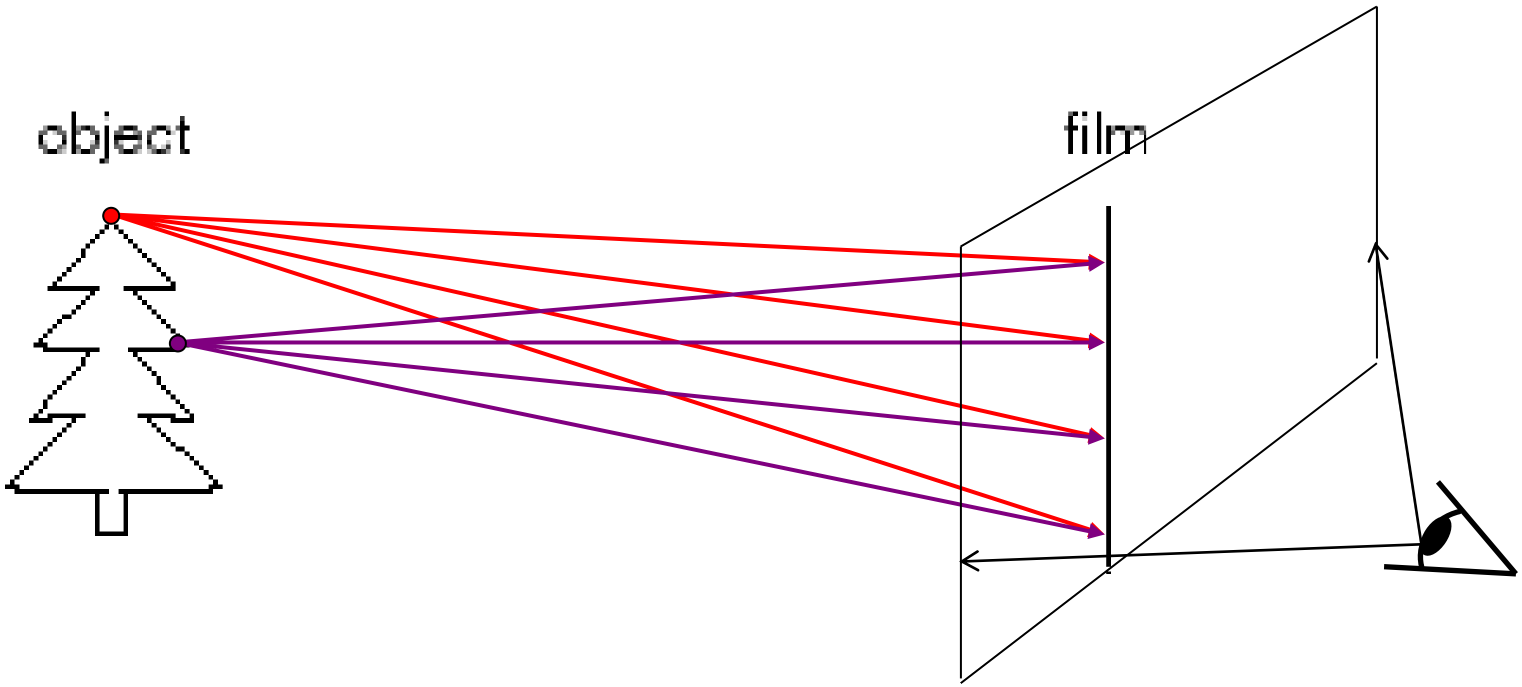
Modeling and Camera Calibration

Overview

- Modeling cameras
  - Pinhole camera
  - Lens
- Projective geometry
  - Homogenous coordinates
- Camera calibration
  - Intrinsic parameters
  - Extrinsic parameters

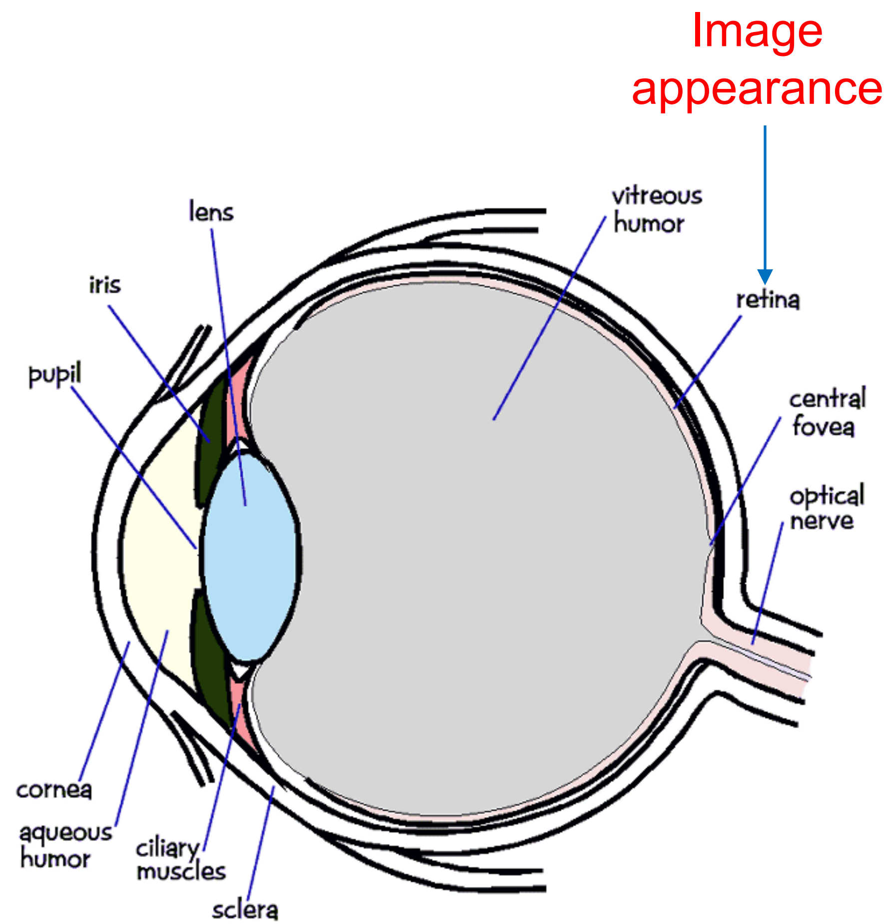
How can we generate one image ?



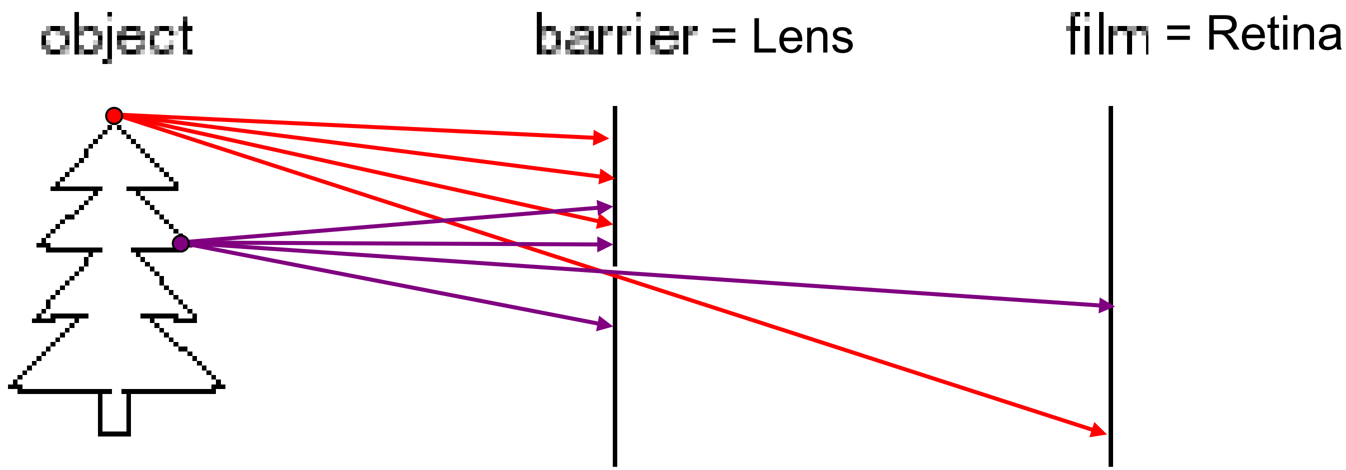
- How to capture one picture.
  - Simple idea: put a piece of transparent film between eye and object.
  - Is it a real image ?

The Eye

- The photosensitive part of the eye is called the **retina**.
- The retina is largely composed of two types of cells
  - ***Rods***: light sensitive
  - ***Cones***: responsible for color perception.

