دانشگاه
خواجه نصیرالدین طوسی
K. N. Toosi University of Technology

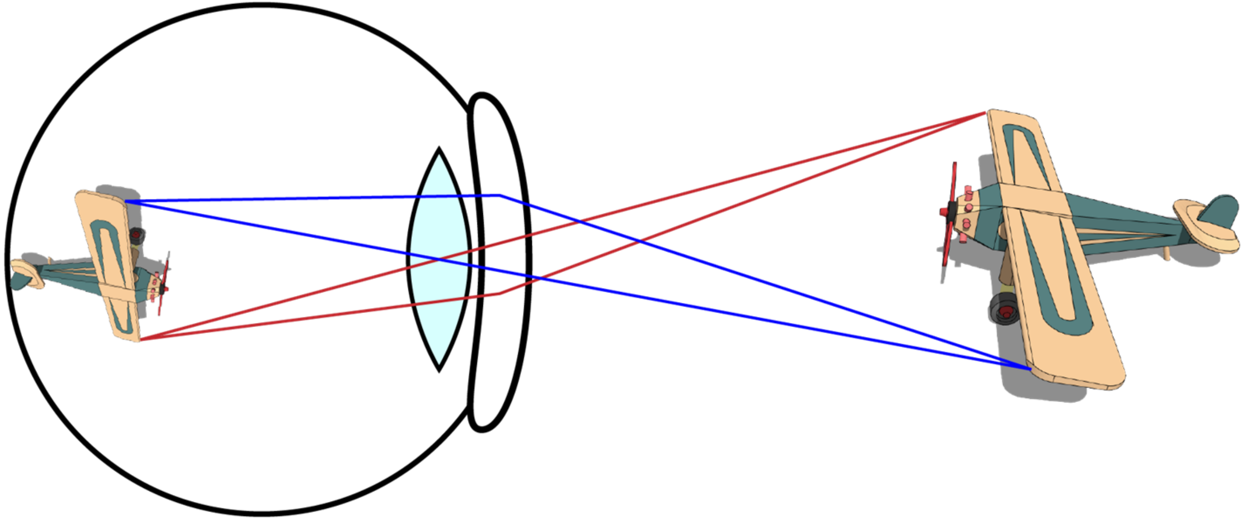
Computer Vision

Lecture 3: 3D-2D Coordinates Transform

Dr. Esmaeil Najafi

MSc. Javad Khoramdel

Eyes: projection onto retina



Model: pinhole camera

- For convenience (to avoid an inverted image) we treat the image plane as if it were in front of the pinhole (i.e. the virtual image).

