)^T$. If we use a thin lens and we place the focal plane at the conjugate distance of $M$ the coordinates of the conjugate pair $\{M, M'\}$ are related by (beware of the sign change in $z$)

$$\frac{1}{Z} - \frac{1}{z} = \frac{1}{Z} + \frac{1}{f_d} = \frac{1}{f};$$

$$M' = (x, y, z) = \left(\frac{-f_d}{Z} X, \frac{-f_d}{Z} Y, -f_d\right)$$

That in matrix form is

$$M' = \begin{bmatrix} -\frac{f_d}{Z} & & \\ & -\frac{f_d}{Z} & \\ & & -f_d \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

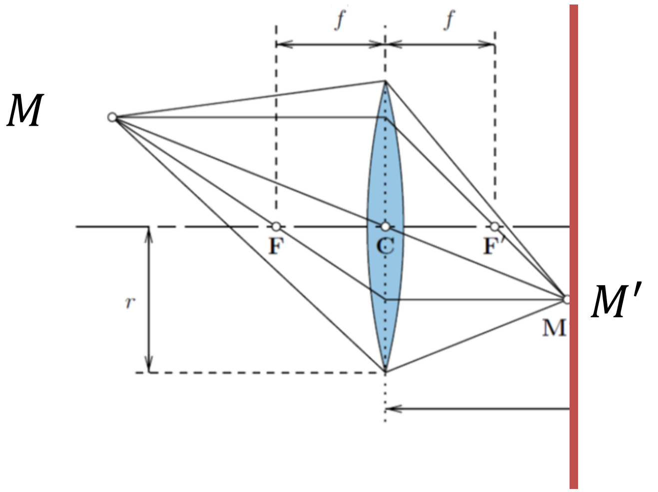



Fig. 2.2 A thin lens (side view)

$z = -f_d$

This imaging process is non-linear

However we can rewrite the former equations in homogenous coordinates

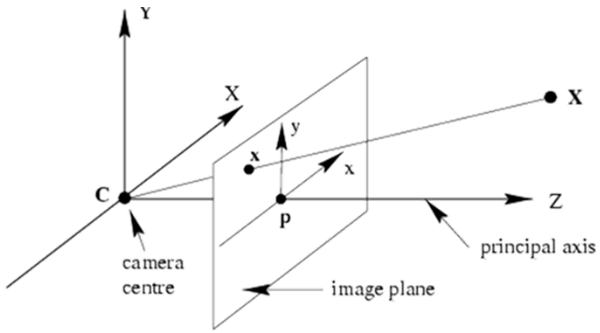
Schreier, M. A. *Image Correlation for Shape, Motion and Deformation Measurements.*

# Perspective projection model

## Homogenous coordinates

In mathematics, homogeneous coordinates, or projective coordinates, are used in projective geometry like Cartesian coordinates are used in Euclidean geometry. They are the natural way to describe the projection process through a central point.

| Cartesian (N) | Homogenous (N+1) |
|---|---|
| $\bar{M} = (x_1, \dots, x_N)$ | $M = (x_1, \dots, x_N, 1)$ |
| $\bar{M} = \left(\frac{h_1}{h_{N+1}}, \dots, \frac{h_N}{h_{N+1}}\right)$ | $M = (h_1, \dots, h_{N+1})$ |




In homogenous coordinates

$$P \cdot (X, Y, Z, 1)^T = P \cdot X = x = (x_1, x_2, x_3)^T$$
$$P \cdot s \cdot X = sx$$

<u>In both cases</u> the Cartesian coordinates in the plane are the same

$$\bar{x} = \left(\frac{x_1}{x_3}, \frac{x_2}{x_3}\right)$$

And we say $x \propto s \cdot x$ or $x \propto P \cdot X$

Hartley, *Multiple View Geometry in Computer Vision*

# Perspective projection model

## Imaging as a perspective projection-the thin lens

In homogenous coordinates the image point in the focal plane $z = -f_d$ is given by $m = (x, y, 1)^T$ , the object point is given by $M = (X, Y, Z, 1)^T$ and we can rewrite the imaging process as

$$sm = \begin{bmatrix} -f_d & & & 0 \\ & -f_d & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where $s = Z$

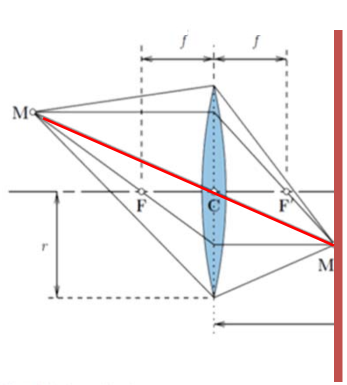



Fig. 2.2 A thin lens (side view)

For two conjugate planes, the use of homogenous coordinates linearize the imaging process. Actually, in this model the image of a point $\bar{M} = (X, Y, Z)$ is the intersection of the principal ray with the focal plane.

Additionally if both planes are conjugate we have

$$\frac{1}{Z} + \frac{1}{f_d} = \frac{1}{f};$$

Schreier, M. A. *Image Correlation for Shape, Motion and Deformation Measurements.*

# Perspective projection model

## Imaging as a perspective projection-the thin lens

$$s \cdot m = \begin{bmatrix} -f_d & & & 0 \\ & -f_d & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$s \cdot m = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} -Xf_d \\ -Yf_d \\ Z \end{bmatrix}$$

From the homogeneous coordinates $s \cdot m$ we can recover the Cartesian coordinates

