I: Image Formation

3D CV

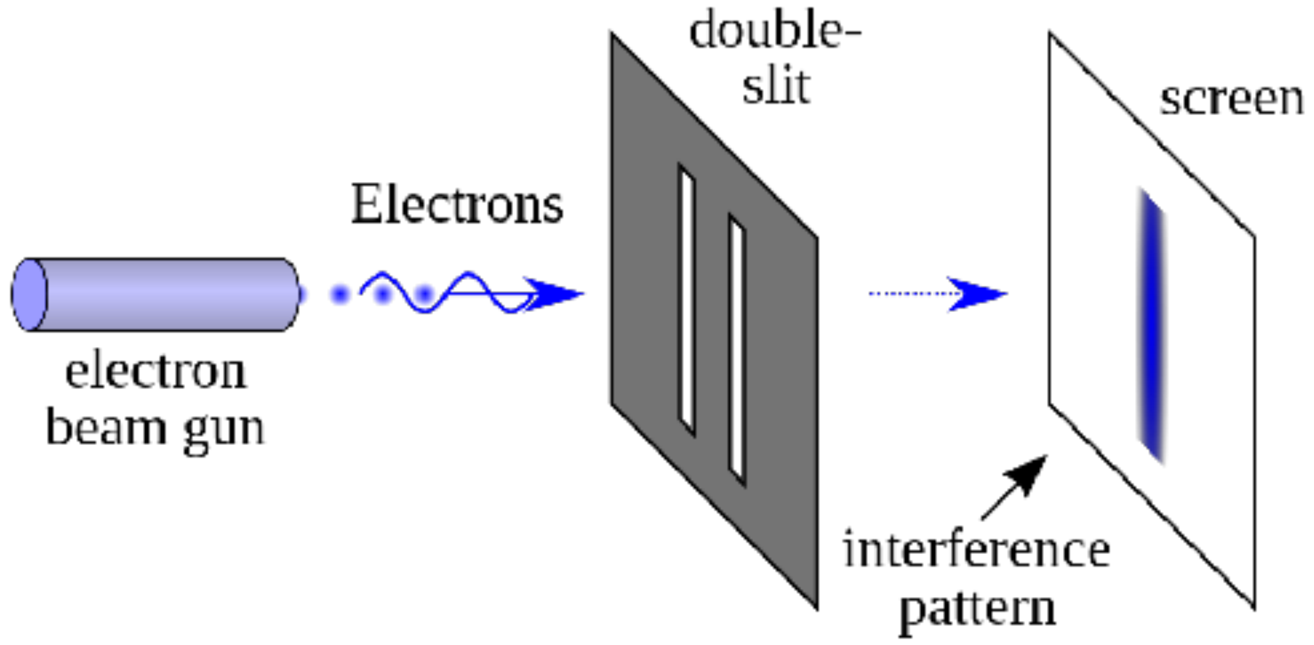
Kirill Struminsky

Course Structure

- Lecture 1+2: multi-view geometry fundamentals (+ assignment)
- Lecture 3: from 2D to 2.5D: depth estimation
- Lecture 4: point cloud processing (+ assignment)
- Lectures 5-8: differentiable rendering for reconstruction and generation
  - Rasterization
  - Implicit representations and neural fields (+ assignment)

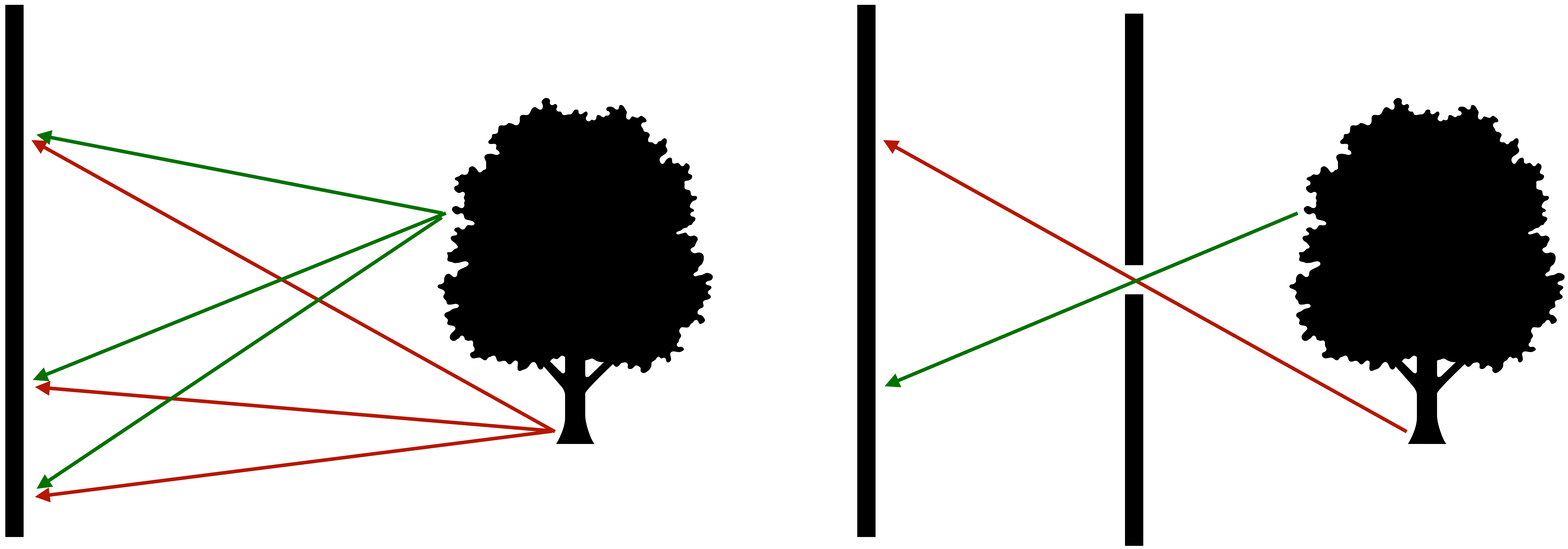
Image Formation

Light

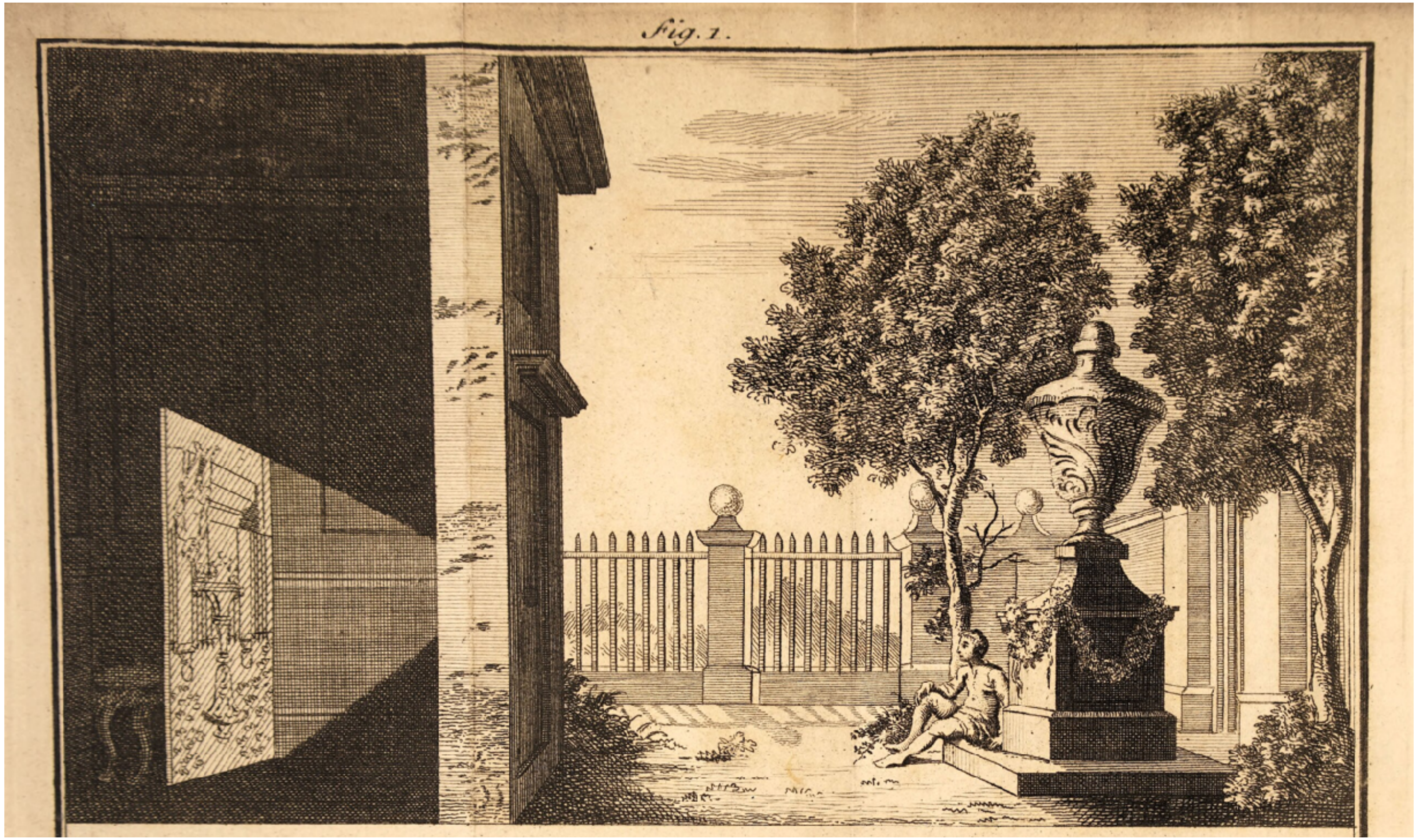


https://en.wikipedia.org/wiki/Double-slit_experiment
https://ru.wikipedia.org/wiki/The_Dark_Side_of_the_Moon
https://www.youtube.com/watch?v=frLwRLS_ZR0