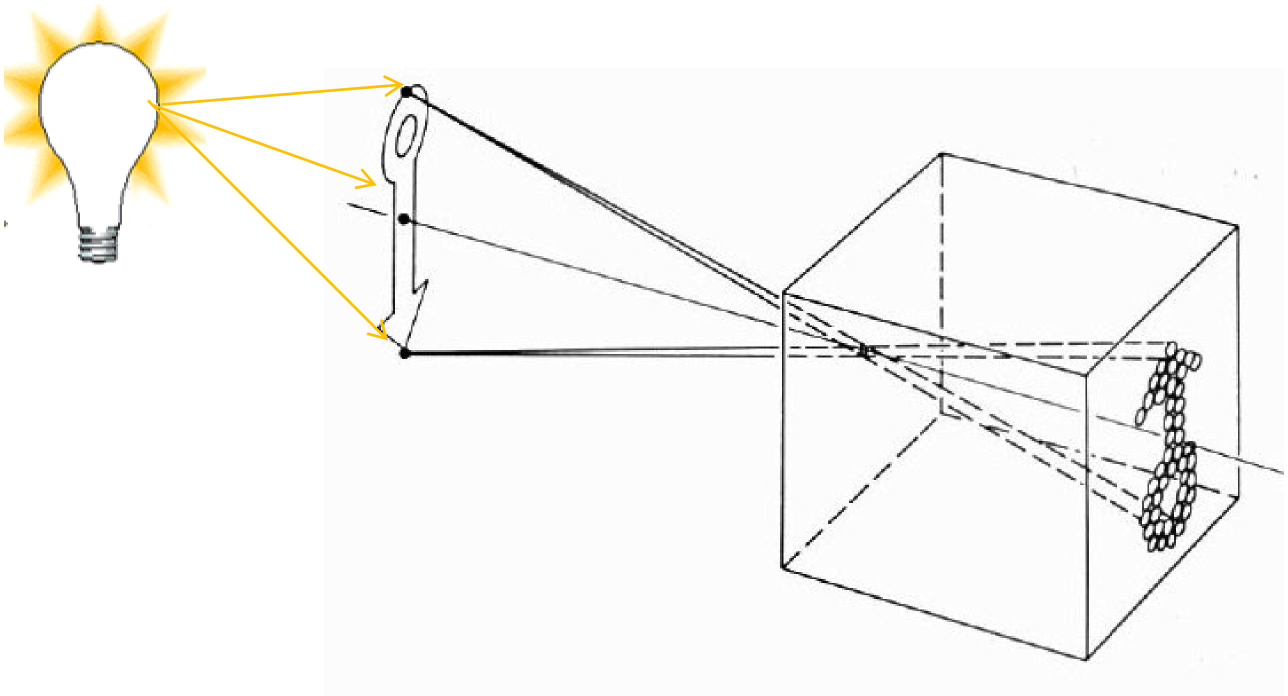


Pinhole camera



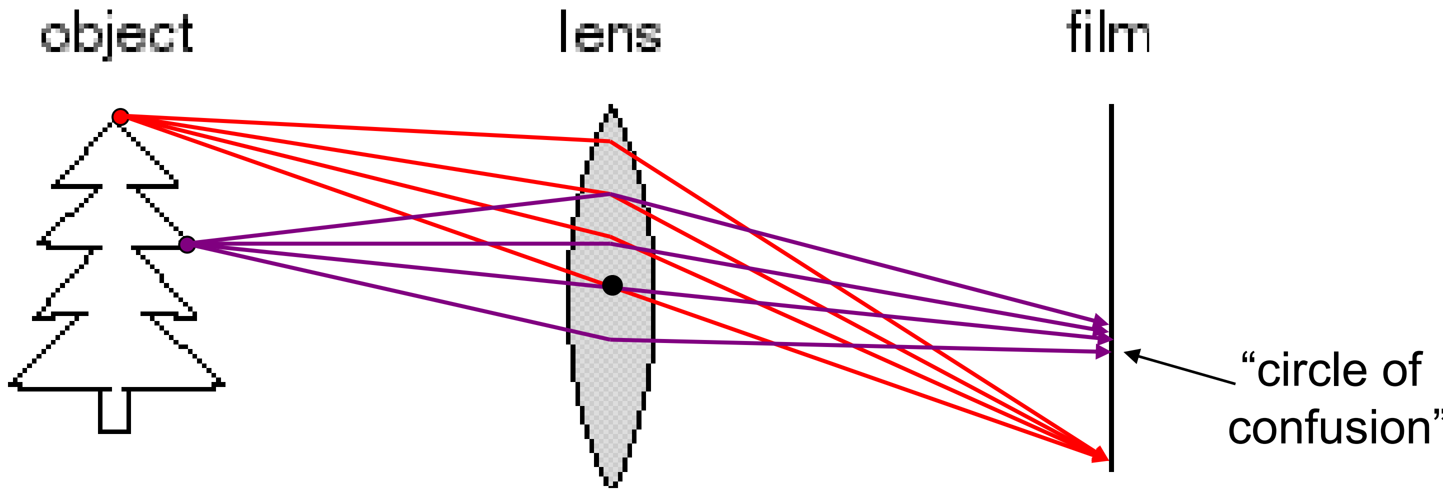
- Add a barrier to block off most of the lighting rays
  - It helps to reduce blurring
  - The opening of barrier is known as the **aperture**
  - **Lighting rays draw an image on film by chemical reaction (silver-halide crystal).**
  - How does this transform the image?

Pinhole Camera's Image



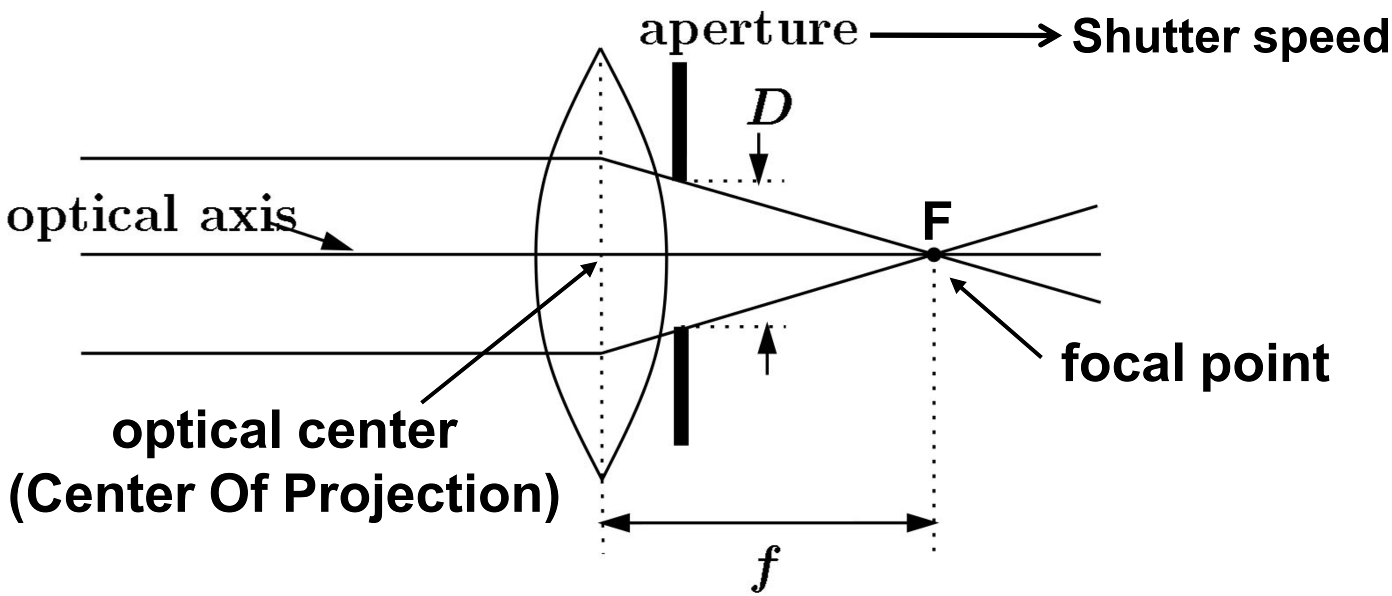
- The first camera
  - Known to Aristotle
  - How does the aperture size affect the image?