$$x = \frac{h_1}{h_3} = -f_d \frac{X}{Z}$$

$$y = \frac{h_2}{h_3} = -f_d \frac{Y}{Z}$$

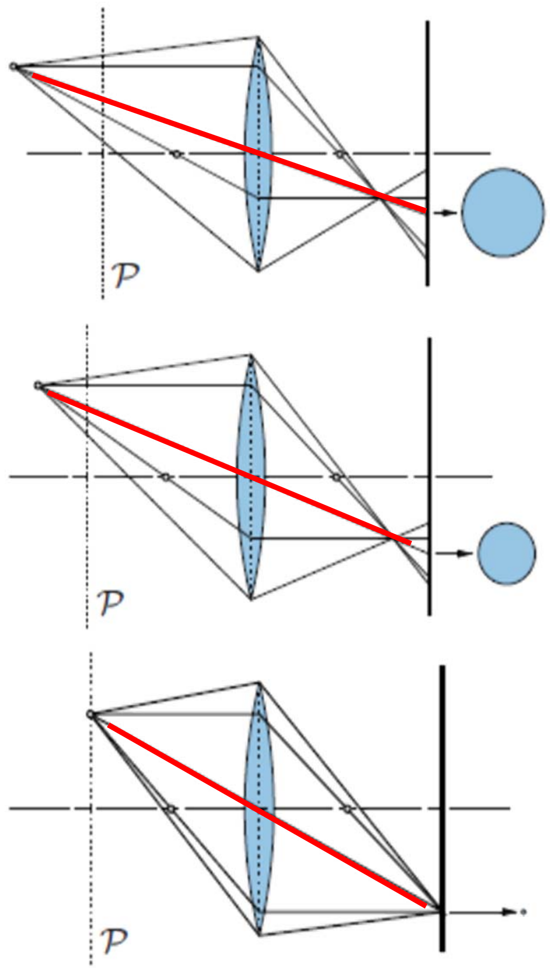
And by definition

$$z = -f_d$$

# Perspective projection model

## Imaging as a perspective projection-the thin lens

- If we position the focal plane in a non-conjugate position, the image of the point $M$ will be a blur circle.
- However the center of the circle, the intersection of the principal ray with the focal plane, will determine a perspective projection image point, for which the perspective imaging model is valid.
- We know that under some circumstances the DOFi+DOFo allows for a tolerable blur in the focal plane.
- Also the image is a blur circle because the EP and XP have the same size and are located at the principal planes.




Schreier, M. A. *Image Correlation for Shape, Motion and Deformation Measurements.*

# Perspective projection model

## Imaging as a perspective projection-the thin lens

In a general case if both planes are not conjugate, the center of the blur circle, $M'' = (x, y, -f_d)$ will follow the perspective projection imaging model

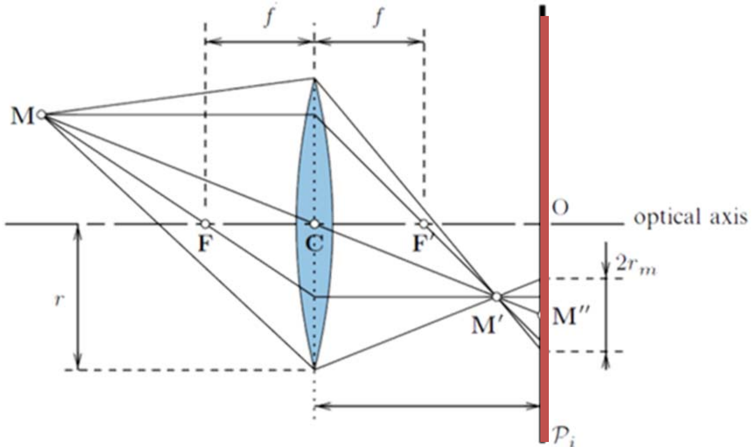
In homogenous coordinates the image point in the focal plane $z = -f_d$ is given by $m = (-x, -y, 1)^T$ and the object point is given by $M = (X, Y, Z, 1)^T$ and we can rewrite the imaging process as

$$sm = \begin{bmatrix} f_d & & & 0 \\ & f_d & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where $s = Z$

But in this general case

$$\frac{1}{Z} + \frac{1}{f_d} \neq \frac{1}{f};$$




$z = -f_d$

**Fig. 2.2** A thin lens (side view)

Schreier, M. A. *Image Correlation for Shape, Motion and Deformation Measurements.*

# Perspective projection model

## Imaging as a perspective projection

In a general case we have seen that an optical system can be described by its cardinal elements, principal planes, focal points and stops.

Image and object distances are measured from the nodal planes that (in air) are the intersection of the optical axis with the principal planes. <u>The reference $C$ is located at the object nodal point</u>

If both planes are not conjugate, the center of the blur circle, $M'' = (x, y, -f_d - d)$ will follow the perspective projection imaging model

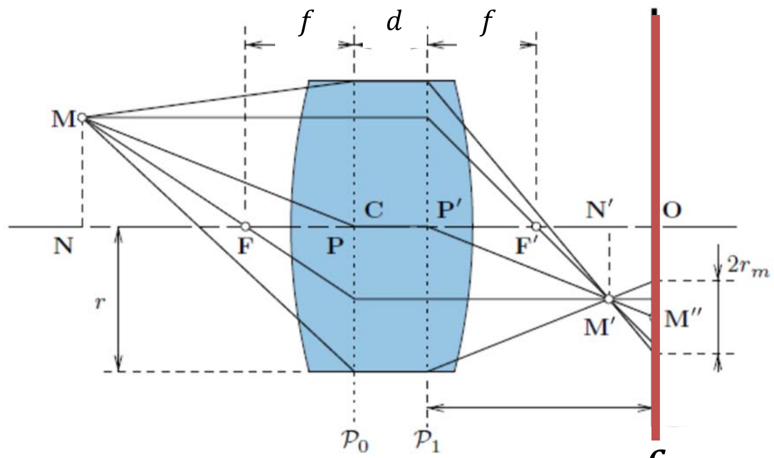
In homogenous coordinates the image point in the focal plane $z = -f_d - d$ is given by $m = (-x, -y, 1)^T$ and the object point is given by $M = (X, Y, Z, 1)^T$ and we can rewrite the imaging process as

$$sm = s\begin{bmatrix}-x\\-y\\1\end{bmatrix} = \begin{bmatrix}f_d & & & 0\\ & f_d & & 0\\ & & 1 & 0\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix}$$

where $s = Z$

But in this general case

$$\frac{1}{Z} + \frac{1}{f_d} \neq \frac{1}{f};$$




$z = -f_d$

Fig. 2.8 Image process and ray-tracing for a thick lens

Schreier, M. A. *Image Correlation for Shape, Motion and Deformation Measurements.*

# Perspective projection model

## Implicit assumptions

| | |
|---|---|
| The center of the blur is the intersection of the principal ray with the focal plane. | • The stops are placed at the principal planes and therefore $m_P = 1$<br>• There is no vignetting, the FS is located in a image plane. |
| The image of the point $\bar{M} = (X, Y, Z)^T$ is located at the intersection of the principal ray and the focal plane image | • We are in the DOFi region for our aperture, focal and object distance, |