Simplest Camera



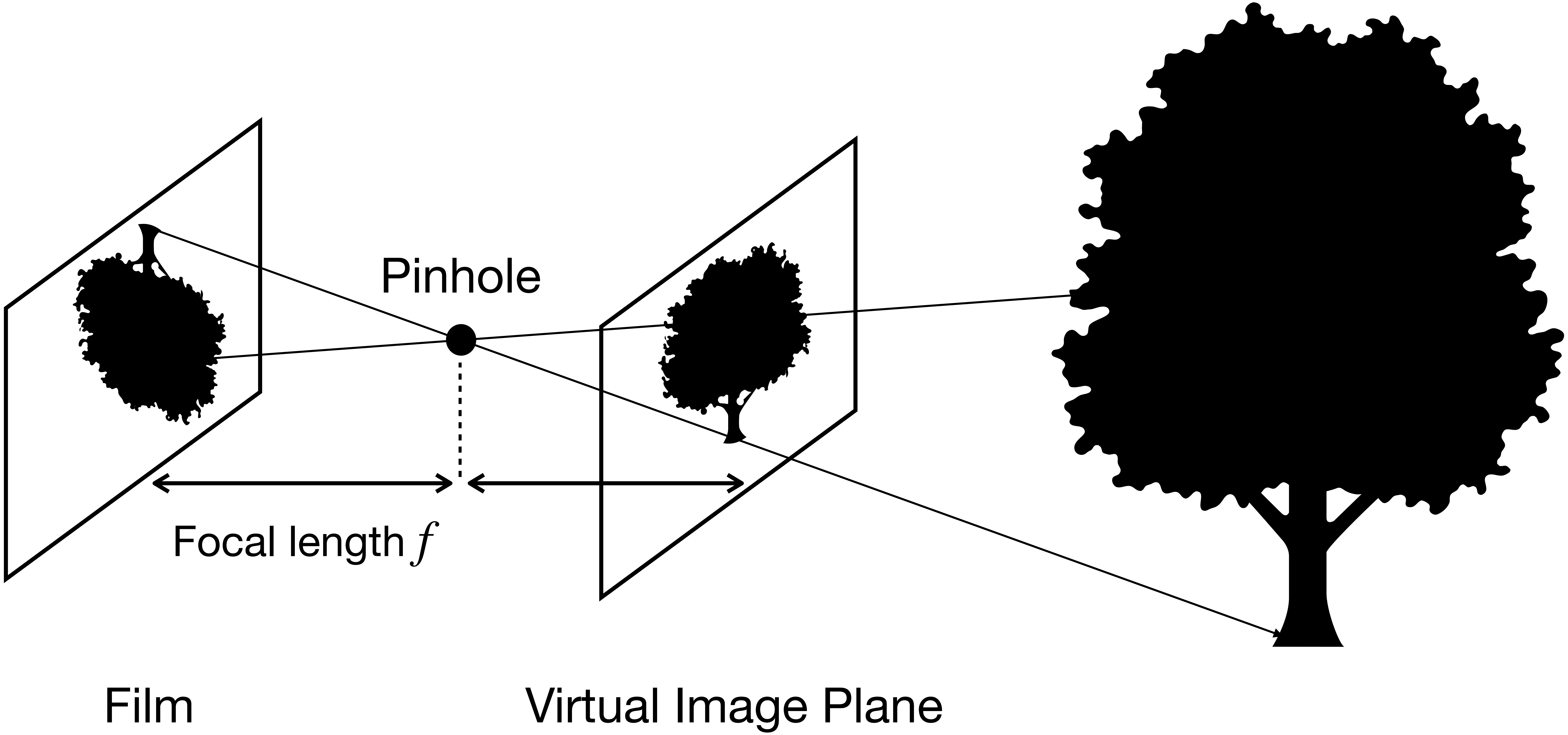
Camera Obscura and Pinhole Camera



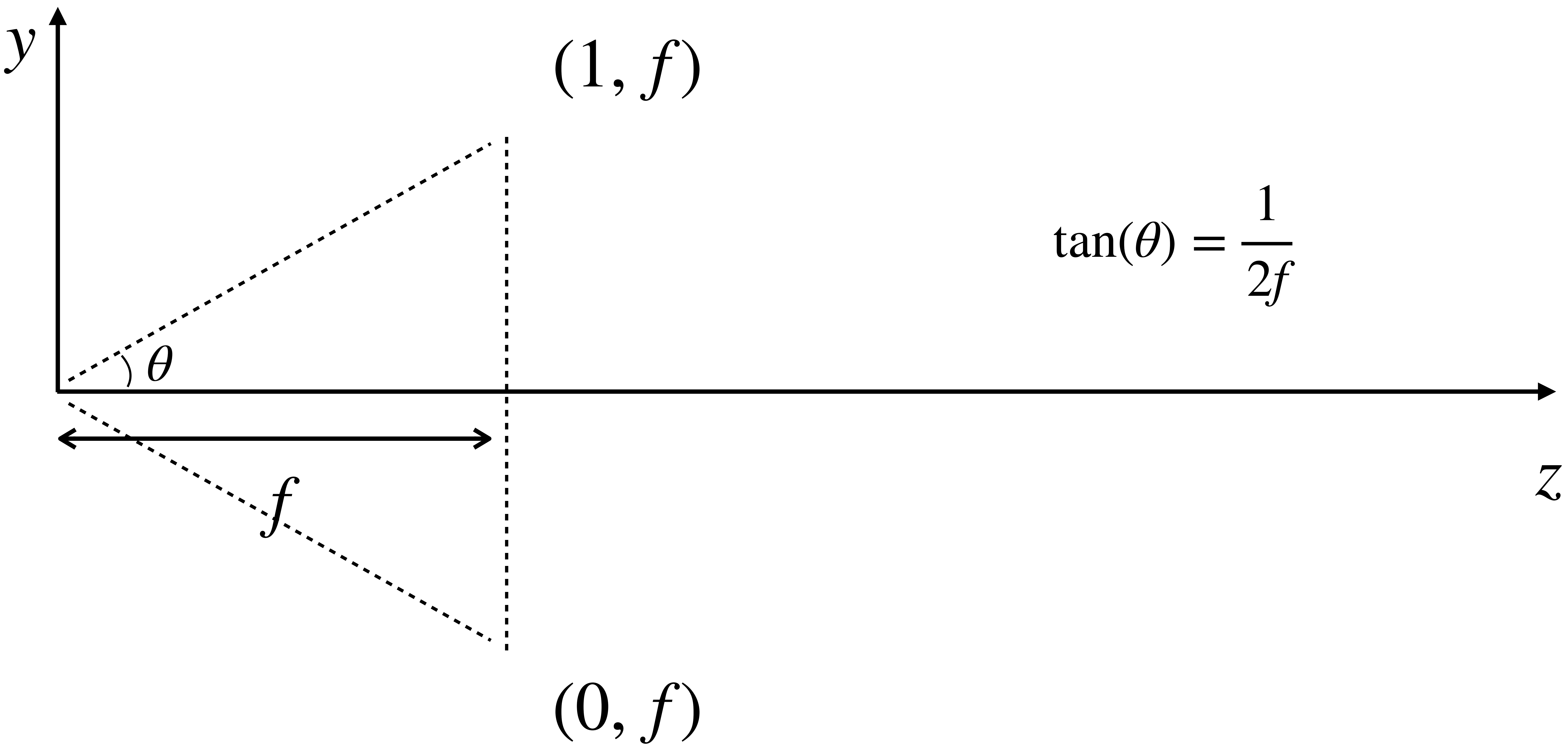
Natural Camera Obscura

https://en.wikipedia.org/wiki/Camera_obscura
https://www.reddit.com/r/pics/comments/3gplnx/sunlight_through_a_tree_during_an_eclipse/
see also https://www.abelardomorell.net/selectedworks/camera-obscura