Camera lens



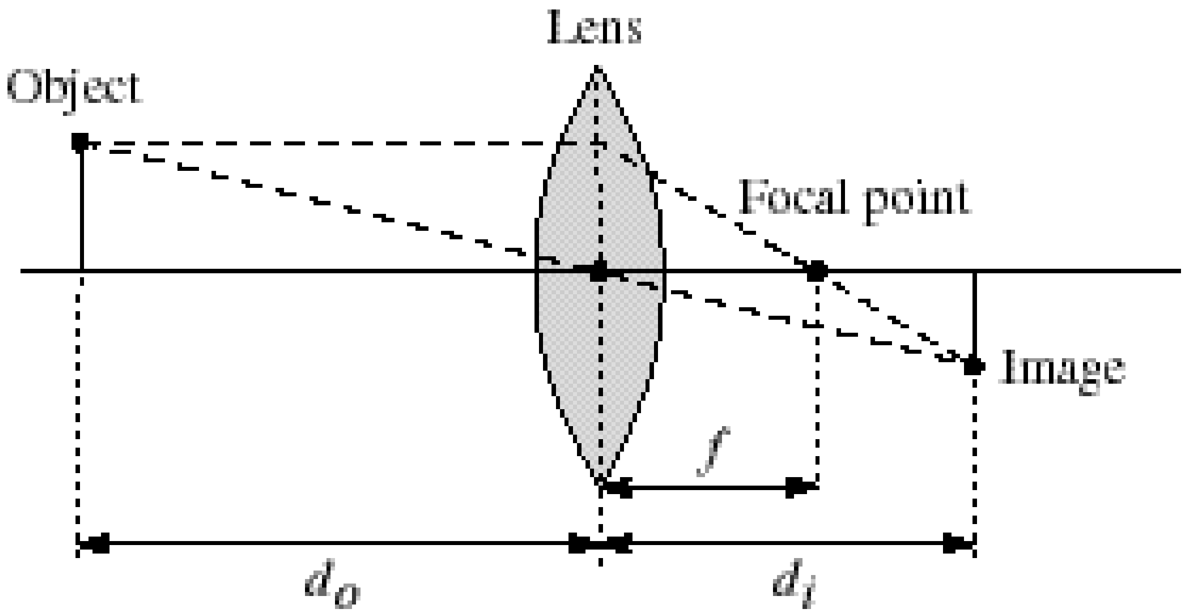
- A lens focuses light onto the film
  - There is a specific distance at which objects are “in focus”
    - When the objects do not stay “in focus” of camera, the object points are projected to a “circle of confusion” instead of point in the image. The result is one blurred image.
  - Changing the shape of the lens changes the focus length

Camera Lens



- A lens focuses parallel rays onto a single focal point
  - focal point at a distance f beyond the plane of the lens
    - f is a function of the shape and index of refraction of the lens
  - Aperture of diameter D restricts the range of rays
    - aperture may be on either side of the lens
  - Lenses are typically spherical (easier to produce)

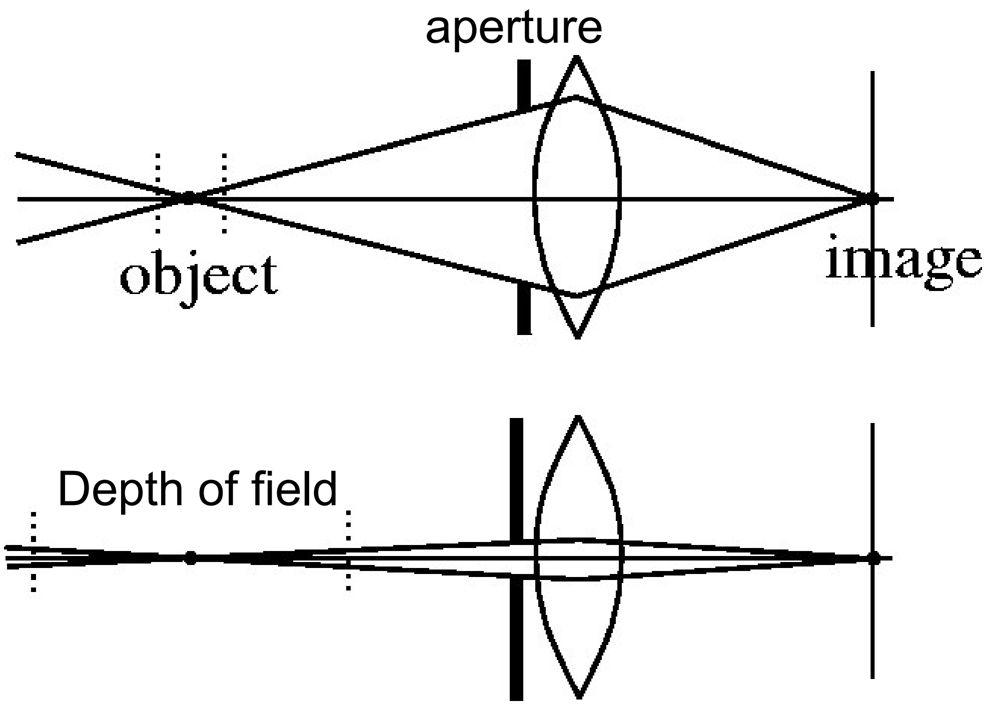
Camera lens



http://www.phy.ntnu.edu.tw/java/Lens/lens_e.html

- Thin lens equation: $$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$
  - Any object point satisfying this equation is in focus

Depth of field



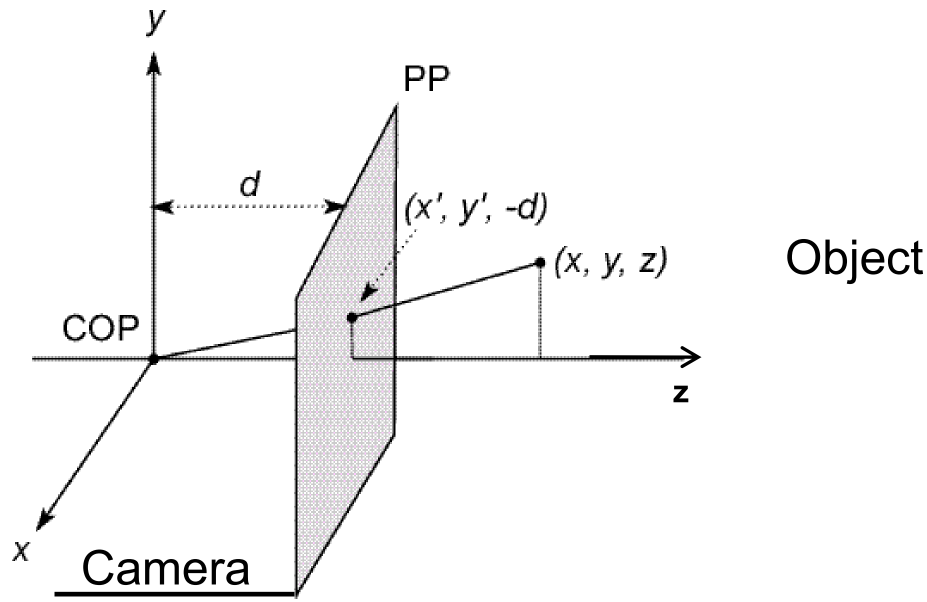
Aperture = *f* / 5.6

f / 32

http://en.wikipedia.org/wiki/Depth_of_field

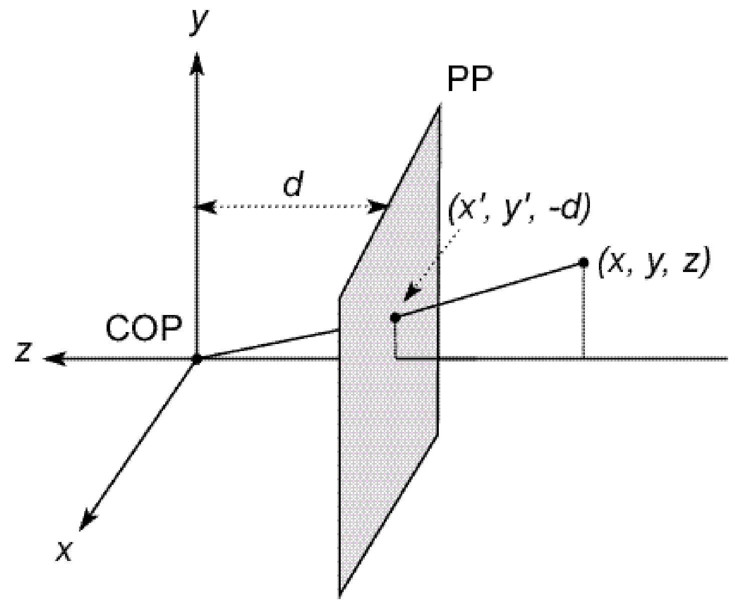
- Changing the aperture size affects depth of field
  - A smaller aperture increases the range in which the object is approximately in focus

Modeling projection



- The coordinate system
  - The pin-hole model is utilized as an approximation
  - The optical center (**C**enter **O**f **P**rojection) is at the origin
  - The image plane (**P**rojection **P**lane) is *in front* of the COP

Modeling projection



- Projection equations
  - Compute intersection with PP of ray from (x,y,z) to COP
  - Derived using similar triangles (on board)

$$(x, y, z) \rightarrow (-d\frac{x}{z},\ -d\frac{y}{z},\ -d)$$

- We get the projection by throwing out the last coordinate:

$$(x, y, z) \rightarrow (-d\frac{x}{z},\ -d\frac{y}{z})$$

Homogeneous coordinates

- Is this a linear transformation?
  - no—division by z is nonlinear

Trick: add one more coordinate:

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

Converting *from* homogeneous coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

Perspective Projection

- Projection is a matrix multiply using homogeneous coordinates:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