If these assumptions are not valid

- There appear distortions.
- The image point $M''$ is at a different position than predicted by the principal ray
- The imaging process is not linear.

$$sm \neq PM$$

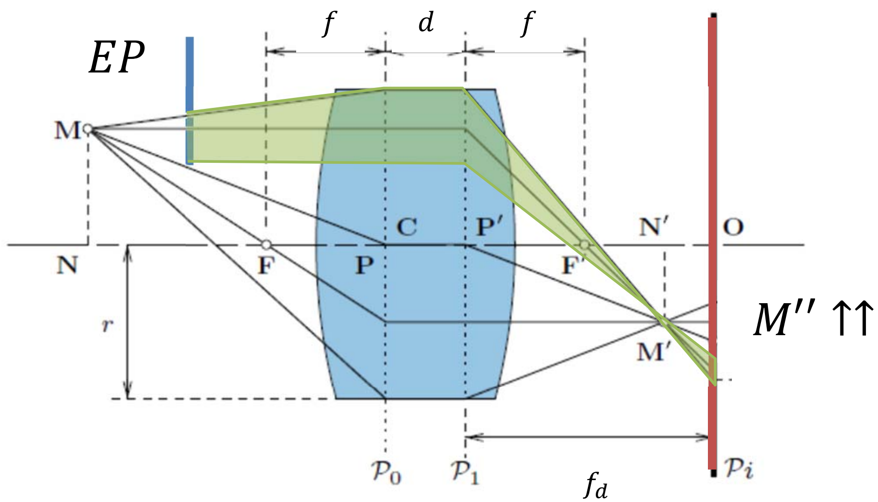



Fig. 2.8 Image process and ray-tracing for a thick lens

Schreier, M. A. *Image Correlation for Shape, Motion and Deformation Measurements.*

# Perspective projection model

## Front image plane

A last step in the model is to locate the focal image plane *in front of* the image center $C$. In this way we do not have image inversion, keep all the projection description, but the physical model it is not so clear.

In homogenous coordinates the object point is given by $M = (X, Y, Z, 1)^T$ measured in the reference system of $C$ and its image point in the focal plane $z = f_d$ is given by $m = (x, y, 1)^T$

$$sm = s\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_d & & & 0 \\ & f_d & & 0 \\ & & 1 & 0 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Where the (usually unknown) scale factor $s = Z$

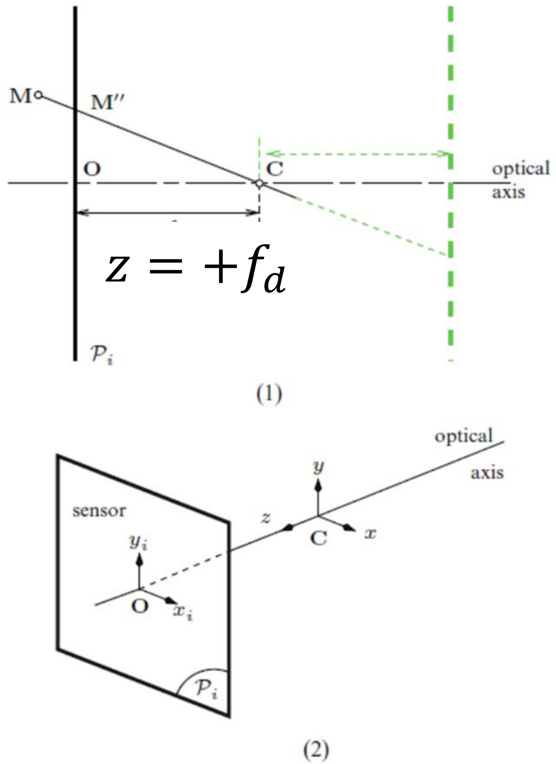


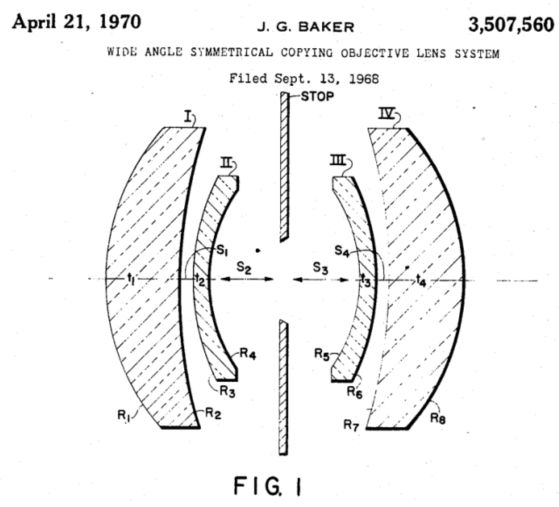

**Fig. 2.10** Front image plane model and associated coordinate frames. Image symmetry with respect to $x$-$y$ plane through the optical center, **C**, is assumed

Hartley, *Multiple View Geometry in Computer Vision*

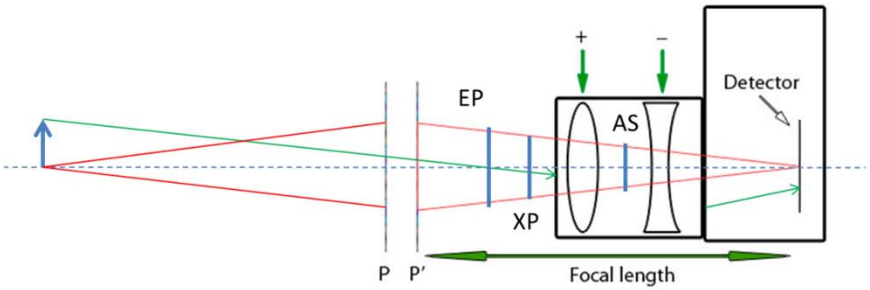
# Perspective projection model

**perspective projection in real world**

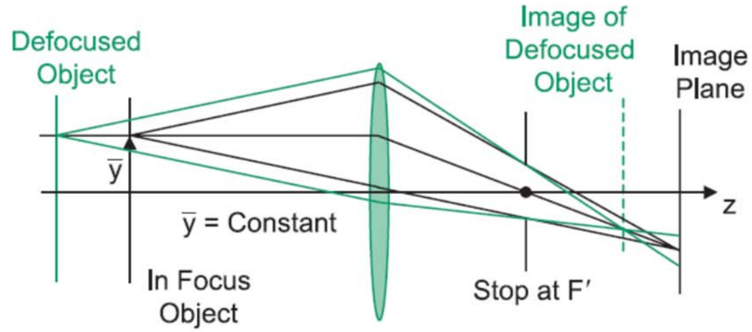
- Web cams (single thin lenses) OK
- Symmetrical objectives OK
- Telephoto NO
- Retro focus NO
- Telecentricity NO