Perspective Projection

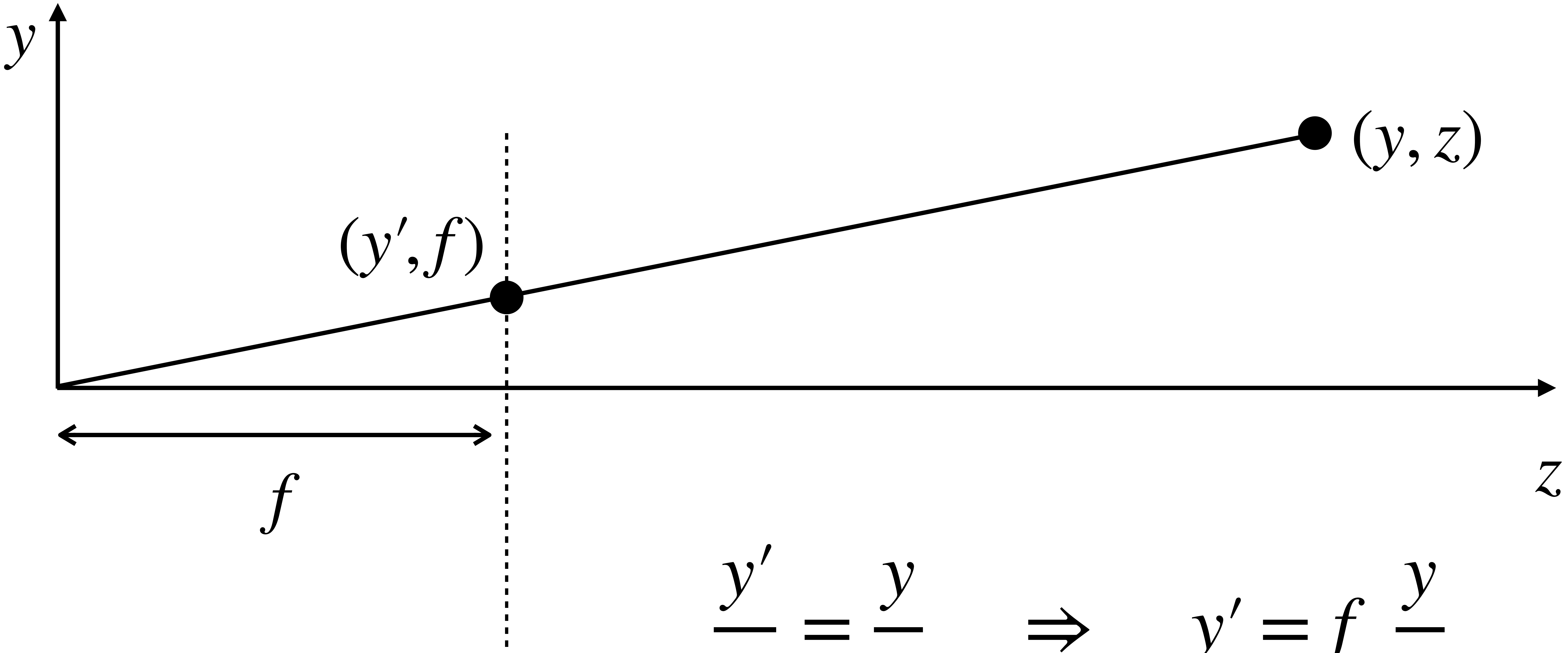


Field of View vs Focal Length



$$\tan(\theta) = \frac{1}{2f}$$

Perspective Projection



$$\frac{y'}{f} = \frac{y}{z} \quad \Rightarrow \quad y' = f\,\frac{y}{z}$$

Homogeneous Coordinates

Perspective Projection as a Linear Map

Inhomogeneous coordinates:

Projection is a non-linear map: $(x, y, z) \rightarrow \left(f\frac{x}{z}, f\frac{y}{z}\right)$

Homogeneous coordinates:

Take $X = (x, y, z)$ and introduce an additional coordinate (homogeneous coordinates):

$$\tilde{X} = (x, y, z, 1)$$

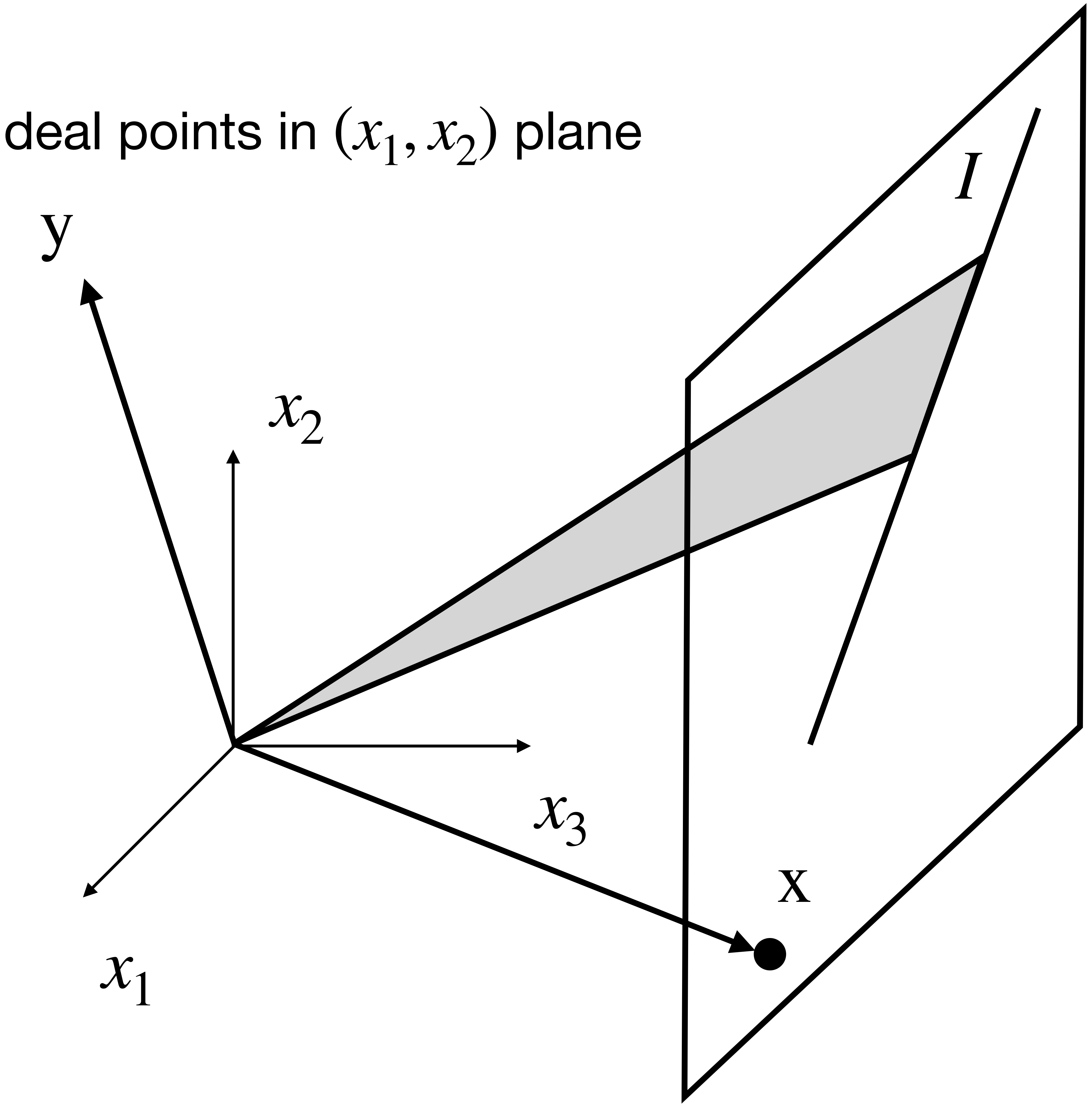
We say $\tilde{Y} = \tilde{X}$ if $\exists \lambda \in \mathbb{R}$ s.t. $\tilde{Y} = (\lambda x, \lambda y, \lambda z, \lambda)$