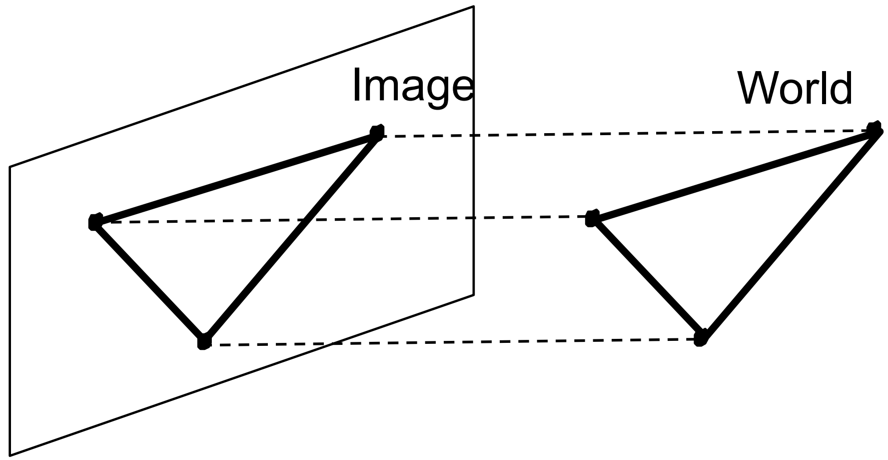
divide by third coordinate

By the other transform

$$\begin{bmatrix} -d & 0 & 0 & 0 \\ 0 & -d & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} -dx \\ -dy \\ z \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

Orthographic projection

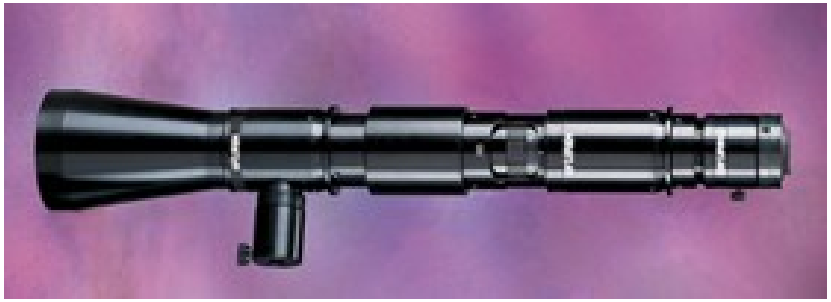
- Special case of perspective projection
  - Distance from the COP to the PP is infinite



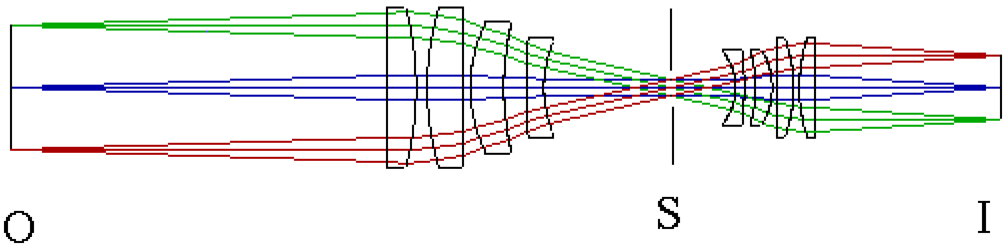
  - Good approximation for telephoto optics
  - Also called "parallel projection": (x, y, z) → (x, y)
  - Projection matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Rightarrow (x, y)$$

Orthographic ("telecentric") lenses



Navitar telecentric zoom lens



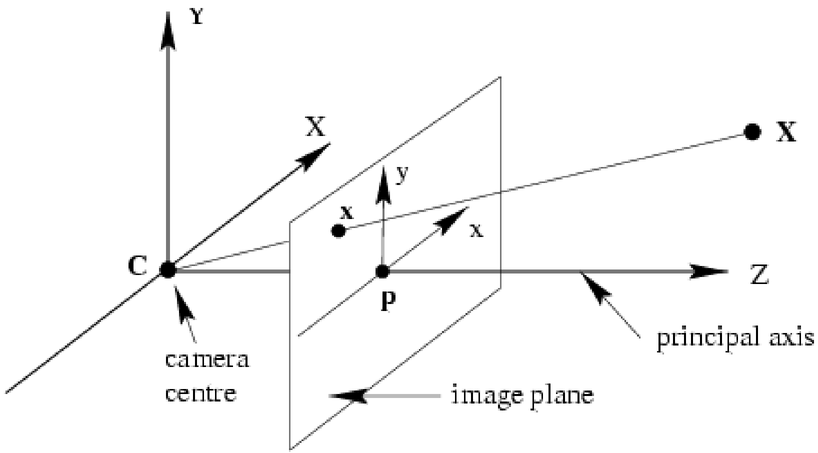
Objects

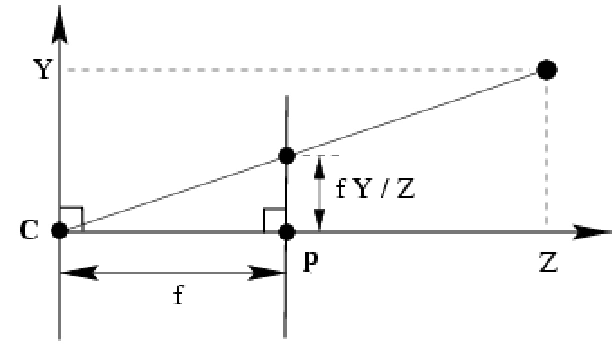
Diaphragm stop

Images

http://www.lhup.edu/~dsimanek/3d/telecent.htm

Pinhole camera: Projection Matrix





$$(X, Y, Z) \mapsto (f X / Z, f Y / Z)$$

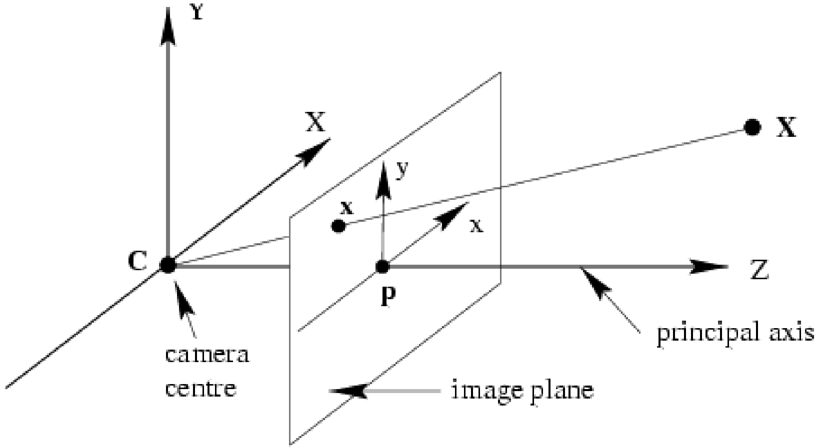
$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

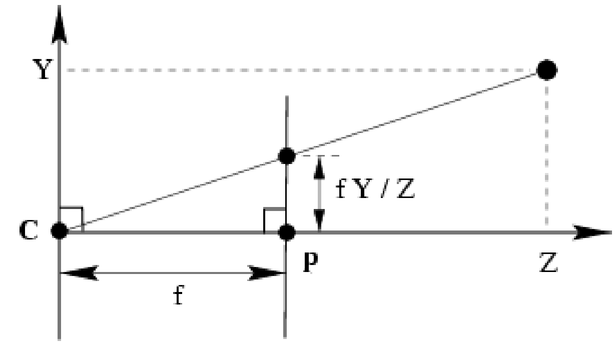
3D real coordinate

$$\mathrm{x} = \mathrm{P}\mathrm{X}$$

Image plane coordinate

Pinhole camera: Projection Matrix

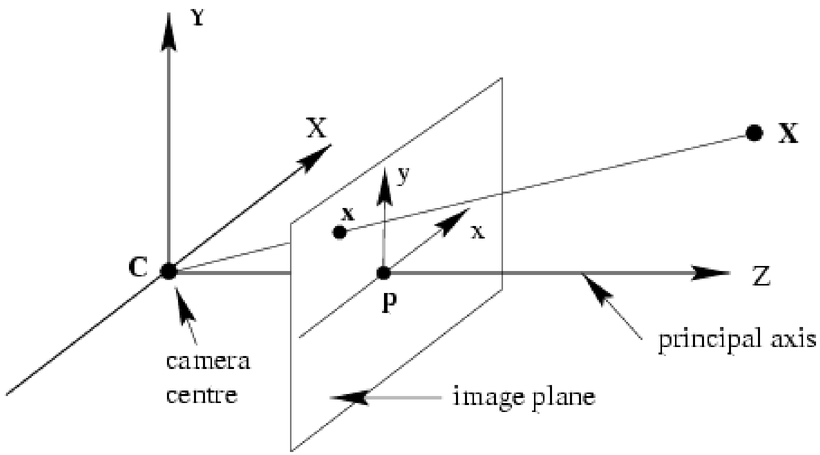




$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \underbrace{\begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix}}_{\text{Camera parameter}} \underbrace{\begin{bmatrix} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{bmatrix}}_{\text{Known}} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