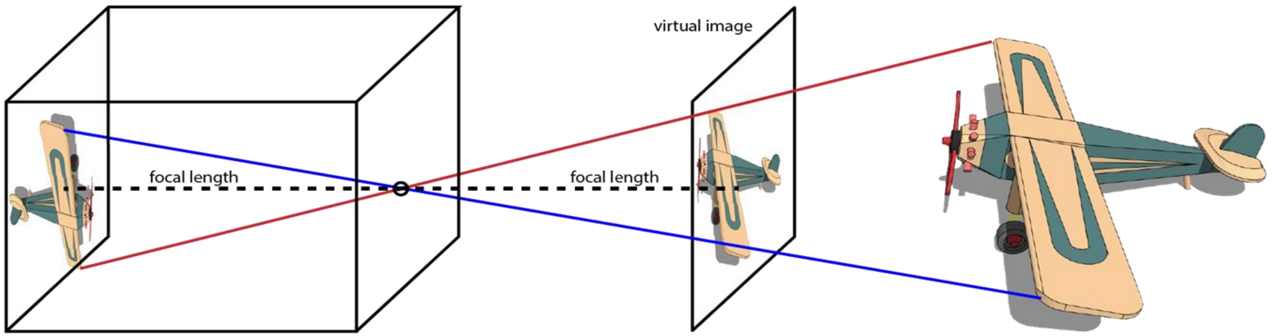
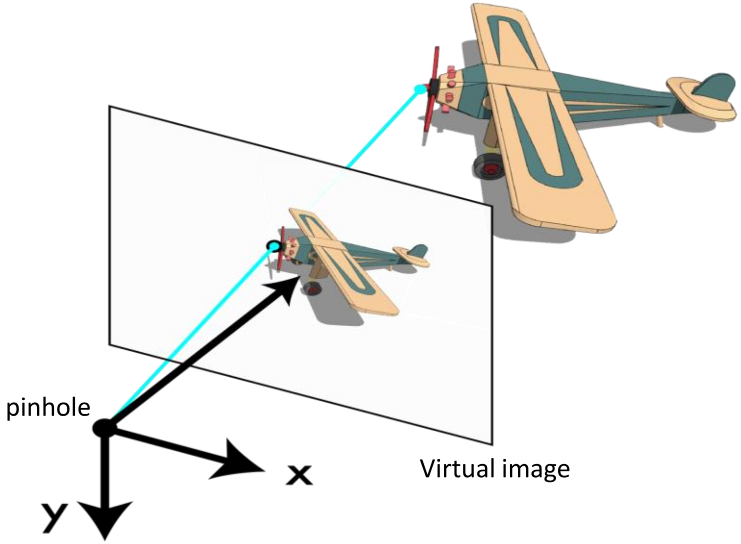
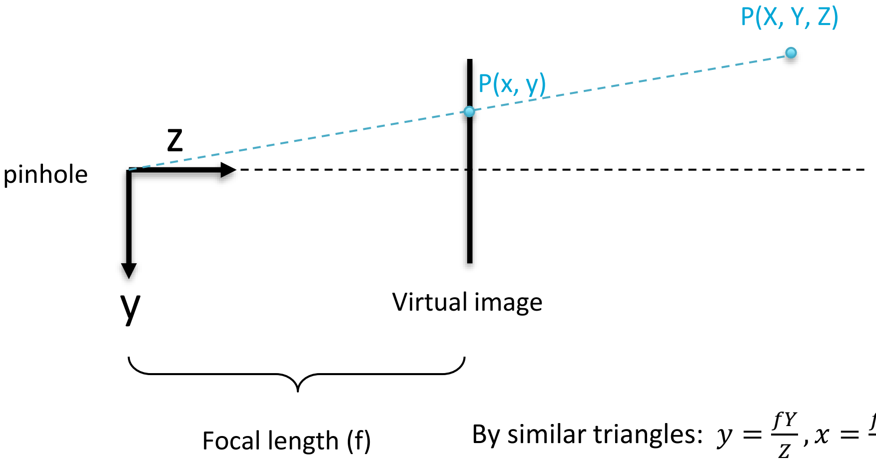


