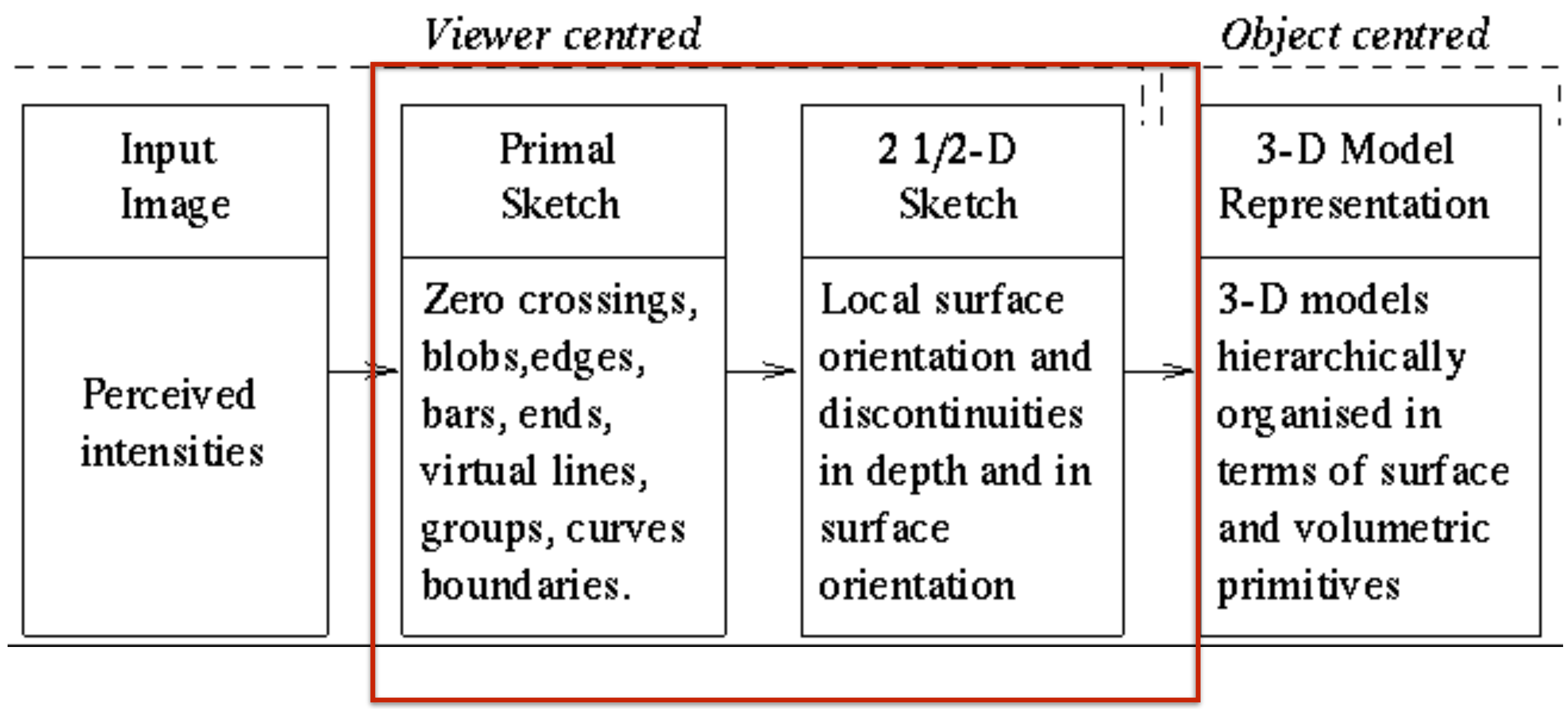
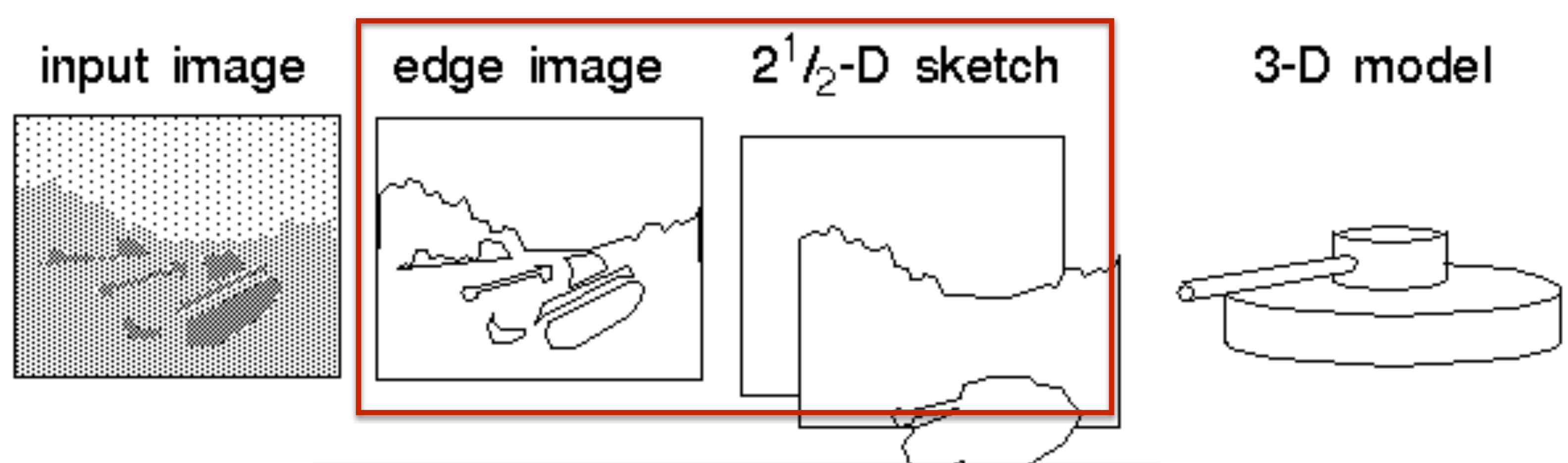
Image formation