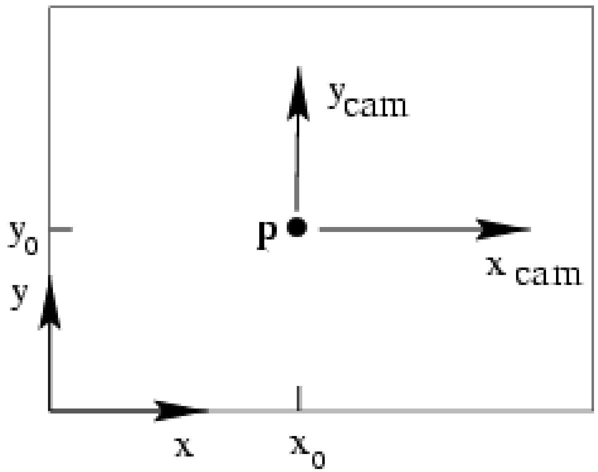
$$\mathrm{x} = \mathrm{PX} \Longrightarrow \mathrm{P} = \mathrm{diag}(f, f, 1)[\mathrm{I} \mid 0]$$

Camera coordinate system



- **Principal axis:** line from the camera center perpendicular to the image plane
- **Normalized (camera) coordinate system:** camera center is at the origin and the principal axis is the z-axis
- **Principal point (p):** point where principal axis intersects the image plane (origin of normalized coordinate system)

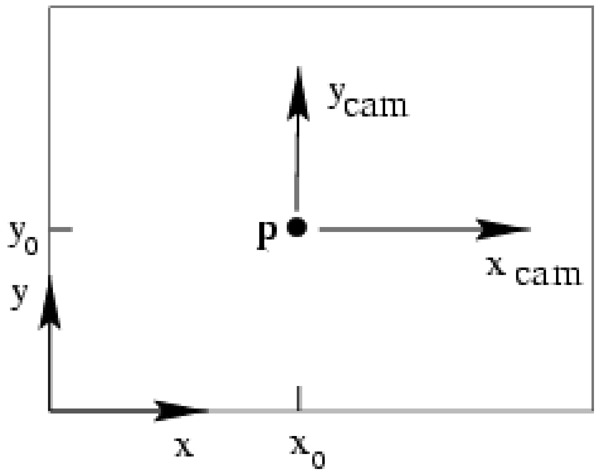
Principal point offset



principal point: (p_x, p_y)

- Camera coordinate system: origin is at the prinicipal point
- Image coordinate system: origin is in the corner

Principal point offset

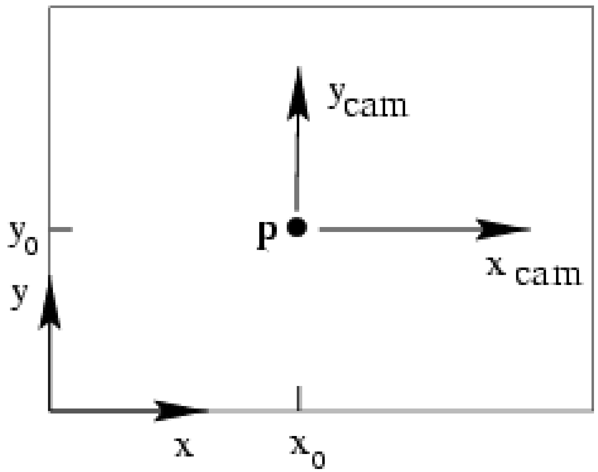


principal point: (p_x, p_y)

$$(X,Y,Z) \mapsto (f\,X/Z + p_x, f\,Y/Z + p_y)$$

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f\,X + Z\,p_x \\ f\,Y + Z\,p_x \\ Z \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Principal point offset

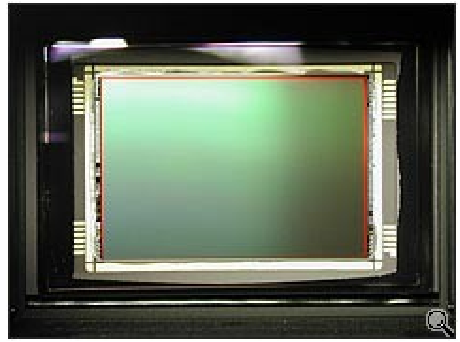


principal point: (p_x, p_y)

$$\begin{pmatrix} fX + Zp_x \\ fY + Zp_x \\ Z \end{pmatrix} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}$$: calibration matrix and $\mathrm{P} = \mathrm{K}[\mathrm{I} \mid 0]$

Intrinsic parameters



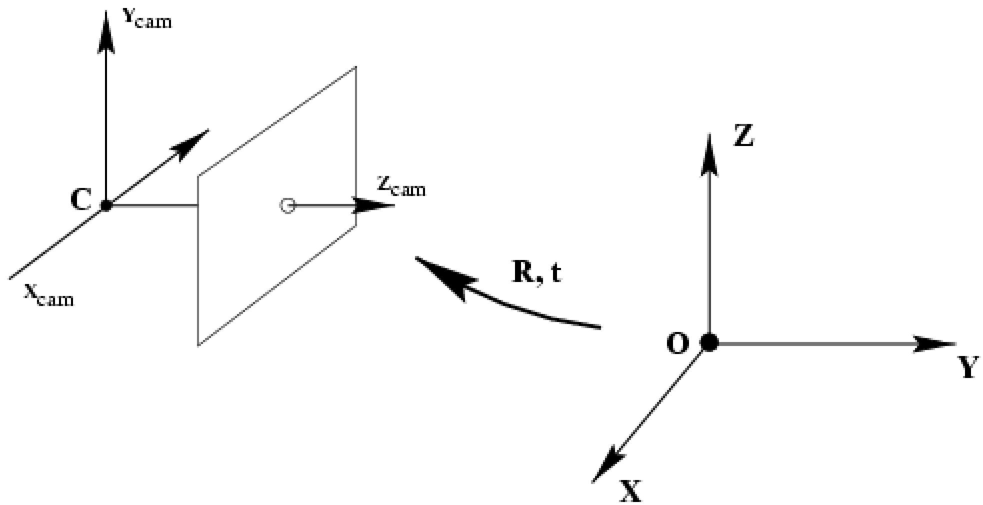
Pixel size: $\frac{1}{m_x} \times \frac{1}{m_y}$

- m_x pixels per meter in horizontal direction, m_y pixels per meter in vertical direction

$$K = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & & \beta_x \\ & \alpha_y & \beta_y \\ & & 1 \end{bmatrix}$$

pixels/m   m   pixels

Camera rotation and translation



- In general, the camera coordinate frame will be related to the world coordinate frame by a rotation and a translation

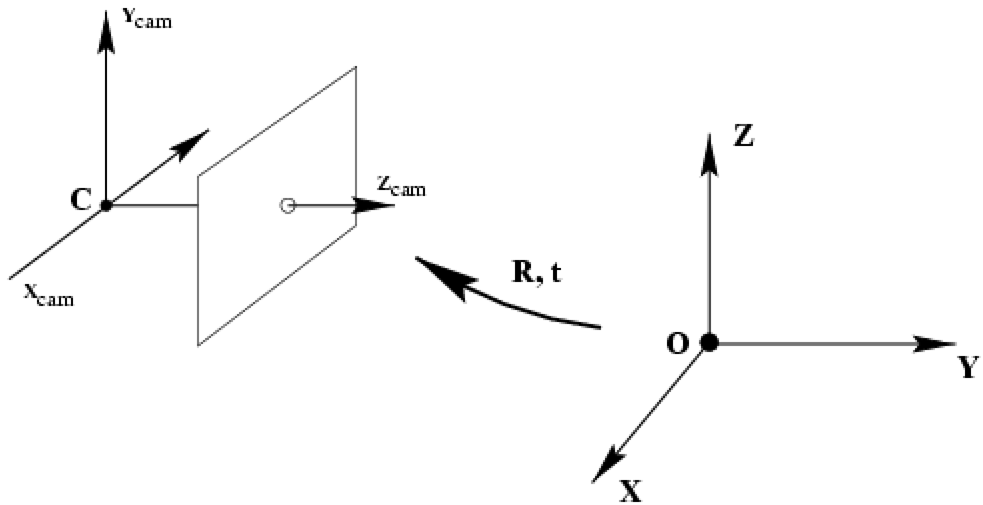
$$\tilde{X}_{cam} = R\left(\tilde{X} - \tilde{C}\right)$$

coords. of point in camera frame

coords. of a point in world frame (nonhomogeneous)

coords. of camera center in world frame

Extrinsic parameters



In non-homogeneous coordinates:

$$\tilde{X}_{cam} = R\left(\tilde{X} - \tilde{C}\right)$$

$$X_{cam} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \tilde{X} \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} X$$

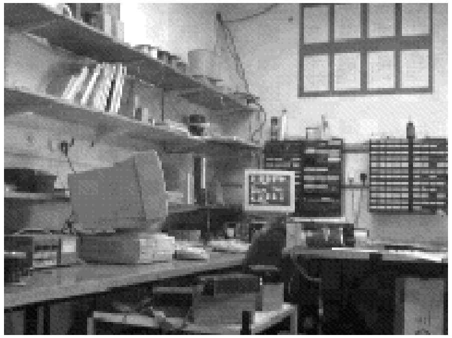
$$x = K[I \mid 0]X_{cam} = K\left[R \mid -R\tilde{C}\right]X \qquad P = K[R \mid t], \qquad t = -R\tilde{C}$$