http://www.google.com.na/patents/US3507560




http://coinimaging.com/telephoto.html




Field Guide to Geometrical Optics, John E. Greivenkamp

# Perspective projection model

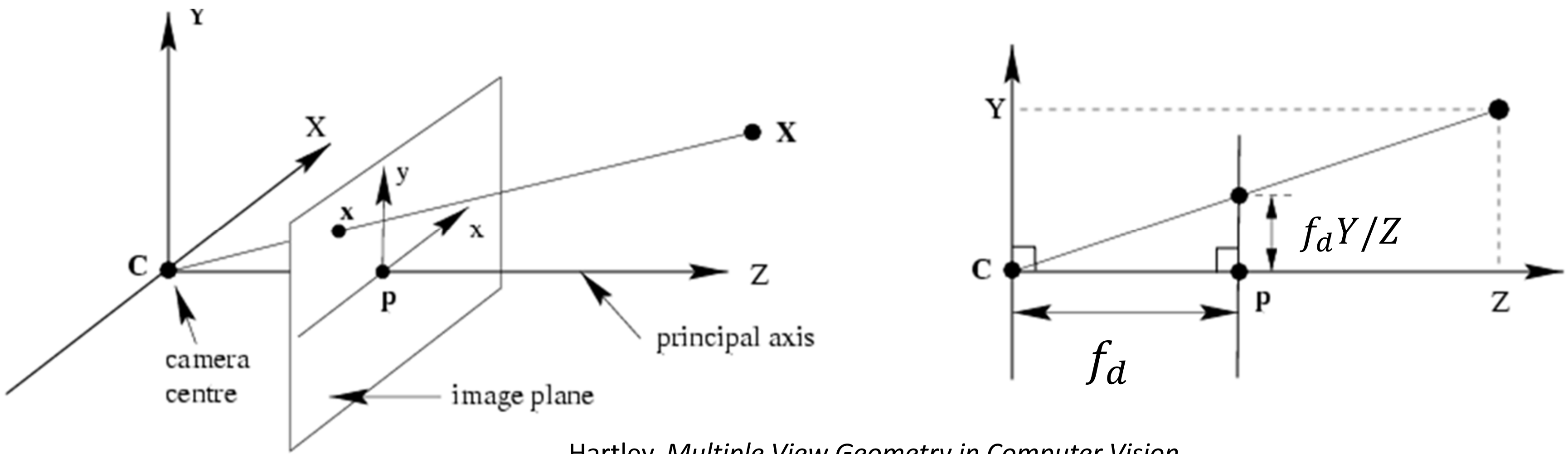
## Camera dissection

Under the projective projection model any imaging system is represented as a pinhole camera, with the image plane located is at distance $z = f_d$ from the camera center. Do not confuse with the focal of the optical system $f$. (most books denote $f_d$ by $f$, beware of the context)

If $M = (X, Y, Z, 1)^T$ is a object point in the camera reference system $(C)$ its image $m = (x, y, 1)^T$ is given by

$$sm = s\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_d & & & 0 \\ & f_d & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

In this equation all units are metric (i.e. $mm$)




Hartley, *Multiple View Geometry in Computer Vision*

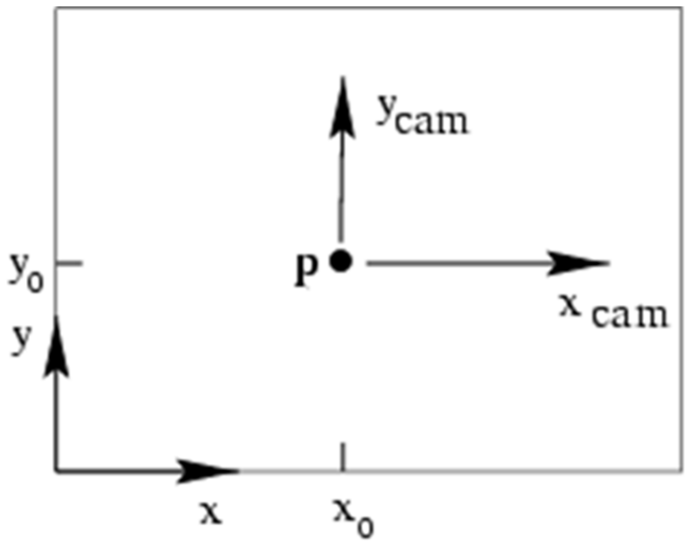
## Perspective projection model

### Principal point

The principal point is the intersection of the optical axis with the image plane. In the case of digital images captured with a CCD or CMOS the origin is not located a the image center but in the lower left corner.

If $p = (p_x, p_y)$ are the metric coordinates (i.e. mm) then

$$sm = s\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_d & & p_x & 0 \\ & f_d & p_y & 0 \\ & & 1 & 0 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$



# Perspective projection model

## From mm to px

The image captured by the CCD is spatially sampled and form a $N \times M$ matrix of irradiance samples