Image: 3d -> 2d projection of the world

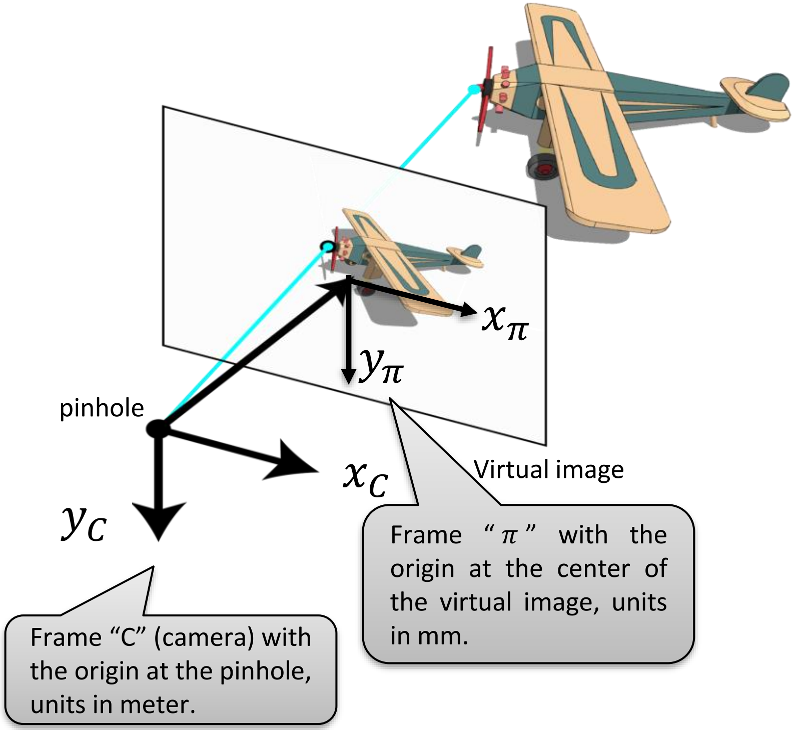


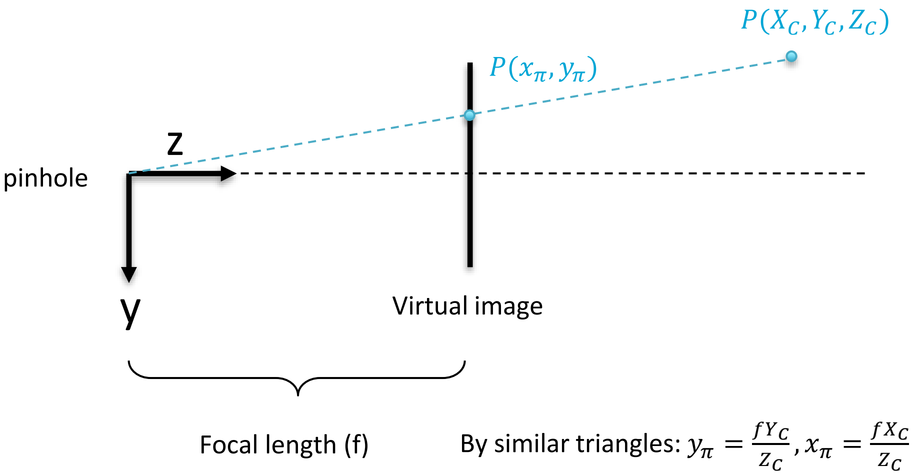


By similar triangles: $y = \frac{fY}{Z}, x = \frac{fX}{Z}$

But, what are $(x, y), (X, Y, Z)$?
What do they represent?