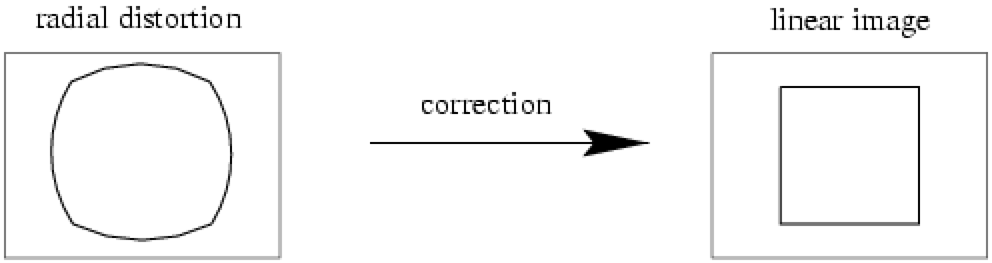
Note: C is the null space of the camera projection matrix (PC=0)

Camera parameters

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - *Skew (non-rectangular pixels)*
  - *Radial distortion*

$$K = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & & \beta_x \\ & \alpha_y & \beta_y \\ & & 1 \end{bmatrix}$$

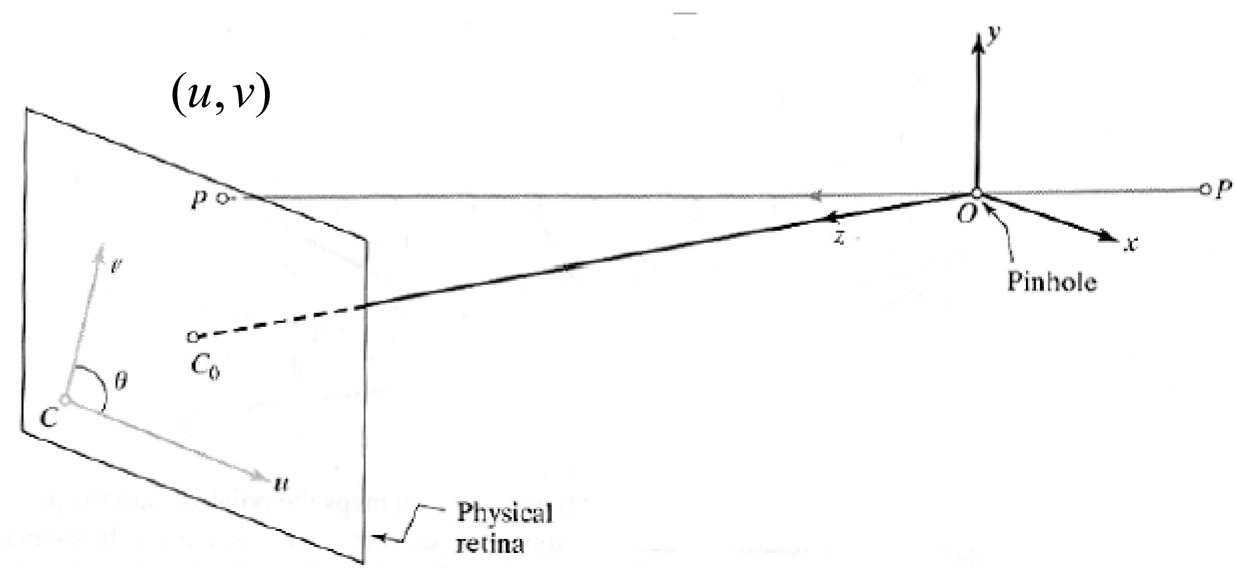




Camera parameters

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - *Skew (non-rectangular pixels)*
  - *Radial distortion*
- Extrinsic parameters
  - Rotation and translation relative to world coordinate system

Intrinsic Camera Parameters



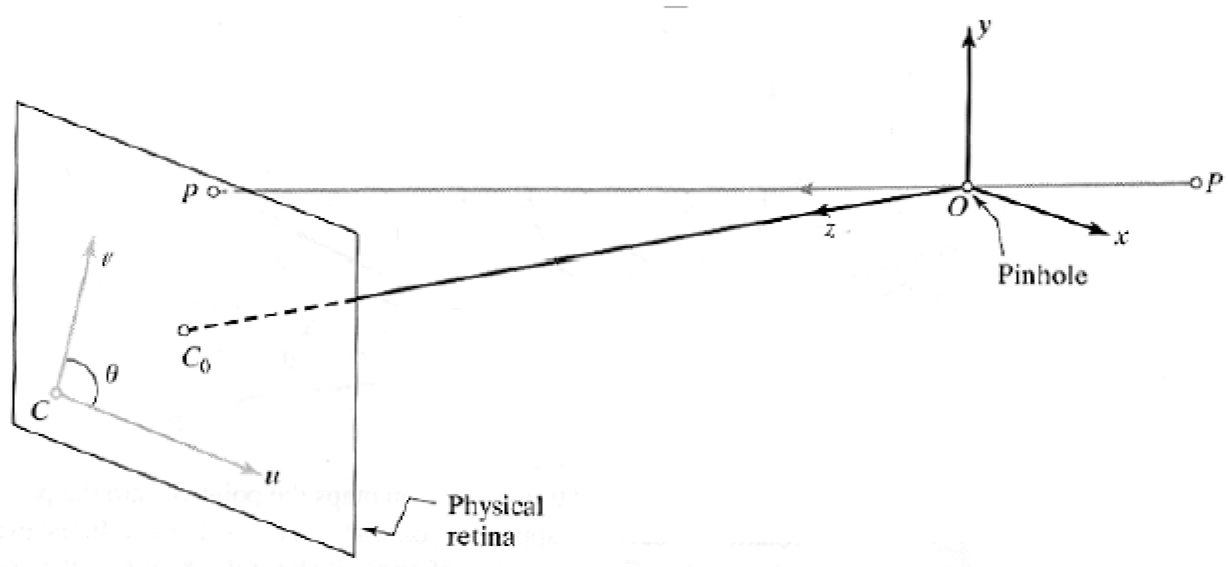
$$P = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Perspective projection

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

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

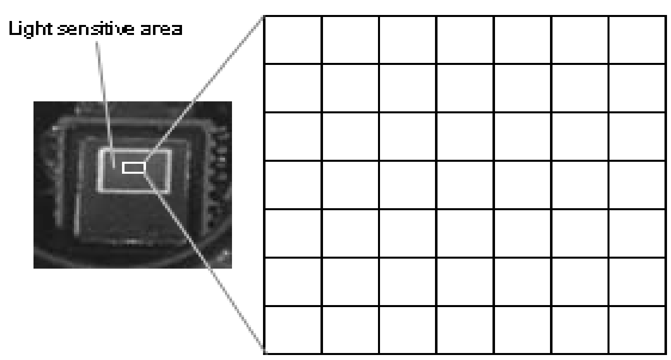
Intrinsic Camera Parameters



We need take into account the dimensions of the pixels.