If $m_{x,y} = pixel/mm$ in $x$ and $y$ direction respectively

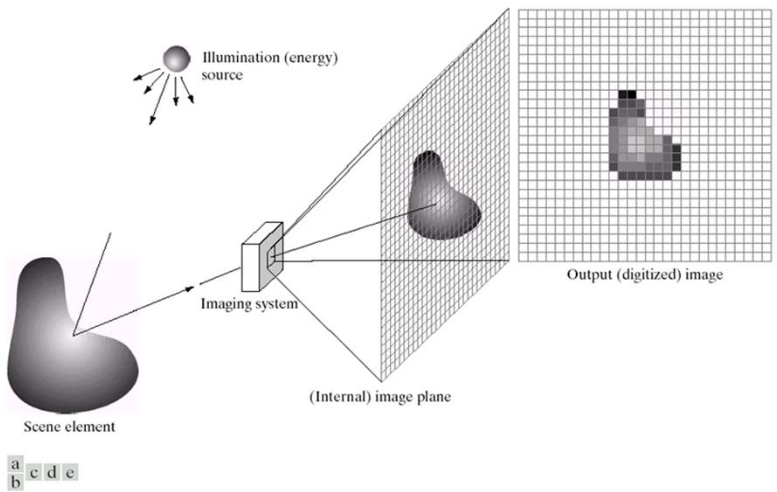
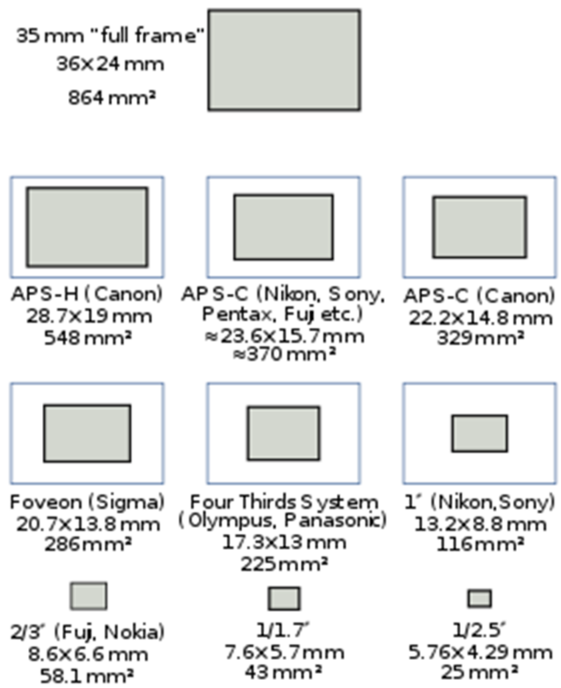



**FIGURE 2.15** An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.




https://en.wikipedia.org/wiki/Image_sensor_format

R. C. Gonzalez, R. E. Woods, Digital Image Processing (3rd Edition) Prentice-Hall, (2006)

# Perspective projection model

**From px to mm**

The image point location in mm is $m = (x, y, 1)$

And in px $m \equiv (u, v, 1) = \left(x \cdot m_x, y \cdot m_y, 1\right)$

$$sm = s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} m_x f_d & & m_x p_x & 0 \\ & m_y f_d & m_y p_y & 0 \\ & & 1 & 0 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

That can be rewritten as

$$\underbrace{sm}_{px} = \underbrace{\begin{bmatrix} f_x & & u_0 \\ & f_y & v_0 \\ & & 1 \end{bmatrix}}_{\substack{px \\ \text{Intrinsic} \\ \text{parameters}}} \left(\begin{array}{ccc|c} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{array}\right) \underbrace{M}_{mm} = K[I|0]M$$

$f_{x,y}$ focal distance in px $f_d = \frac{f_{x,y}}{m_{x,y}}$ in mm

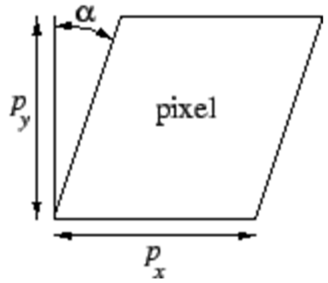
$(u_0, v_0)$ principal point in px and $K$ is the intrinsic parameters matrix

# Perspective projection model

## Skew

In some old imaging devices there was the possibility of non-square pixels, in this case we must add an extra intrinsic parameter named <u>the skew $\gamma = f_y \tan\alpha$</u>, with this parameter the intrinsic parameters matrix is

$$K = \begin{bmatrix} f_x & \gamma & u_0 \\ & f_y & v_0 \\ & & 1 \end{bmatrix}$$




$f_{x,y}$ focal distance in px

$(u_0, v_0)$ principal point in px

$\gamma$ pixel skew in px

Modern CCD cameras have $\gamma \approx 0$

# Perspective projection model

## Reference systems

It is very useful to have the capability to use a different reference system than the camera's reference. A point $\bar{M} = (X, Y, Z)^T$ in the world system has coordinates $\bar{M}_C = (X_C, Y_C, Z_C)^T$ in the camera frame. The camera has coordinates $\bar{C}$ in the world system. $R$ ($3 \times 3$ matrix) and $t$ ($3 \times 1$ vector) define the roto-traslation from world to camera reference systems

## Change of reference system

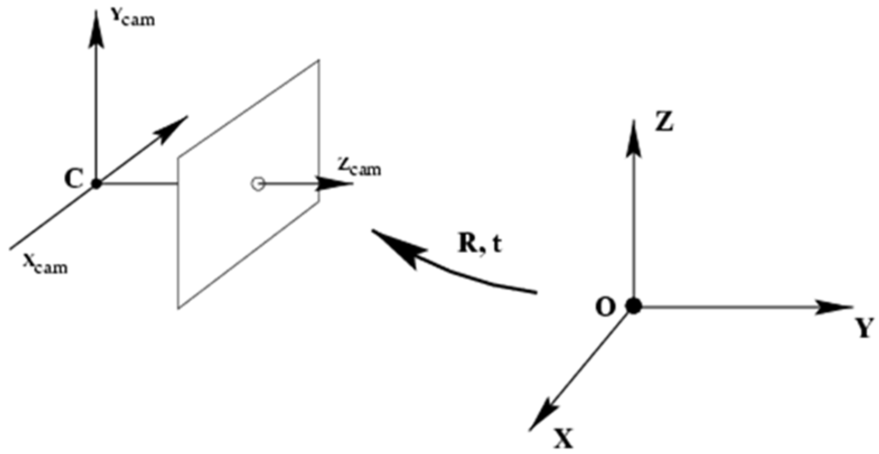
In Cartesian coordinates

$$\bar{M}_C = R \cdot (\bar{M} - \bar{C}) = R \cdot \bar{M} + t$$

with $t = -R \cdot \bar{C}$

In homogenous coordinates

$$M_C = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} M$$




$$sm = K[I|0]M_C = K[R|t]M = PM$$

$K \equiv$ intrinsic matrix, $3 \times 3$

$[R|t] \equiv$ extrinsic matrix, $3 \times 4$

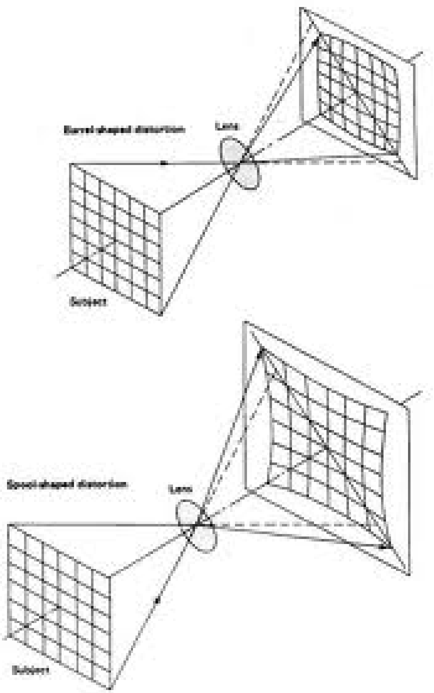
$P = K[R|t] \equiv$ Projection or camera matrix, $3 \times 4$