Projection turns into $(x, y, z, w) \rightarrow (x, y, \frac{z}{f})$

A Glimpse into Projective Geometry

A model of the Projective Plane

Ideal points in (x_1, x_2) plane

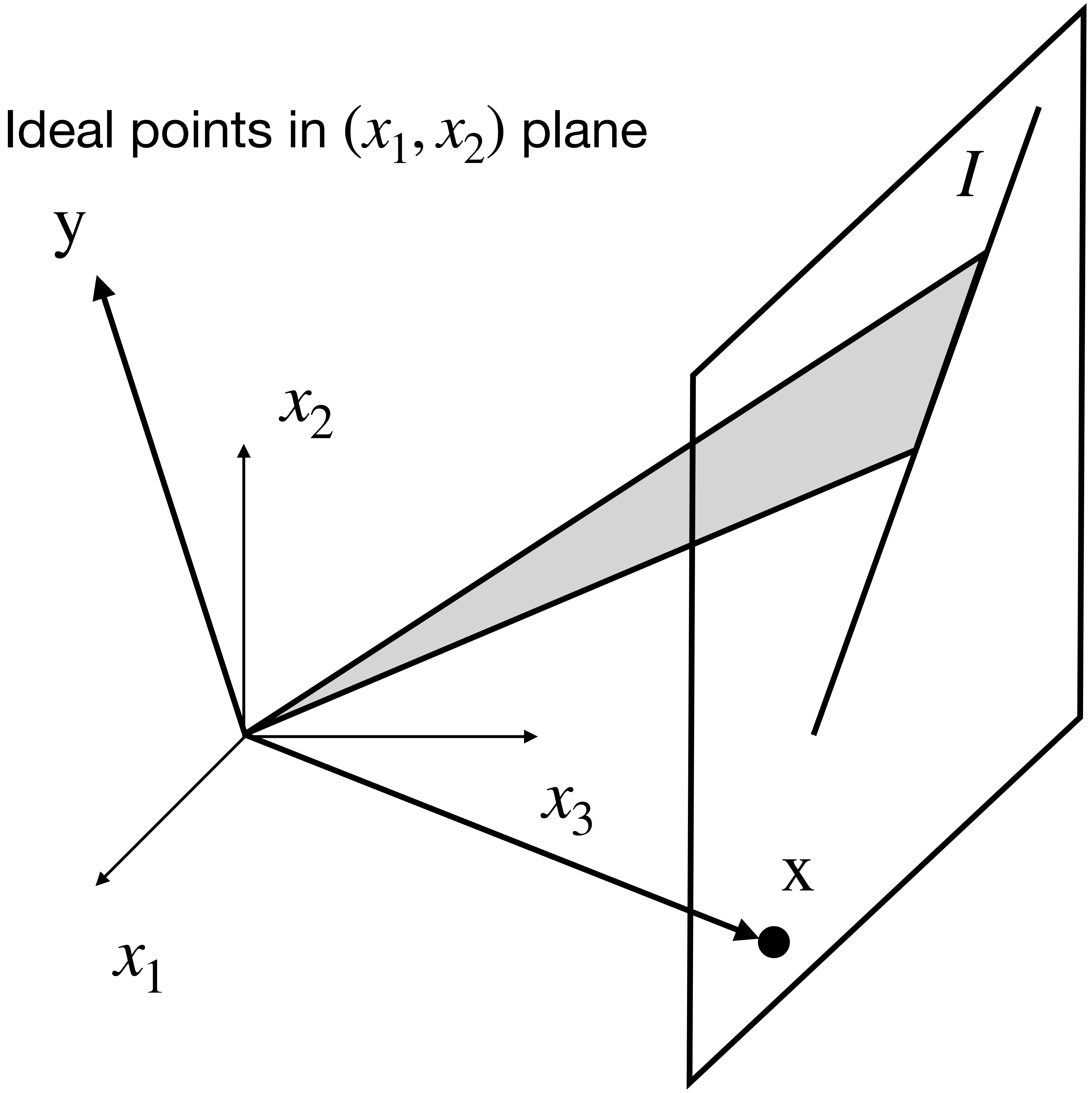


$\mathrm{I} \in \mathbb{R}^3 \backslash \{0\}$ *defines a line*

- *Point* x *lies on a line iff* $\mathrm{x} \cdot \mathrm{I} = 0$
- *Lines* I_1 *and* I_2 *intersect at* $\mathrm{I}_1 \times \mathrm{I}_2$
- *Line* $\mathrm{x} \times \mathrm{y}$ *goes through points* x *and* y

A Glimpse into Projective Geometry

A model of the Projective Plane



Projective space $\mathbb{P}^n$ is the set of equivalence classes of $\mathbb{R}^{n+1} \backslash \{0\}$ under relation $\sim$, where $x \sim y$ if $\exists \lambda : x = \lambda y$

Linear map $\mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$ induce transformations of projective space

Example: simple projection

$$H = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Projective Transformation Properties

Transformation:	Invariant:
Translation	Orientation
Rigid Movement	Lengths
Similarity	Angles
Affine Transformation	Parallelism
Projective Transformation	Collineation

Let $\mathrm{x}_1, \mathrm{x}_2$ and x_3 lie on a line I. Let H be a non-singular 3×3 matrix.

Then points Hx_i lie on the line $\mathrm{H}^{-T}\mathrm{I}$

Indeed, $\mathrm{H}^{-T}\mathrm{I} \cdot \mathrm{H}x_i = \mathrm{I}^T\mathrm{H}^{-1}\mathrm{H}x_i = 0$

https://www.youtube.com/watch?v=e4qsX4XDtYY

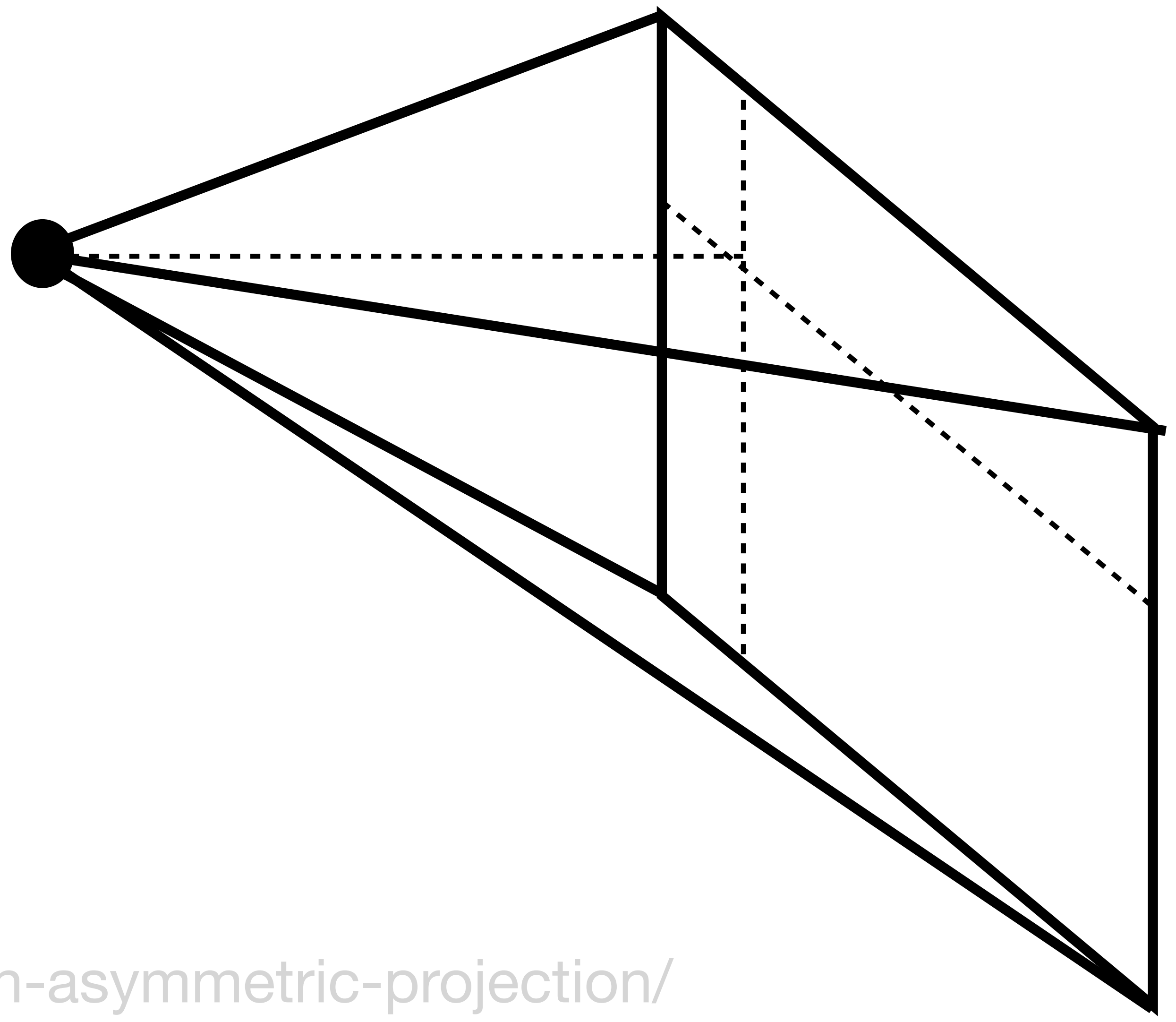
https://www.youtube.com/watch?v=QWMFpxkGO_s