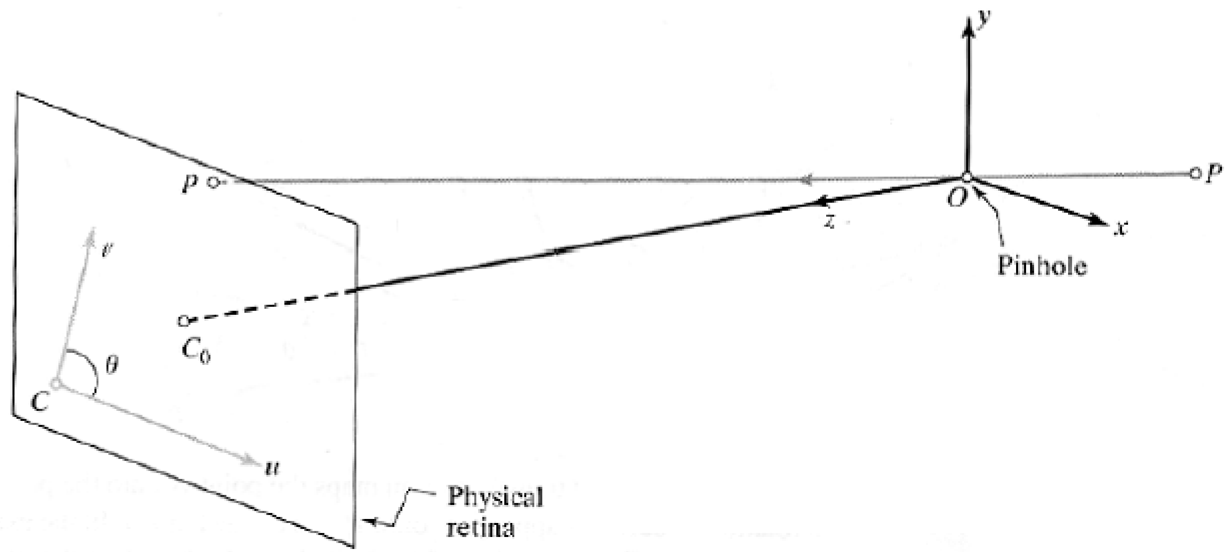
$$u = \alpha \frac{x}{z}$$

$$v = \beta \frac{y}{z}$$



CCD sensor array

Intrinsic Camera Parameters

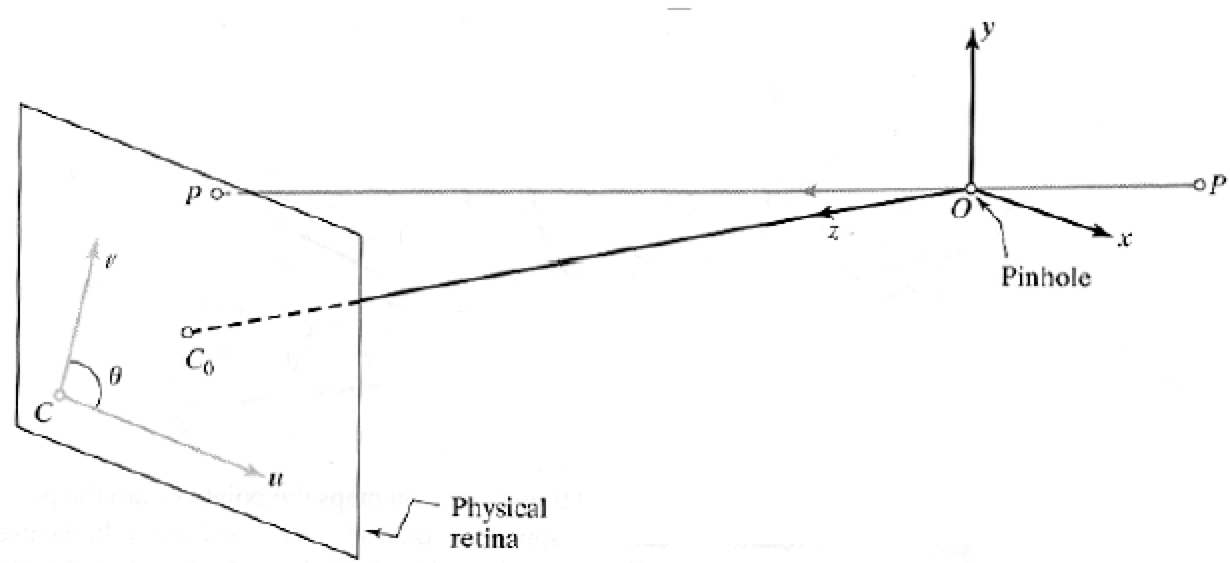


The center of the sensor chip may not coincide with the pinhole center.

$$u = \alpha \frac{x}{z} + u_0$$

$$v = \beta \frac{y}{z} + v_0$$

Intrinsic Camera Parameters

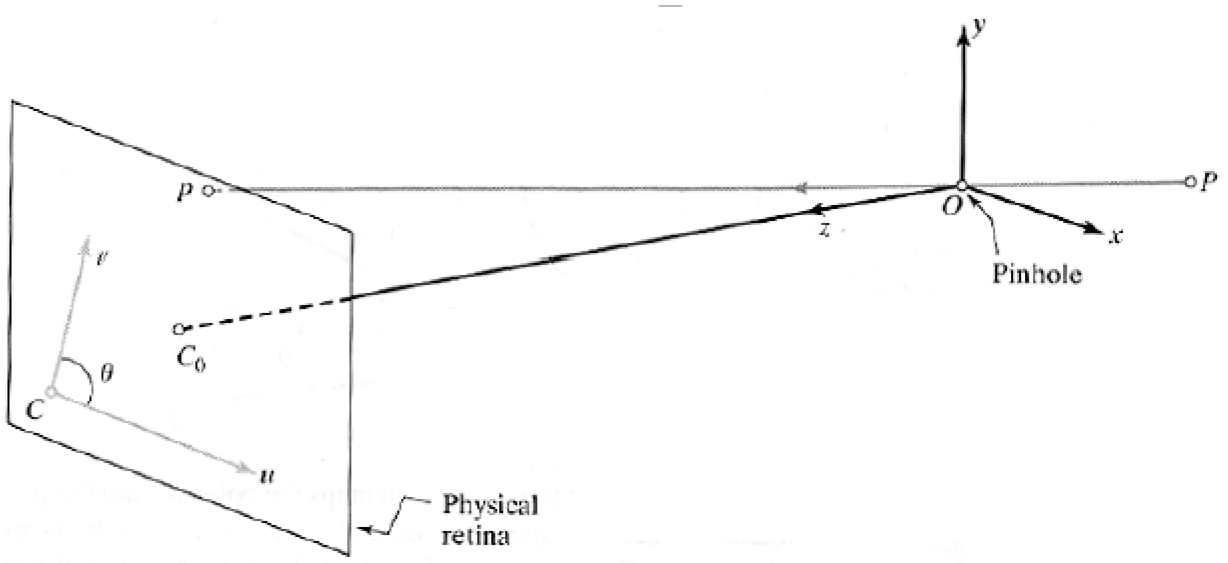


The camera coordinate system may be skewed due to some manufacturing error.

$$u = \alpha \frac{x}{z} - \alpha \cot(\theta) \frac{y}{z} + u_0$$

$$v = \frac{\beta}{\sin(\theta)} \frac{y}{z} + v_0$$

Intrinsic Camera Parameters



In homogeneous coordinates

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \begin{pmatrix} \alpha & -\alpha\cot(\theta) & u_0 & 0 \\ 0 & \dfrac{\beta}{\sin(\theta)} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

These five parameters are known as *intrinsic parameters*

Intrinsic Camera Parameters

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \begin{pmatrix} \alpha & -\alpha\cot(\theta) & u_0 & 0 \\ 0 & \dfrac{\beta}{\sin(\theta)} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

In a simpler notation: $\mathbf{p} = \dfrac{1}{z}\begin{pmatrix} K & \mathbf{0} \end{pmatrix}\mathbf{P}$

With respect to the camera coordinate system

Extrinsic Camera Parameters

- Translation and rotation of the camera frame with respect to the world frame

In homogeneous coordinates

$$X_{cam} = \begin{bmatrix} R & -R\widetilde{C} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \widetilde{X} \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\widetilde{C} \\ 0 & 1 \end{bmatrix} X$$

Using $x_{image} = \frac{1}{z}\begin{pmatrix} K & \mathbf{0} \end{pmatrix} X_{cam}$, we get $\boxed{\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z}\begin{pmatrix} K & \mathbf{0} \end{pmatrix} \begin{pmatrix} R & -R\tilde{C} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} W_X \\ \\ W_Z \\ 1 \end{pmatrix}}$

Combine Intrinsic & Extrinsic Parameters

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \begin{pmatrix} K & \mathbf{0} \end{pmatrix} \begin{pmatrix} R & -R\tilde{C} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} W_X \\ W_Y \\ W_Z \\ 1 \end{pmatrix}$$

We can further simplify to

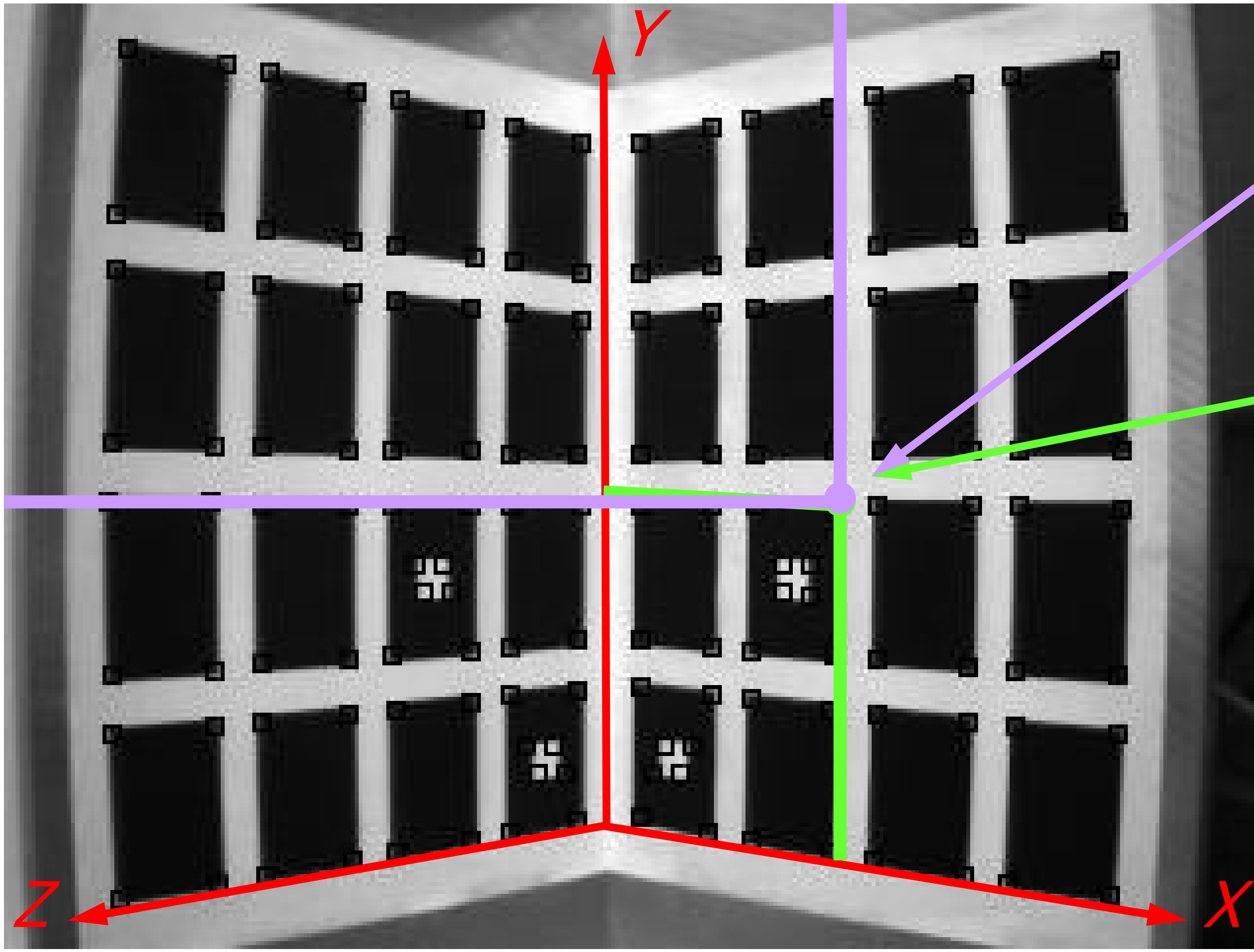
$$\boxed{\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \begin{pmatrix} K \end{pmatrix} \begin{pmatrix} R & -R\tilde{C} \end{pmatrix} \begin{pmatrix} W_X \\ W_Y \\ W_Z \\ 1 \end{pmatrix}}$$