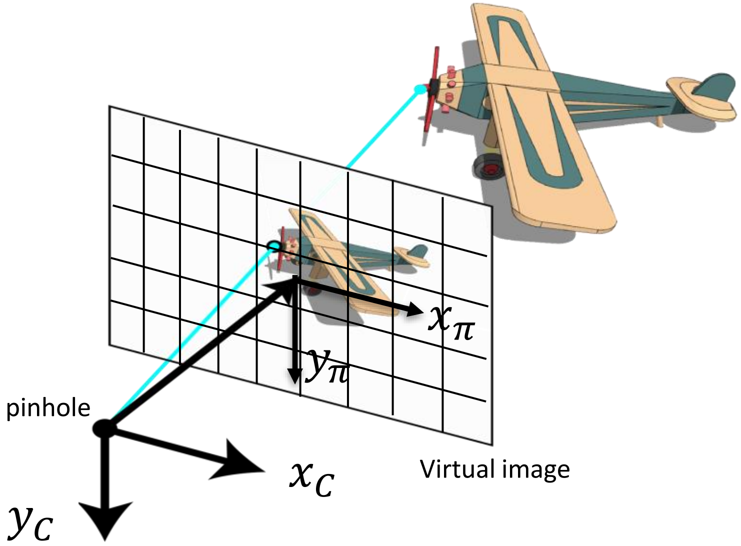
Image: 3d -> 2d projection of the world





By similar triangles: $y_\pi = \frac{fY_C}{Z_C}, x_\pi = \frac{fX_C}{Z_C}$

Digital image is a matrix

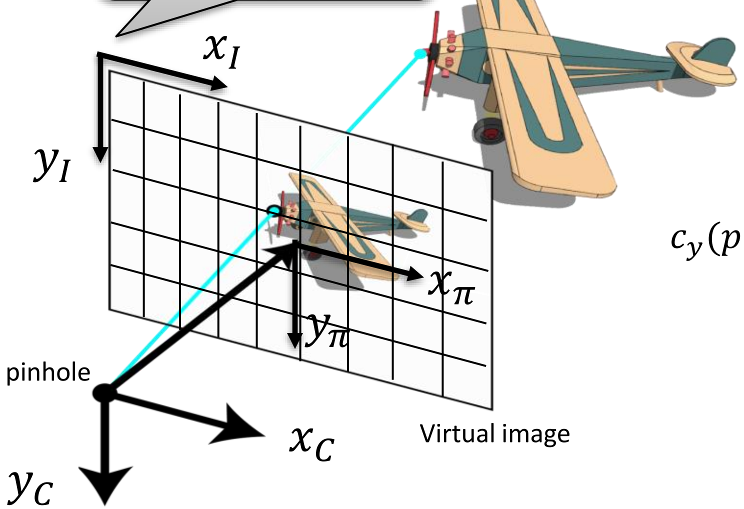


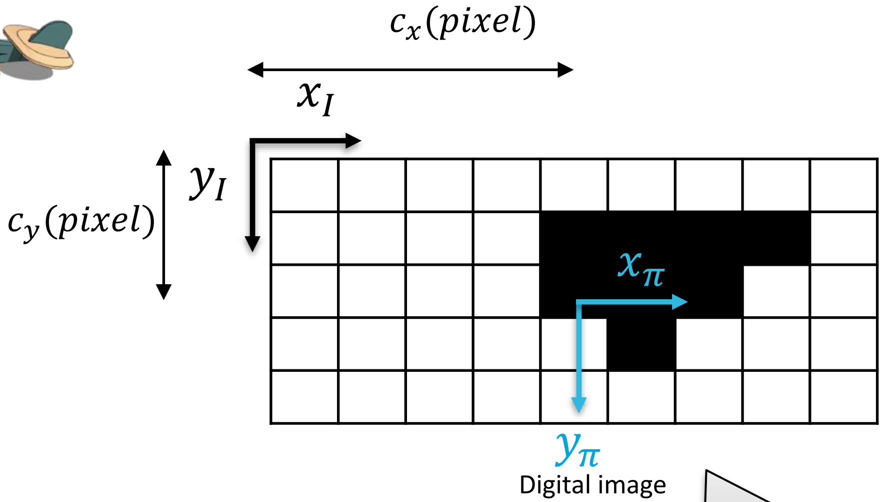
255	255	255	255	255	255	255	255	255
255	255	255	255	0	0	0	0	255
255	255	255	255	0	0	0	255	255
255	255	255	255	255	0	255	255	255
255	255	255	255	255	255	255	255	255

Digital image

Digital image is a matrix

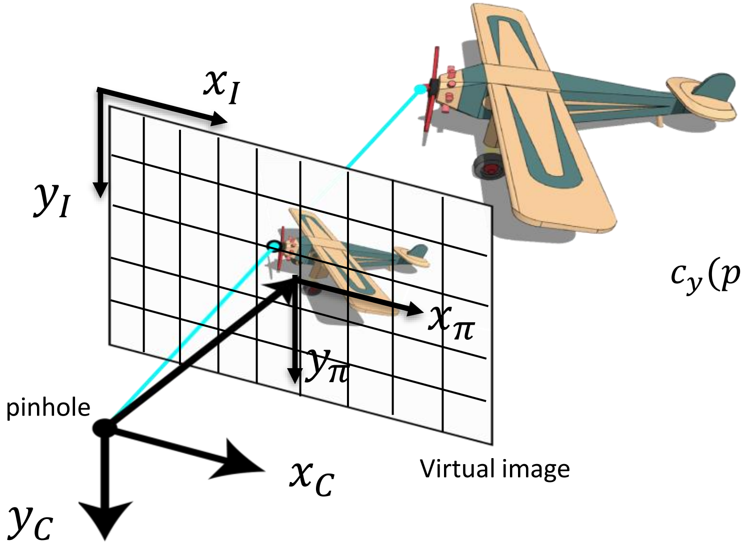
Frame “I” (image) with the origin at the top left corner of the image, units in pixels.