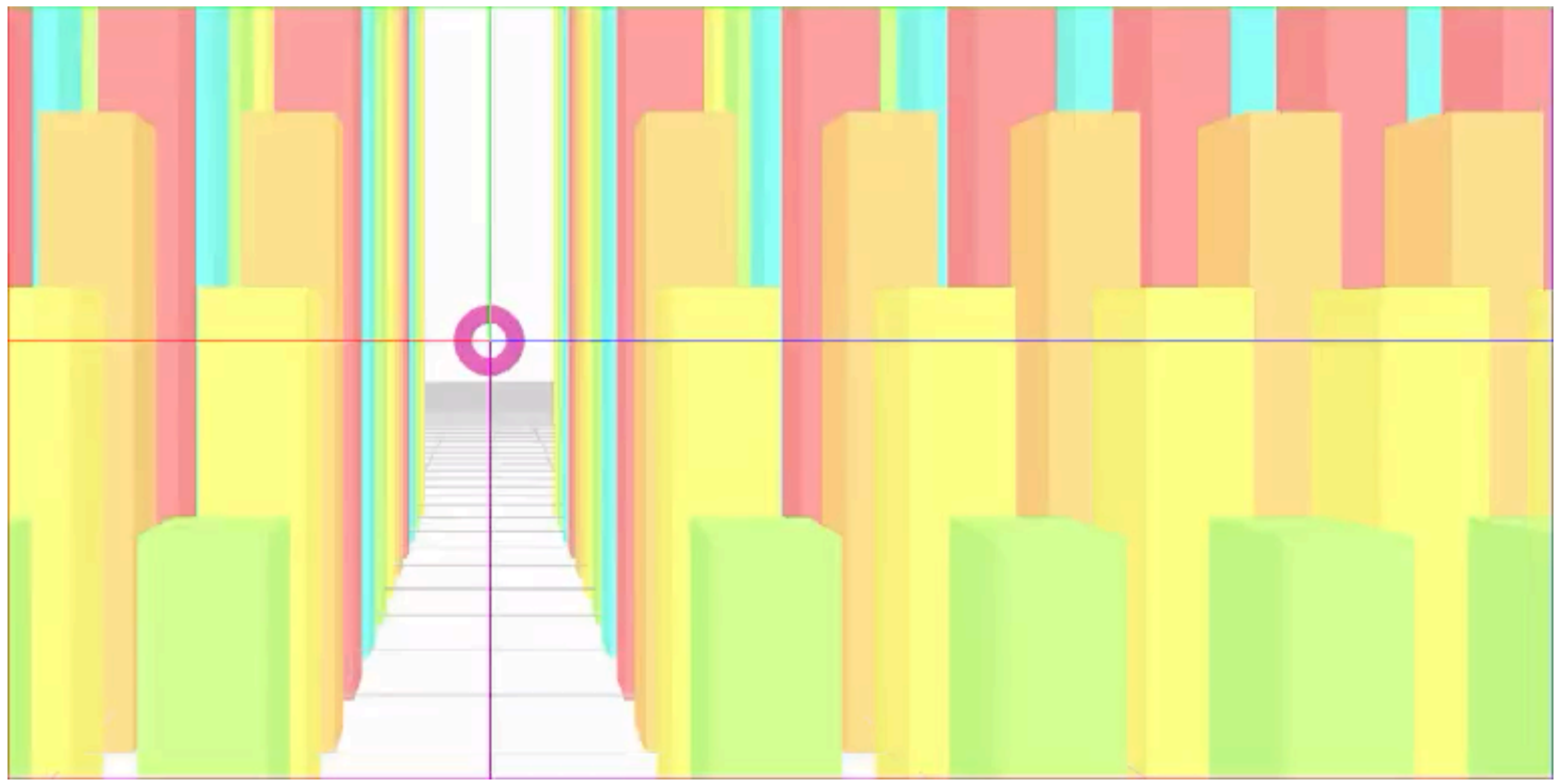
Camera Intrinsics / Projection Matrix

- There is one more component to the standard projection matrix

$$H = \begin{pmatrix} f_x & 0 & p_x & 0 \\ 0 & f_y & p_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = K\ [I \mid 0]$$

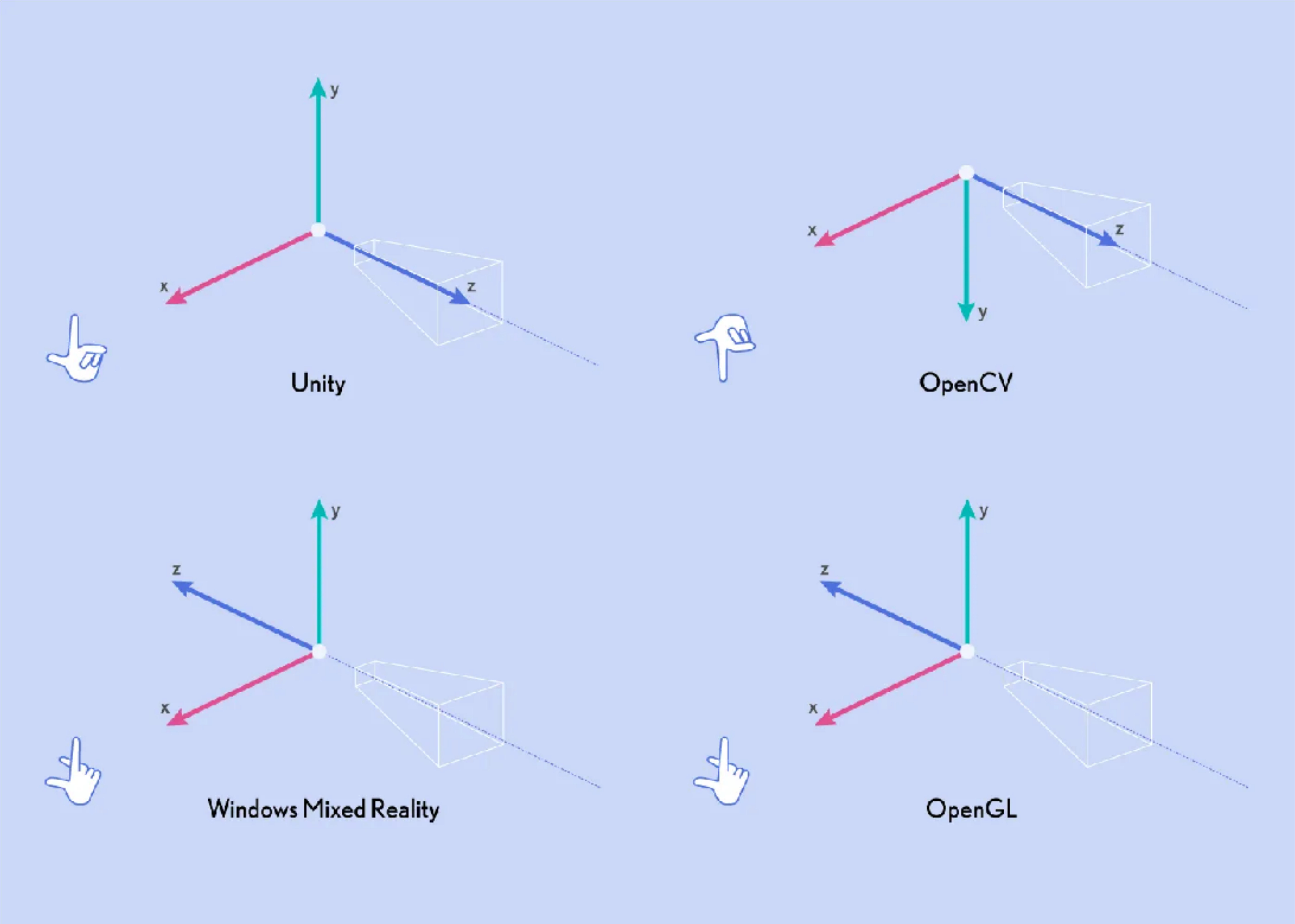
- K represents *intrinsic* parameters of camera (“intrinsic calibration” = estimation of K)