3x4 matrix with 11 degrees of freedom: **5 intrinsic, 3 rotation, and 3 translation parameters.**

Camera Calibration

- Camera's intrinsic and extrinsic parameters are found using a setup with known positions in some fixed world coordinate system.

Camera Calibration



$$\begin{bmatrix} u_i \\ v_i \end{bmatrix}$$

$$\begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \end{bmatrix}$$

courtesy of B. Wilburn

Camera Calibration

- Mathematically, we are given *n* points

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \text{ and } \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \end{bmatrix} \text{ where } i = 1, \ldots, n$$

- We want to find **M**

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \frac{1}{z_i^c} \mathrm{M} \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix}$$

Camera Calibration

- We can write

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \frac{1}{z_i^c} \mathrm{M} \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix}$$

$$z_i^c \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix}$$

$$\begin{aligned} z_i^c u_i &= m_{11} x_i^w + m_{12} y_i^w + m_{13} z_i^w + m_{14} \\ z_i^c v_i &= m_{21} x_i^w + m_{22} y_i^w + m_{23} z_i^w + m_{24} \\ z_i^c &= m_{31} x_i^w + m_{32} y_i^w + m_{33} z_i^w + m_{34} \end{aligned}$$

Camera Calibration

- Scale and subtract last row from first and second rows

$$
\begin{aligned}
z_i^c u_i &= m_{11} x_i^w + m_{12} y_i^w + m_{13} z_i^w + m_{14} \\
z_i^c v_i &= m_{21} x_i^w + m_{22} y_i^w + m_{23} z_i^w + m_{24} \\
z_i^c &= m_{31} x_i^w + m_{32} y_i^w + m_{33} z_i^w + m_{34}
\end{aligned}
$$

to get

$$
\begin{aligned}
x_i^w m_{11} + y_i^w m_{12} + z_i^w m_{13} + m_{14} - u_i x_i^w m_{31} - u_i y_i^w m_{32} - u_i z_i^w m_{33} &= u_i m_{34} \\
x_i^w m_{21} + y_i^w m_{22} + z_i^w m_{23} + m_{24} - v_i x_i^w m_{31} - v_i y_i^w m_{32} - v_i z_i^w m_{33} &= v_i m_{34}
\end{aligned}
$$

Camera Calibration

- Write in matrix form for n points

$$
\begin{aligned}
x_i^w m_{11} + y_i^w m_{12} + z_i^w m_{13} + m_{14} - u_i x_i^w m_{31} - u_i y_i^w m_{32} - u_i z_i^w m_{33} &= u_i m_{34} \\
x_i^w m_{21} + y_i^w m_{22} + z_i^w m_{23} + m_{24} - v_i x_i^w m_{31} - v_i y_i^w m_{32} - v_i z_i^w m_{33} &= v_i m_{34}
\end{aligned}
$$

to get

$$
\begin{bmatrix}
x_1^w & y_1^w & z_1^w & 1 & 0 & 0 & 0 & 0 & -u_1 x_1^w & -u_1 y_1^w & -u_1 z_1^w \\
0 & 0 & 0 & 0 & x_1^w & y_1^w & z_1^w & 1 & -v_1 x_1^w & -v_1 y_1^w & -v_1 z_1^w \\
\vdots & & & \vdots & & & \vdots & & & \vdots & \vdots \\
x_n^w & y_n^w & z_n^w & 1 & 0 & 0 & 0 & 0 & -u_n x_n^w & -u_n y_n^w & -u_n z_n^w \\
0 & 0 & 0 & 0 & x_n^w & y_n^w & z_n^w & 1 & -v_n x_n^w & -v_n y_n^w & -v_n z_n^w
\end{bmatrix}
\begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ \vdots \\ m_{32} \\ m_{33} \end{bmatrix}
=
\begin{bmatrix} u_1 m_{34} \\ v_1 m_{34} \\ u_2 m_{34} \\ v_2 m_{34} \\ \vdots \\ u_n m_{34} \\ v_n m_{34} \end{bmatrix}
$$

$\Rightarrow$ $\mathbf{Km} = \mathbf{U}$

Let m34=1; that is, scale the projection matrix by m34.

Camera Calibration

- The least square solution of $\mathbf{K}\mathbf{m} = \mathbf{U}$ is

$$\mathbf{m} = (\mathbf{K}^{\mathrm{T}}\mathbf{K})^{-1}\mathbf{K}^{T}\mathbf{U}$$

- From the matrix **M**, we can find the intrinsic and extrinsic parameters.

Camera Calibration

- Consider the case where skew angle is 90. Since we set *m34=1*, we need to take that into account at the end.

$$m_{34}\begin{bmatrix}\mathbf{m}_1^T & m_{14}\\ \mathbf{m}_2^T & m_{24}\\ \mathbf{m}_3^T & m_{34}\end{bmatrix} = \begin{bmatrix}\alpha & 0 & u_0 & 0\\ 0 & \beta & v_0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}\mathbf{r}_1^T & t_x\\ \mathbf{r}_2^T & t_y\\ \mathbf{r}_3^T & t_z\\ 0^T & 1\end{bmatrix} = \begin{bmatrix}\alpha\mathbf{r}_1^T + u_0\mathbf{r}_3^T & \alpha t_x + u_0 t_x\\ \beta\mathbf{r}_2^T + v_0\mathbf{r}_3^T & \beta t_y + v_0 t_x\\ \mathbf{r}_3^T & t_x\end{bmatrix}$$

Notice that $m_{34}\mathbf{m}_3 = \mathbf{r}_3$

Since **R** is a rotation matrix, $|\mathbf{r}_i| = 1$

Therefore, $m_{34} = \dfrac{1}{|\mathbf{m}_3|}$

Camera Calibration

- We get

$$
\begin{aligned}
\mathbf{r}_3 &= m_{34}\mathbf{m}_3 \\
u_0 &= (\alpha\mathbf{r}_1^T + u_0\mathbf{r}_3^T)\mathbf{r}_3 = m_{34}^2\mathbf{m}_1^T\mathbf{m}_3 \\
v_0 &= (\beta\mathbf{r}_2^T + v_0\mathbf{r}_3^T)\mathbf{r}_3 = m_{34}^2\mathbf{m}_2^T\mathbf{m}_3 \\
\alpha &= m_{34}^2|\mathbf{m}_1 \times \mathbf{m}_3| \\
\beta &= m_{34}^2|\mathbf{m}_2 \times \mathbf{m}_3|
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{r}_1 &= \frac{m_{34}}{\alpha}(\mathbf{m}_1 - u_0\mathbf{m}_3) \\
\mathbf{r}_2 &= \frac{m_{34}}{\beta}(\mathbf{m}_2 - v_0\mathbf{m}_3) \\
t_z &= m_{34} \\
t_x &= \frac{m_{34}}{\alpha}(m_{14} - u_0) \\
t_y &= \frac{m_{34}}{\beta}(m_{24} - v_0)
\end{aligned}
$$

See Forsyth & Ponce for details and skew-angle case.

See http://www.vision.caltech.edu/bouguetj/calib_doc/ for more useful toolbox

References

[1] Computer Vision: A Modern Approach, Forsyth, David A., and Ponce, Prentice Hall, 2003.

[2] CV Course slide, Rob Fergus, New York Unv., http://cs.nyu.edu/~fergus/teaching/vision/index.html

[3] http://www.vision.caltech.edu/bouguetj/calib_doc

[4] http://www.ece.lsu.edu/gunturk/EE4780/EE4780.html