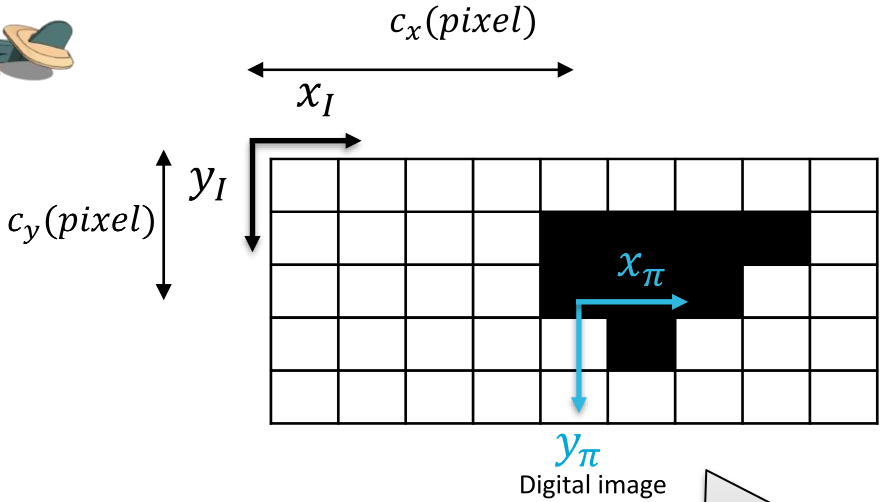




Considering the units for each frames, what are the coordinates of the origin of the frame "π“ in frame “I”?