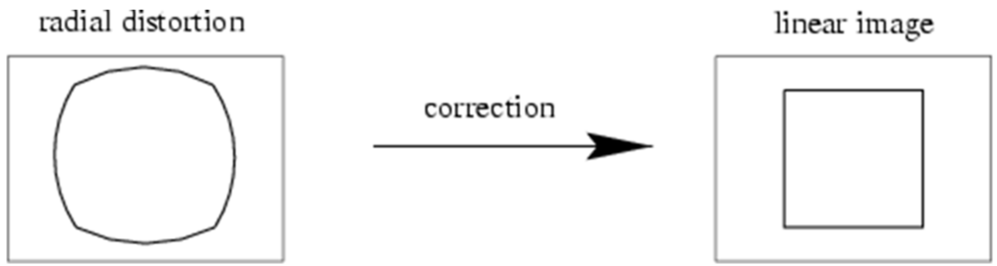
# Perspective projection model

## Distortion

Real camera system with inexpensive lenses suffer from distortion. The main reason is that for cheap systems the image of a plane is a curved surface

This makes the imaging process non-linear and introduces new parameters




$m \neq PM$ $m = PM$

## Perspective projection model

### Distortion model

### Normalized coordinates

We can divide the image formation in two steps. First we can project the object point $M_C$ in the camera reference system on a normalized camera with

$$f_d = 1 \Rightarrow K_N = I \Rightarrow P_N = K_N[I|0]$$

and obtain the normalized coordinates $N(M) = (x_n, y_n, 1)$ and then we can project the normalized units on the acual camera

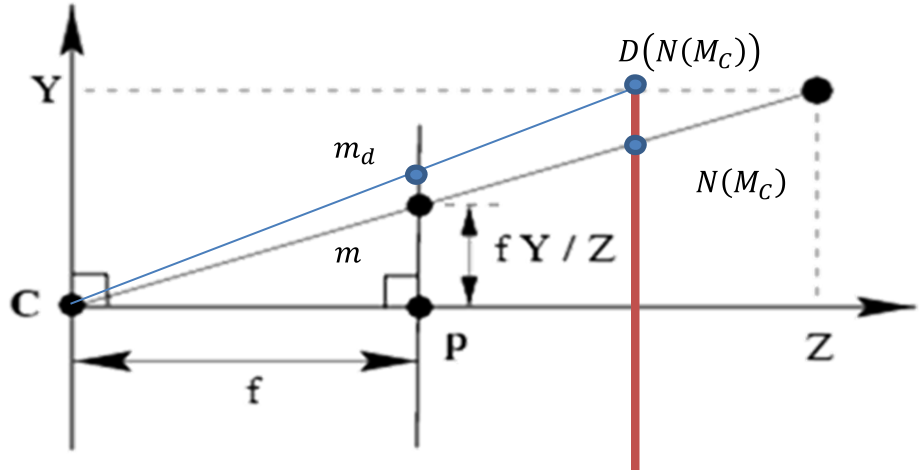
$$sm = K(K_N[I|0])M_C = K \begin{bmatrix} X_C/Z_C \\ Y_C/Z_C \\ 1 \end{bmatrix} = K \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix} = KN(M_C)$$

### Distortion

The distortion first change the normalized coordinates, and then they are projected on the linear camera

$$N(M_C) \rightarrow D(N)$$

$$sm_d = KD\big(N(M_C)\big)$$



# Perspective projection model

## Distortion function

In first approximation we can assume radial distortion with the center of the distortion located in the principal point of the camera

$$N(M_C) = \begin{bmatrix} X_C/Z_C \\ Y_C/Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix}$$

$$D(N) = \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = \begin{bmatrix} (1 + k_1 r^2 + k_2 r^4) x_n \\ (1 + k_1 r^2 + k_2 r^4) y_n \\ 1 \end{bmatrix} = \begin{bmatrix} L(r) x_n \\ L(r) y_n \\ 1 \end{bmatrix}$$

Where $r^2 = x_n^2 + y_n^2$

The more general model assumes radial as well as tangential distortion

$$\begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = \begin{bmatrix} (1 + k_1 r^2 + k_2 r^4) x_n + 2p_1 x_n y_n + p_2(r^2 + 2x_n) \\ (1 + k_1 r^2 + k_2 r^4) y_n + p_1(r^2 + 2y_n) + 2p_2 x_n y_n \\ 1 \end{bmatrix}$$

# Perspective projection model

## Distortion correction

Once the distortion parameters have been determined the distortion projection procedure is as follows

1) Retro-project the distorted image px to metric units

$$M_d = K^{-1} m_d = [X_d, Y_d, Z_d]^T$$

2) Normalize the metric retro-projected coordinates

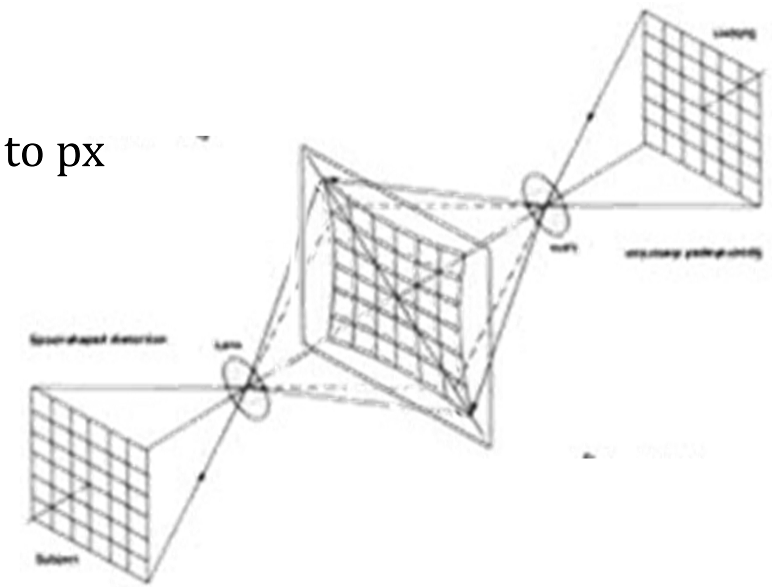
$$N_d = \left[\frac{X_d}{Z_d}, \frac{Y_d}{Z_d}, 1\right]^T$$