https://jsantell.com/portals-with-asymmetric-projection/

A Note on Conventions



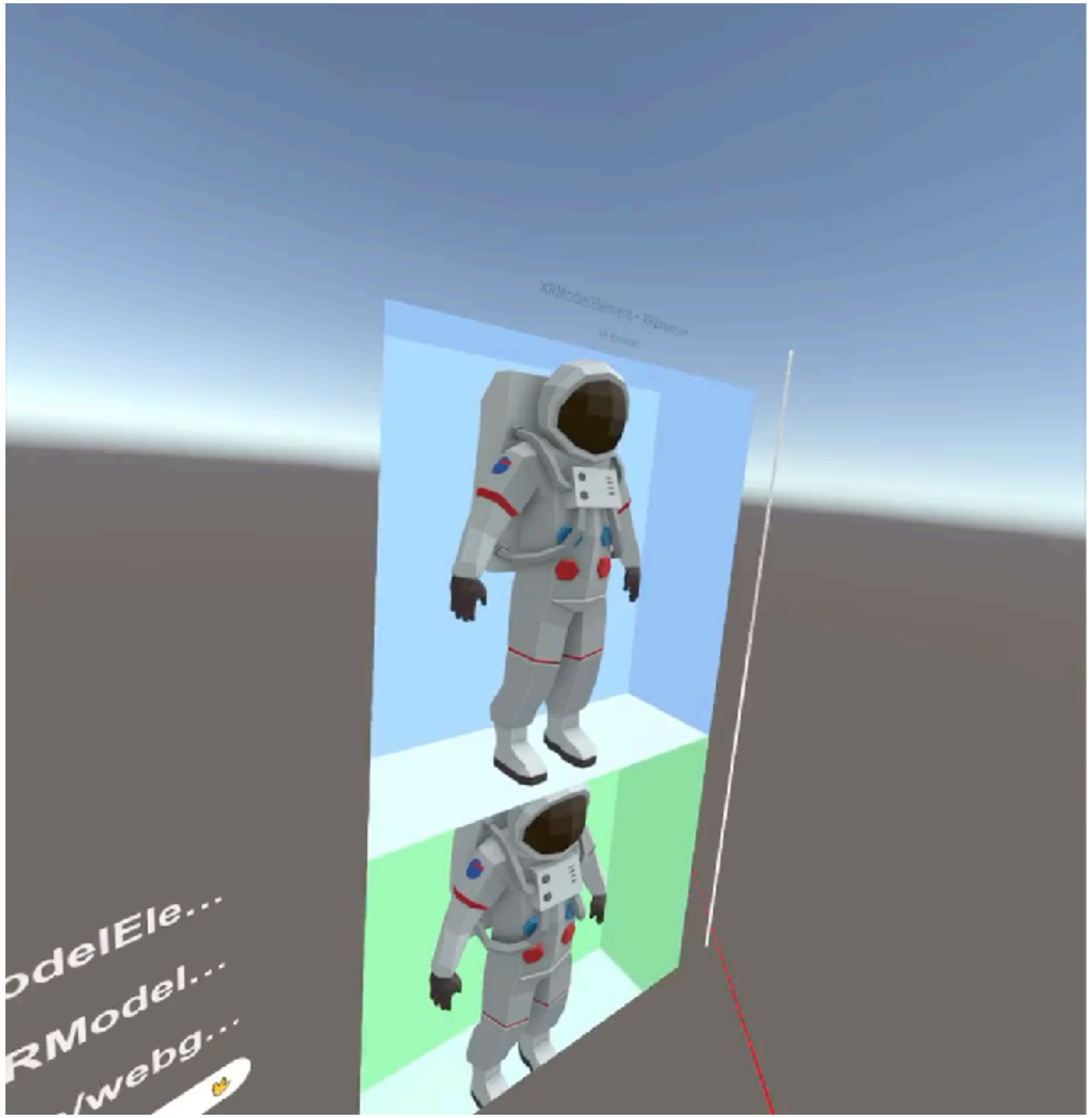
https://medium.com/@christophkrautz/what-are-the-coordinates-225f1ec0dd78