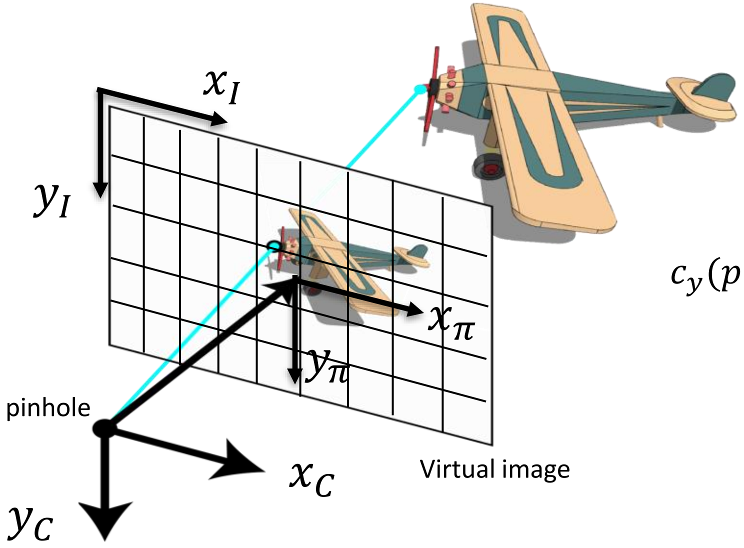
Digital image is a matrix

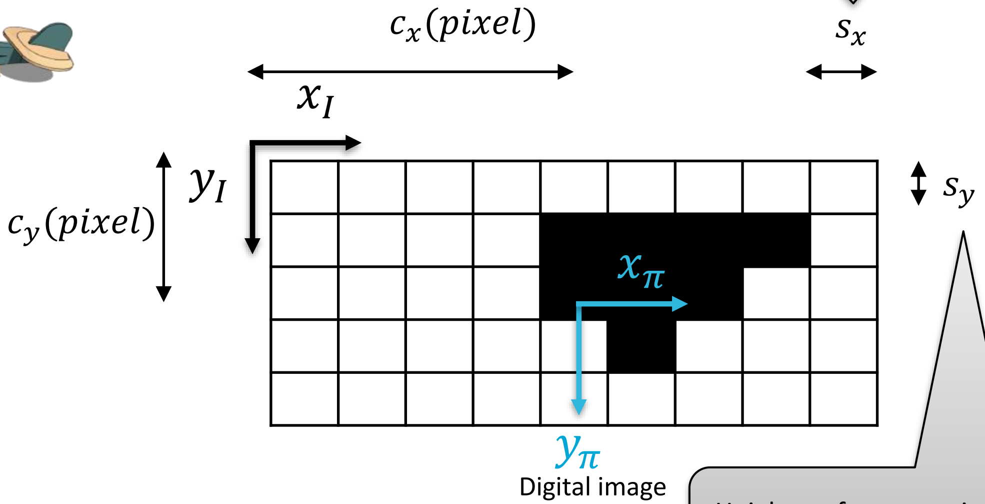




Considering the units for each frames, what are the coordinates of the origin of the frame "I" in frame "π"?

Conversion between virtual image and digital image





Width of one pixel in millimeters, unit (mm/pixel)

Height of one pixel in millimeters, unit (mm/pixel)

Frame "I" to frame "π"

$$x_\pi = (x_I - c_x)s_x$$

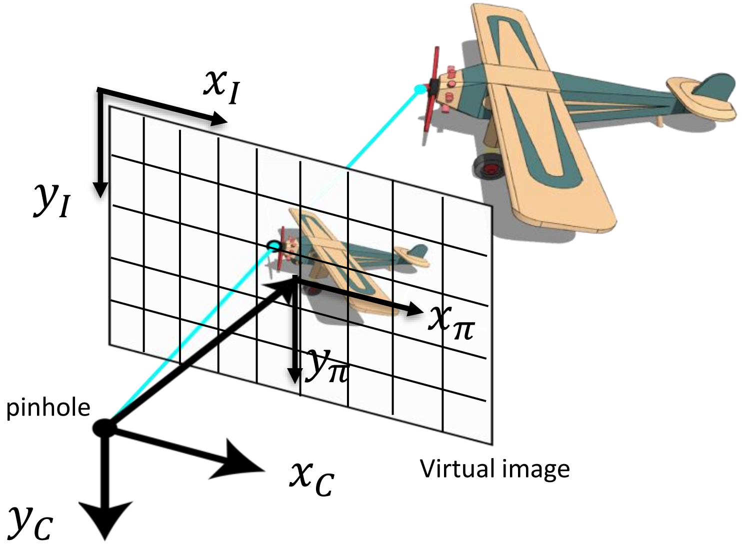
$$y_\pi = (y_I - c_y)s_y$$

$$x_I = \frac{x_\pi}{s_x} + c_x$$

$$y_I = \frac{y_\pi}{s_y} + c_y$$

Frame "π" to frame "I"

From camera frame to digital image frame



$$x_I = \frac{x_\pi}{s_x} + c_x$$

$$y_I = \frac{y_\pi}{s_y} + c_y$$

$$x_\pi = \frac{fX_C}{Z_C}$$

$$y_\pi = \frac{fY_C}{Z_C}$$

$$x_I = \frac{fX_C}{s_x Z_C} + c_x$$

$$y_I = \frac{fY_C}{s_y Z_C} + c_y$$

All we really need is :

$$f_x = \frac{f}{s_x}, f_y = \frac{f}{s_y}$$

We don’t need to know the actual values of f, s_x and s_y; just their ratios.

Intrinsic Camera Parameters

- Camera intrinsic parameters for a pinhole camera model:
  - Focal length f and sensor element sizes s_x, s_y.
    - ✓ Or, just focal lengths in pixels f_x, f_y.
  - Optical center of the image at pixel location c_x, c_y.

All we really need is :

$$f_x = \frac{f}{s_x}, f_y = \frac{f}{s_y}$$

$$x_I = \frac{X_C}{Z_C} f_x + c_x$$

$$y_I = \frac{Y_C}{Z_C} f_y + c_y$$

We can alternatively express focal length in units of pixels.