3) Apply distortion to $N_d$ ,

$$dN_d = D(N_d)$$

4) Project back the distorted image points to px

$$sm_c = K \cdot dN_d$$



# Perspective projection model

**Camera calibration**

The process to determine the intrinsic and extrinsic parameters of the perspective projection model is denominated camera calibration

Intrinsic parameters $K \rightarrow$ 5 DOF, $distortion \rightarrow$ 4 DOF

$$K = \begin{bmatrix} f_x & \gamma & u_0 \\ & f_y & v_0 \\ & & 1 \end{bmatrix}$$

$$\begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = \begin{bmatrix} (1 + k_1 r^2 + k_2 r^4) x_n + 2 p_1 x_n y_n + p_2 (r^2 + 2 x_n) \\ (1 + k_1 r^2 + k_2 r^4) y_n + p_1 (r^2 + 2 y_n) + 2 p_2 x_n y_n \\ 1 \end{bmatrix}$$

Extrinsic parameter $[R|t] \rightarrow$ 6 DOF

**Calibration process**

Given a set of known object and image points $\{m, M\}_l$

Determine the projection matrix $P = K\ [R|t]$ and distortion parameters that minimize

$$U = \sum_{l=1}^{L} \left(m_l - K \cdot D\left(N([R|t] M_l)\right)\right)^2$$

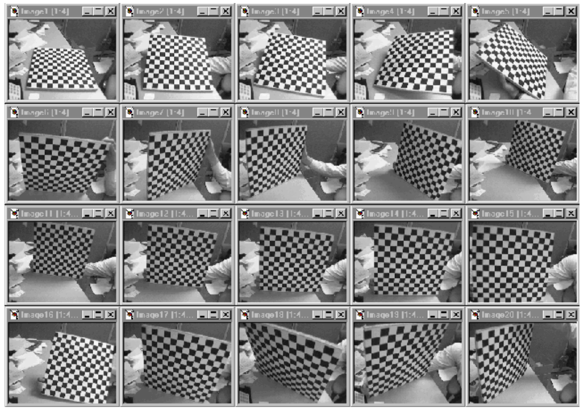
Camera calibration toolbox for MATLAB

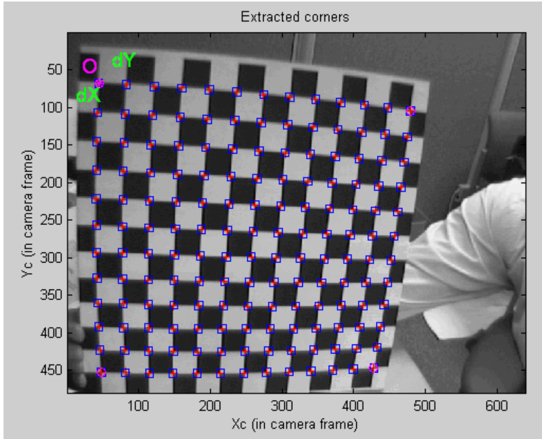
http://www.vision.caltech.edu/bouguetj/calib_doc/

# Perspective projection model

## Camera calibration

http://www.vision.caltech.edu/bouguetj/calib_doc/

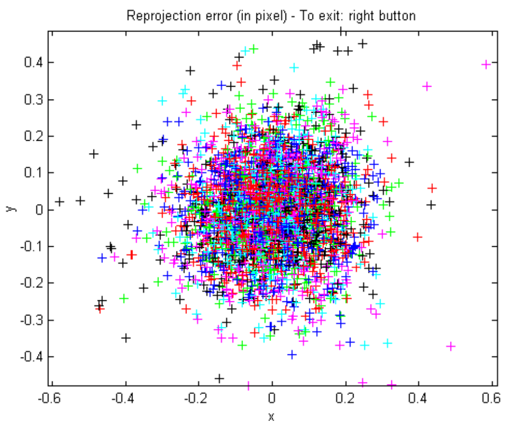


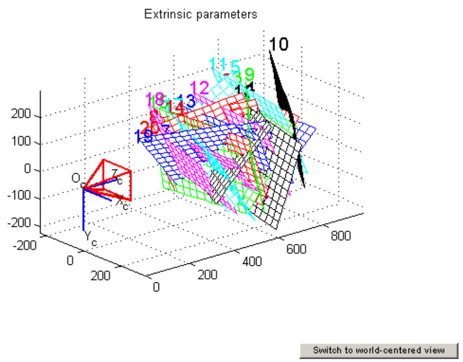


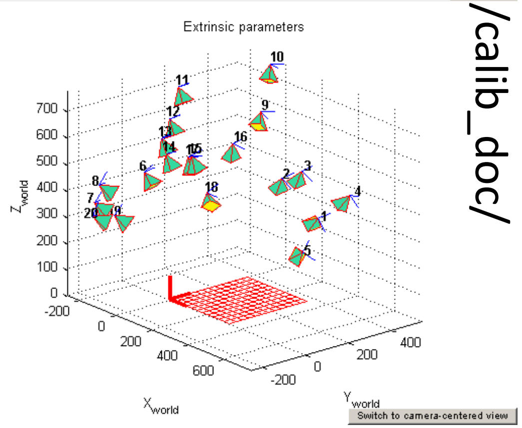

```
Calibration results after optimization (with uncertainties):

Focal Length:          fc = [ 657.46290   657.94673 ] ± [ 0.31819   0.34046 ]
Principal point:       cc = [ 303.13665   242.56935 ] ± [ 0.64682   0.59218 ]
Skew:             alpha_c = [ 0.00000 ] ± [ 0.00000  ]   => angle of pixel axes = 90.00000 ± 0.00000 degrees
Distortion:            kc = [ -0.25403   0.12143   -0.00021   0.00002  0.00000 ] ± [ 0.00248   0.00986   0.00013   0.00013  0.00000 ]
Pixel error:          err = [ 0.11689   0.11500 ]

Note: The numerical errors are approximately three times the standard deviations (for reference).
```