General Projections

https://www.flickr.com/photos/alsal/8865025942

General Projections

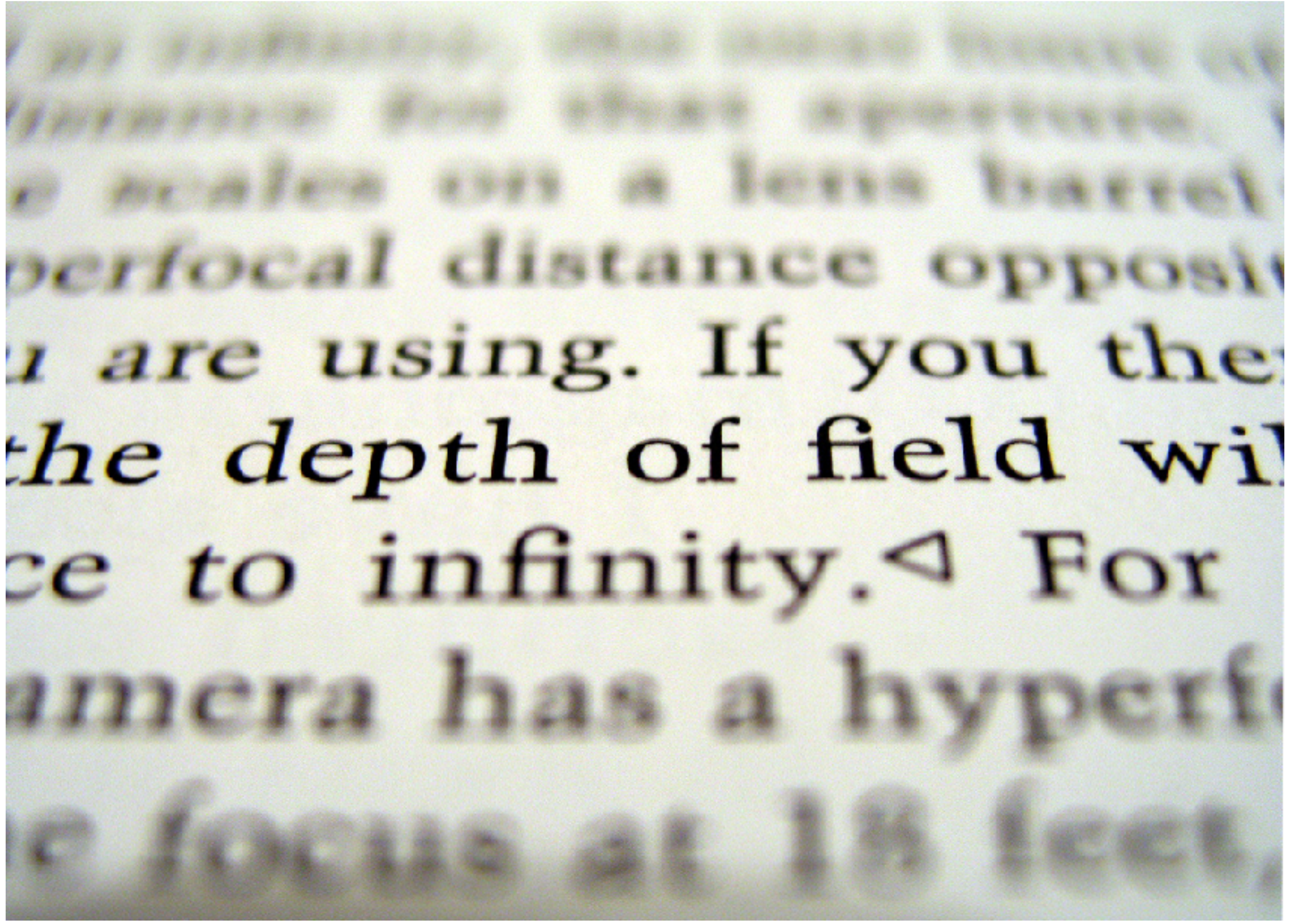
General Projections



https://jsantell.com/portals-with-asymmetric-projection/

Model Limitations

No depth of field



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

Lens Distortions

Locating the Camera

Camera Position

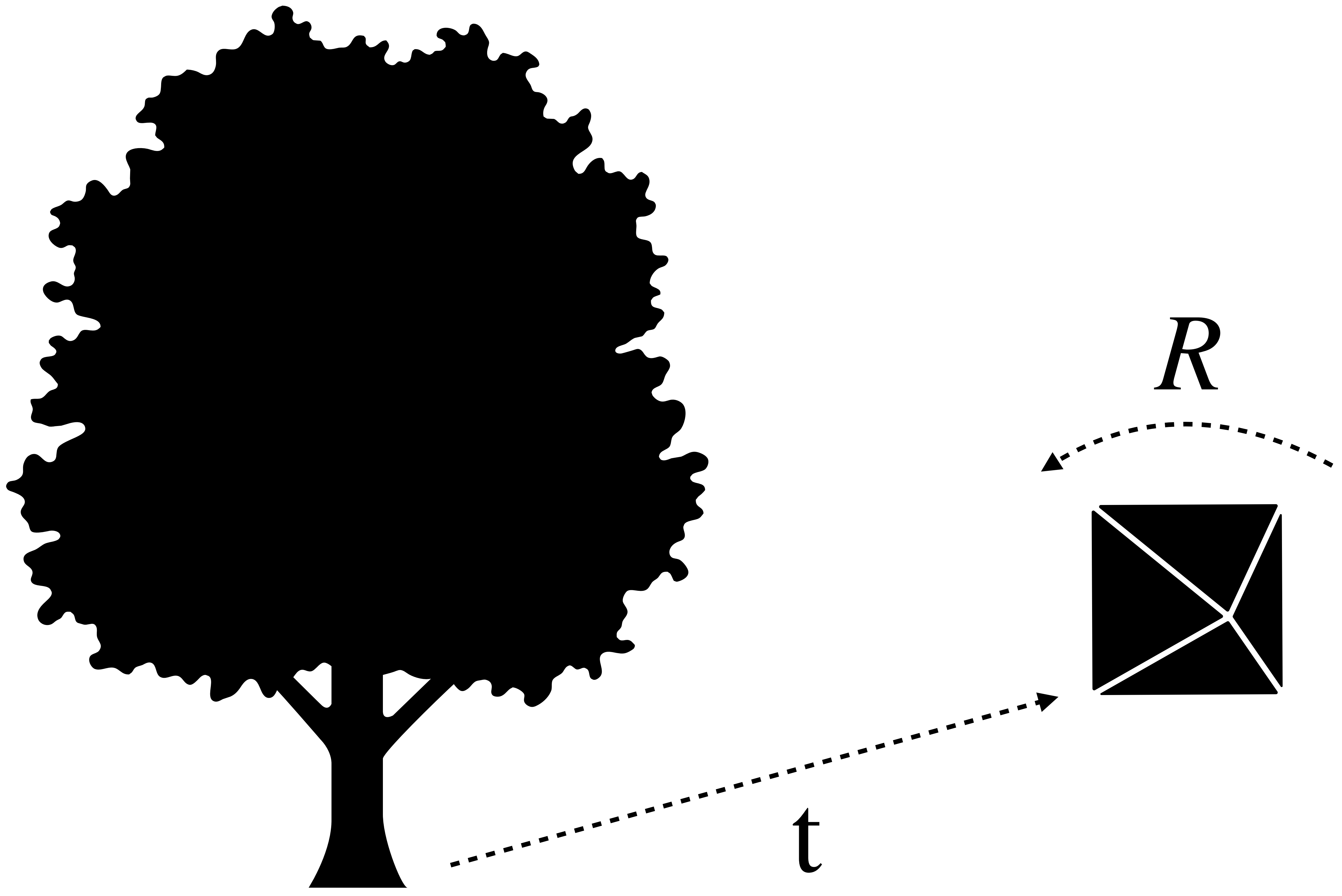
Extrinsic Parameters

- Rigid movement $\mathrm{y} = R(\mathrm{x} - \mathrm{t})$
- In homogeneous coordinates translates into

$$\mathrm{y} = \begin{bmatrix} R & -R\mathrm{t} \\ 0 & 1 \end{bmatrix} \mathrm{x} = C^{W2C}\mathrm{x}$$

- Perspective transformation turns into

$$H = \begin{bmatrix} f_x & 0 & p_x & 0 \\ 0 & f_y & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -Rt \\ 0 & 1 \end{bmatrix}$$

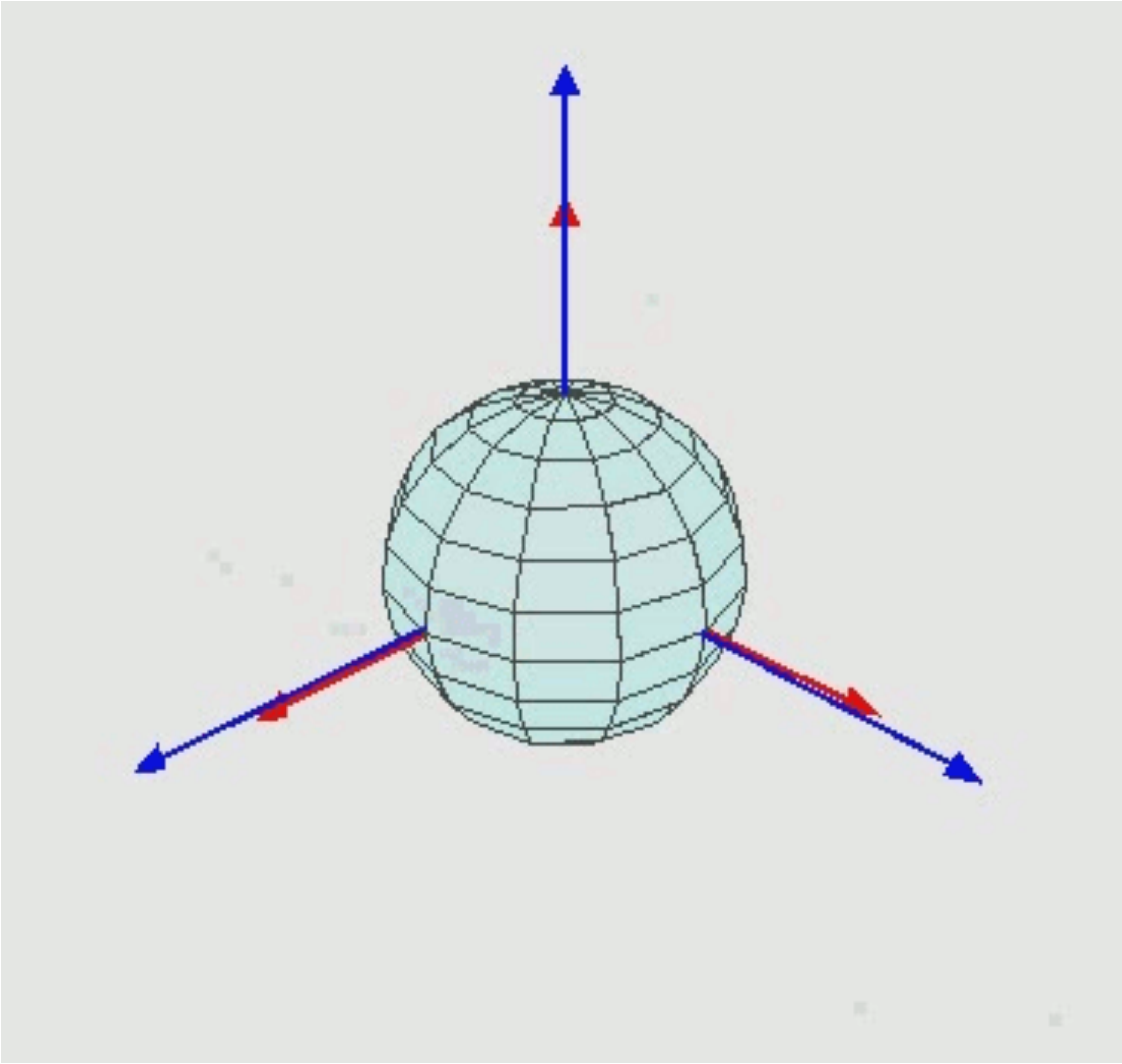


Degrees of freedom?

Rotation Representations

Euler Angles

- Composition of rotations along $z - x - z$ axes
- Can represent any rotation
- Is not injective
    - Loss of degree of freedom, gimbal lock