Intrinsic camera matrix

- We can capture all the intrinsic camera parameters in a matrix **K**:

$$K = \begin{pmatrix} f/s_x & 0 & c_x \\ 0 & f/s_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

- The coordinates of a point in camera frame can be converted to the image frame by matrix multiplication:

$$X_{un-norm} = K.\begin{pmatrix} X_C \\ Y_C \\ Z_C \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} X_C \\ Y_C \\ Z_C \end{pmatrix} = \begin{pmatrix} f_x.X_C + c_x.Z_C \\ f_y.Y_C + c_y.Z_C \\ Z_C \end{pmatrix}$$

$$X_{norm} = \frac{1}{Z_C}.X_{un-norm} = \begin{pmatrix} f_x.X_C + c_x.Z_C \\ f_y.Y_C + c_y.Z_C \\ Z_C \end{pmatrix} = \begin{pmatrix} f_x.\frac{X_C}{Z_C} + c_x \\ f_y.\frac{Y_C}{Z_C} + c_y \\ 1 \end{pmatrix} = X_I$$

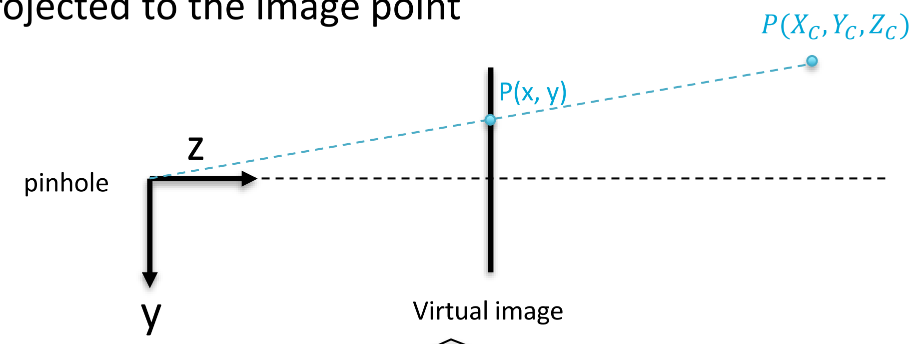
Back projection

- If you have an image point, you can "back project" that point into the scene.

- However, the resulting 3D point is not uniquely defined
  - It is actually a ray emanating from the camera center, out through the image point, to infinity
  - Any 3D point along that ray could have projected to the image point

$$\begin{pmatrix} \frac{X_C}{Z_C} \\ \frac{Y_C}{Z_C} \\ 1 \end{pmatrix} = K^{-1}X_I = K^{-1}\begin{pmatrix} x_I \\ y_I \\ 1 \end{pmatrix}$$

Z_C (distance to camera) is unknown.



Since we don't know Z_C, any point on the dotted blue ray can be the actual point.

Extrinsic camera matrix

- If 3D points are in world coordinates, we first need to transform them to camera coordinates with homogeneous transformation matrix (${}^{C}_{W}T$):

$$P_C = {}^{C}_{W}T\, P_W = \begin{pmatrix} {}^{C}_{W}R_{3\times3} & {}^{C}_{W}t_{org\,3\times1} \\ \mathbf{0}_{1\times3} & 1_{1\times1} \end{pmatrix} . P_W = \begin{pmatrix} {}^{C}_{W}R_{3\times3} & {}^{C}_{W}t_{org\,3\times1} \\ \mathbf{0}_{1\times3} & 1_{1\times1} \end{pmatrix} . \begin{pmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{pmatrix} = \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix}$$

- We can write this as an extrinsic camera matrix (M_{ext}):

$$P_C = M_{ext}\, P_W = \begin{pmatrix} {}^{C}_{W}R_{3\times3} & {}^{C}_{W}t_{org\,3\times1} \end{pmatrix} . P_W = \begin{pmatrix} {}^{C}_{W}R_{3\times3} & {}^{C}_{W}t_{org\,3\times1} \end{pmatrix} . \begin{pmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{pmatrix} = \begin{pmatrix} X_C \\ Y_C \\ Z_C \end{pmatrix}$$