# Perspective projection model

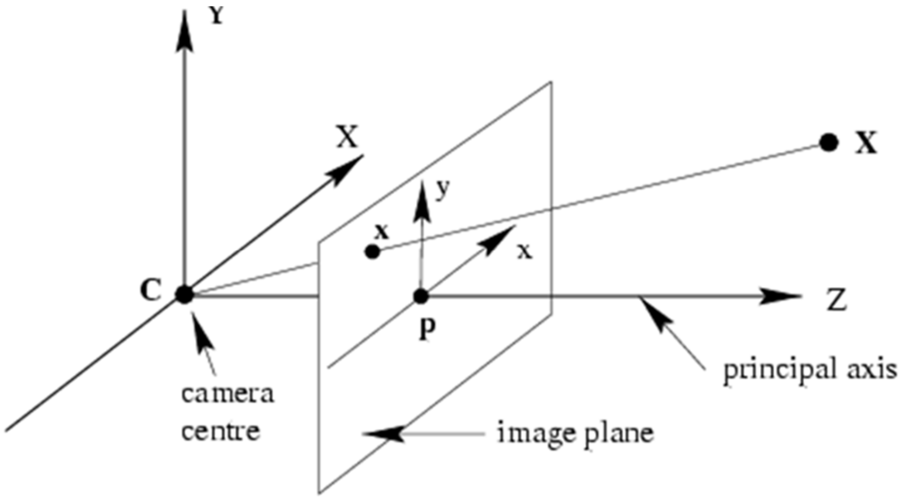
## Task

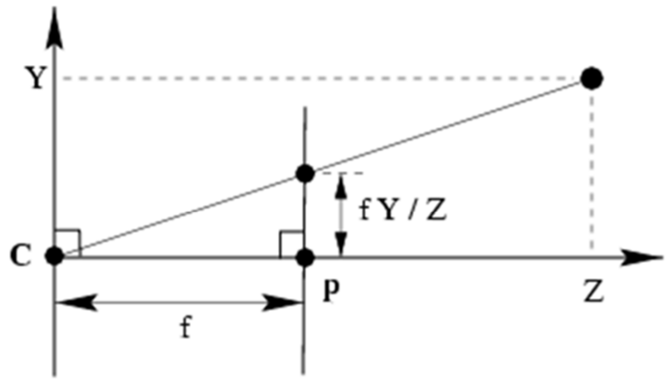
use the next intrinsic parameters to determine the image position in pixels of the corners of a square with size 10 $mm$ normal to the optical axis located at $z = 50\ cm$ from the camera. Repeat with the square located at $z = 75\ cm$. If the camera pixel size is 5 $\mu m$ what is the focal distance in mm? (for simplicity do not include distortion)

```
Calibration results after optimization (with uncertainties):

Focal Length:          fc = [ 657.46290   657.94673 ] ± [ 0.31819   0.34046 ]
Principal point:       cc = [ 303.13665   242.56935 ] ± [ 0.64682   0.59218 ]
Skew:             alpha_c = [ 0.00000 ] ± [ 0.00000  ]   => angle of pixel axes = 90.00000 ± 0.00000 degrees
Distortion:            kc = [ -0.25403   0.12143   -0.00021   0.00002  0.00000 ] ± [ 0.00248   0.00986   0.00013   0.00013  0.00000 ]
Pixel error:          err = [ 0.11689   0.11500 ]

Note: The numerical errors are approximately three times the standard deviations (for reference).
```



# Perspective projection model

**Linear triangulation**

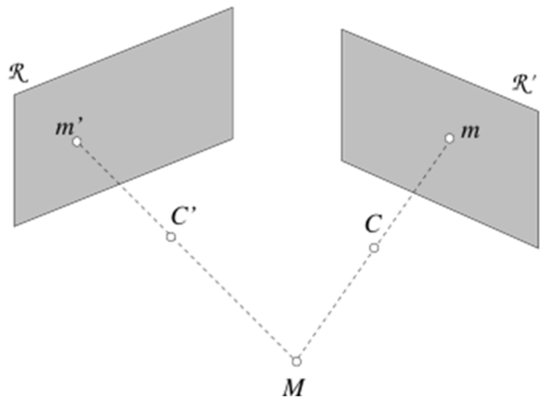
Using calibration procedures it is possible the triangulation of an object point from its image points. The calibration will determine the two projection matrixes $P_1 = K_1[I|0]$ and $P_2 = K_2[R|t]$, where $[R|t]$ is the roto-traslation from $C_2$ to $C_1$

The triangulation can be stated as to find the $M = [X, Y, Z, 1]^T$ that fulfil

$$m_1 = P_1 M; \;\; m_2 = P_2 M$$

Taking into account that $m_1 \times P_1 M = 0 \Rightarrow$

$$u(p_1^{3T} M) - (p_1^{1T} M) = 0$$
$$v(p_1^{3T} M) - (p_1^{2T} M) = 0$$
$$u(p_1^{2T} M) - v(p_1^{1T} M) = 0$$




where $p_n^{iT}$ are the rows of $P_n$. These equations are linear in the components of $M$. If we use $m_2 \times P_2 M = 0$, an equation of the form $AM = 0$ can be composed with

$$A = \begin{bmatrix} up_1^{3T} - p_1^{1T} \\ vp_1^{3T} - p_1^{2T} \\ up_2^{3T} - p_2^{1T} \\ up_2^{3T} - p_2^{2T} \end{bmatrix}$$

Where two equations have been included for each image, giving a total of 4 equations for 3 unknowns

# Perspective projection model

## Applications

- Visión estéreo para seguir los estorninos en un bando, Link

## References


- Schreier, H., Orteu, J.-J., & Sutton, M. A. (2009). *Image Correlation for Shape, Motion and Deformation Measurements*. Boston, MA: Springer US. http://doi.org/10.1007/978-0-387-78747-3
- Hartley, R., & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press. Retrieved from https://books.google.es/books?id=si3R3Pfa98QC