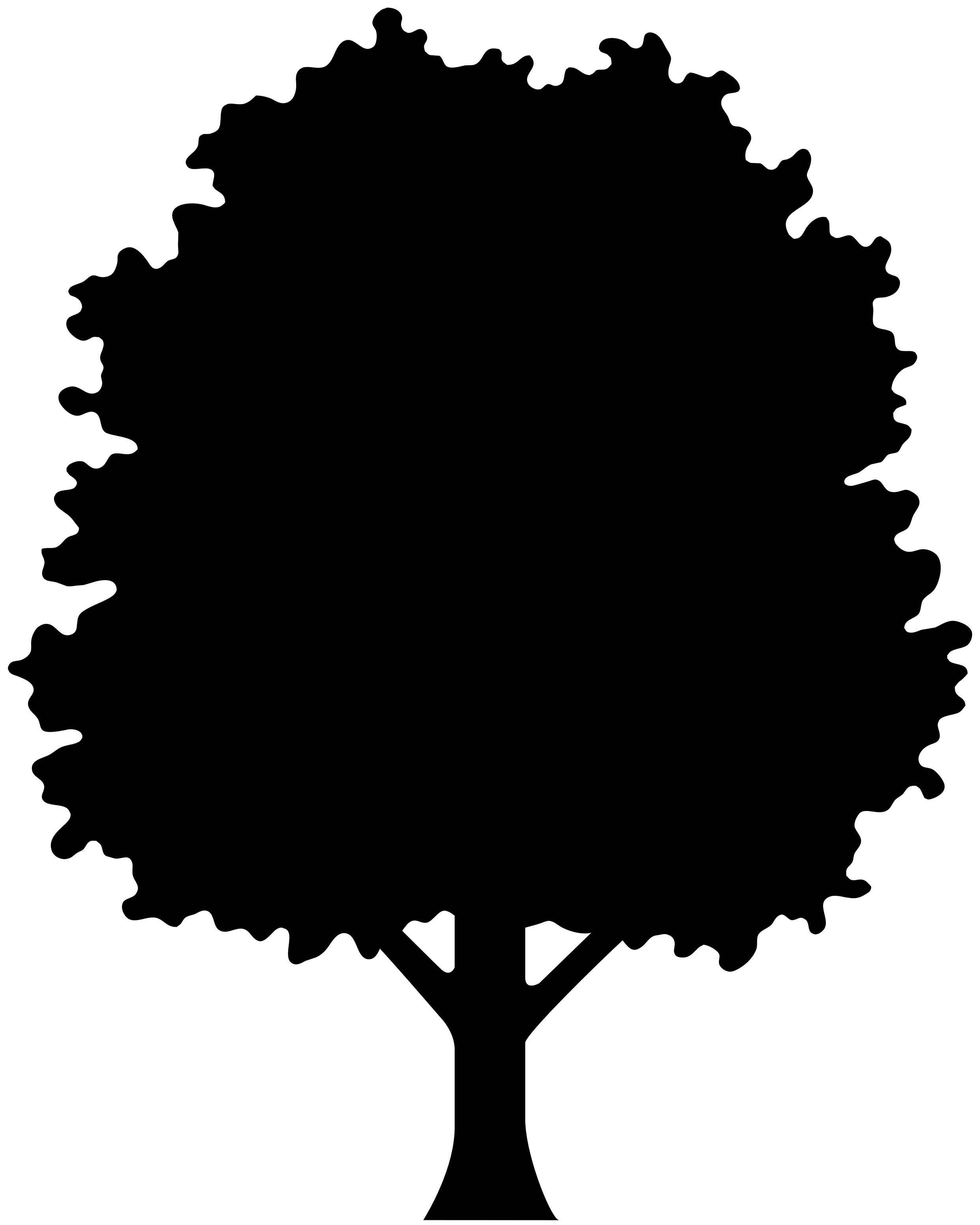
Rotation Representations

Angle Axis Representation

- Consider rotation $R \in SO(3)$
  - Eigenvalues lie on a unit circle in $\mathbb{C}$
  - $\|\mathrm{x}\|^2 = (R\mathrm{x}, R\mathrm{x}) = \lambda^2 \|x\|^2 \Rightarrow \lambda^2 = 1$
  - Eigenvector n with $\lambda = 1$ defines rotation axis
  - Rotation angle θ
- Unit quaternions also represent $SO(3)$: $(\mathrm{n} \cdot \sin\frac{\theta}{2}, \cos\frac{\theta}{2})$

“Look At” Camera

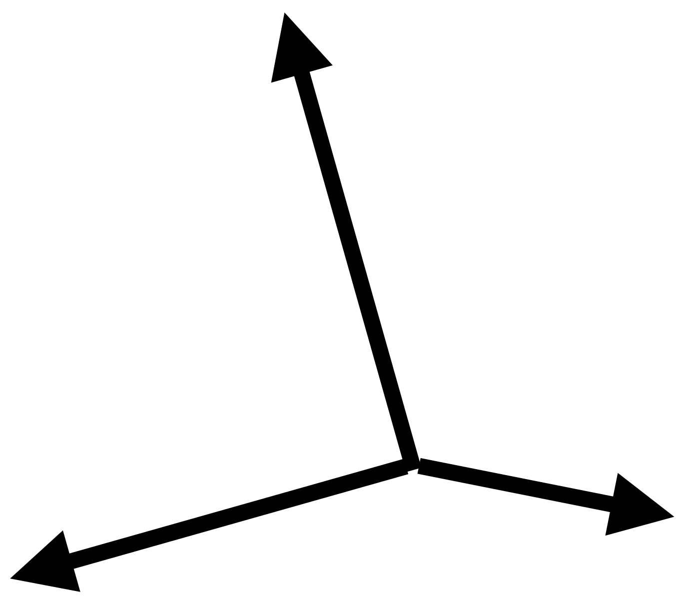
Looking at o from c



o

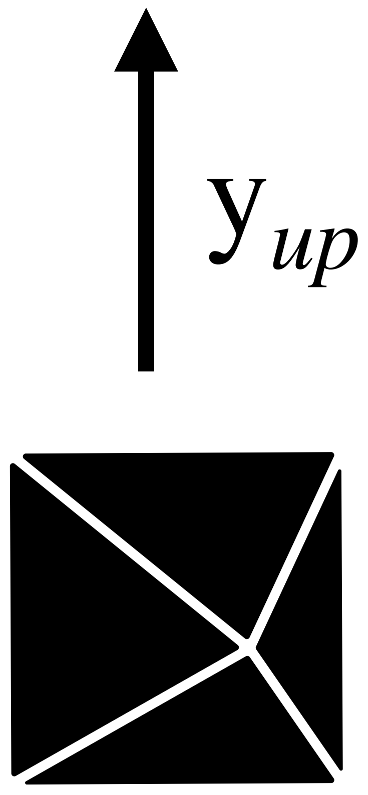
$$R = \left[\mathrm{x}^T \mathrm{y}^T \mathrm{z}^T\right]$$

$\mathrm{y} = \mathrm{x} \times \mathrm{z}$



$\mathrm{x} = \mathrm{z} \times \mathrm{y}_{up}$

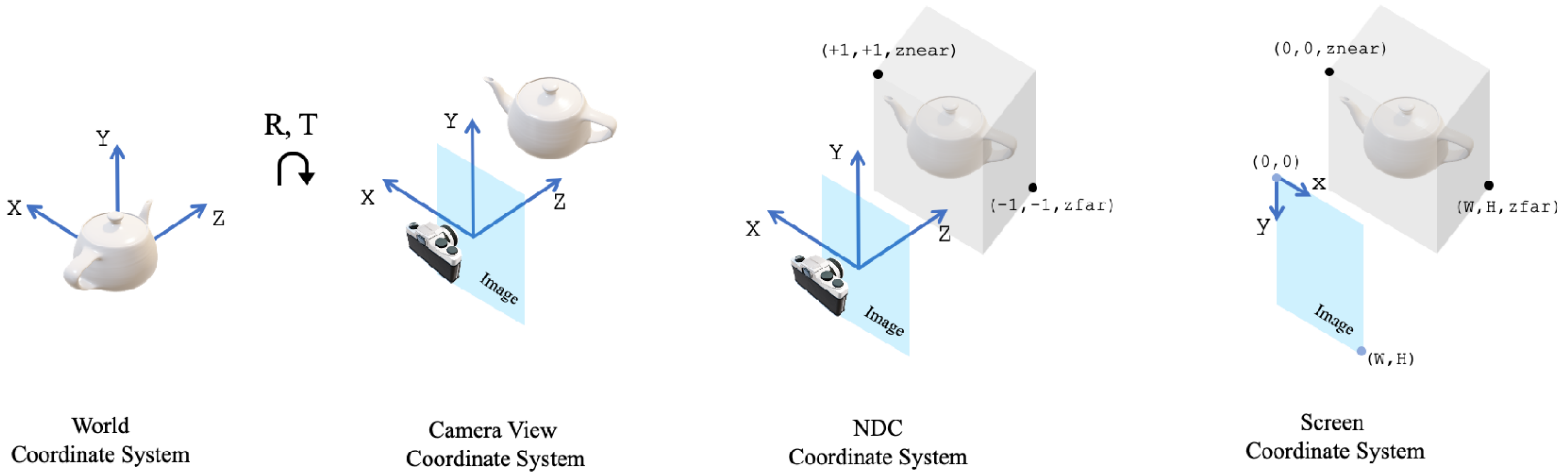
$\mathrm{z} = \|\mathrm{o} - \mathrm{c}\|^{-1}(\mathrm{o} - \mathrm{c})$



y_{up}

c

Another Note on Conventions



https://pytorch3d.org/docs/cameras

Key Takeaways

- Pinhole camera models
- Perspective projections and homogeneous coordinates
- Extrinsic and intrinsic parameters of a camera

- Next lecture: multi-view geometry