Slide credits: Deva Ramanan, Jitendra Malik

David Marr's Taxonomy of Vision





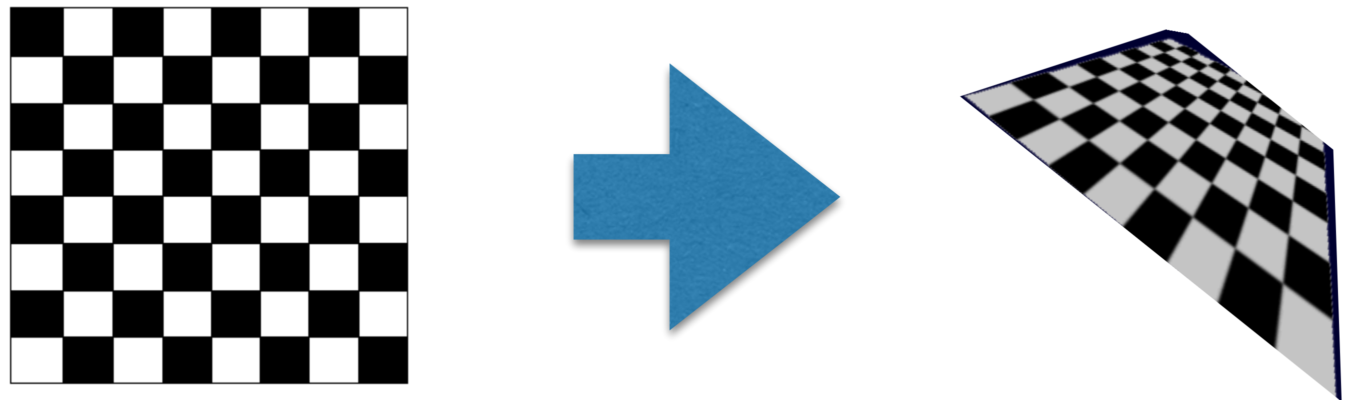
Low-level Mid-level High-level

Agenda

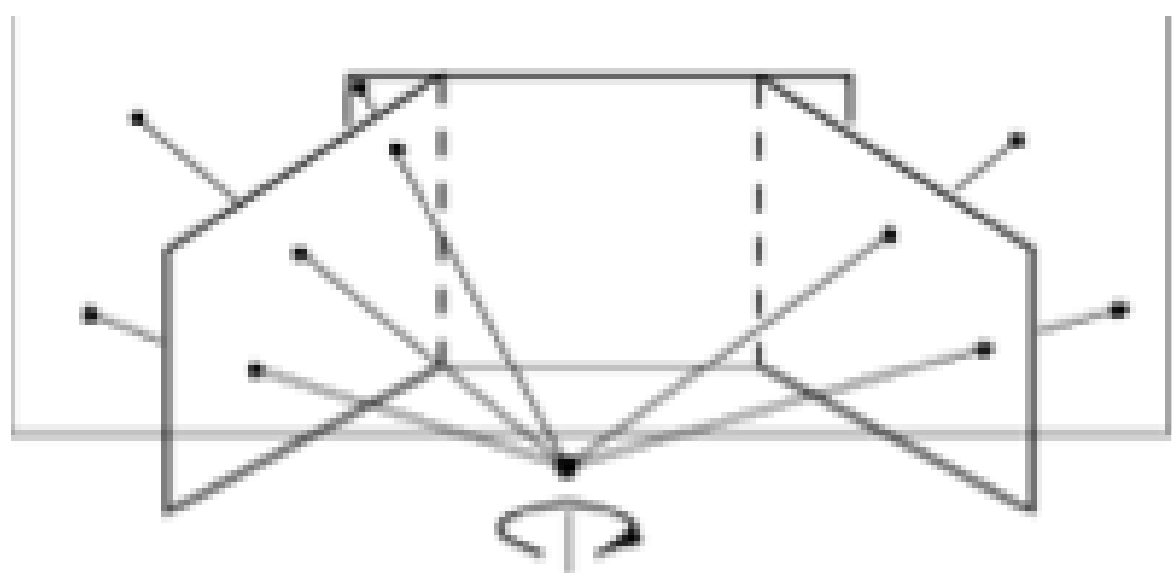
- Pinhole optics
  - Perspective projection (vanishing points, horizon, object height)
  - Projection models (orthographic, scaled orthographic, paraperspective)

Homography transformations

1. Models image projection of an arbitrary planar scene

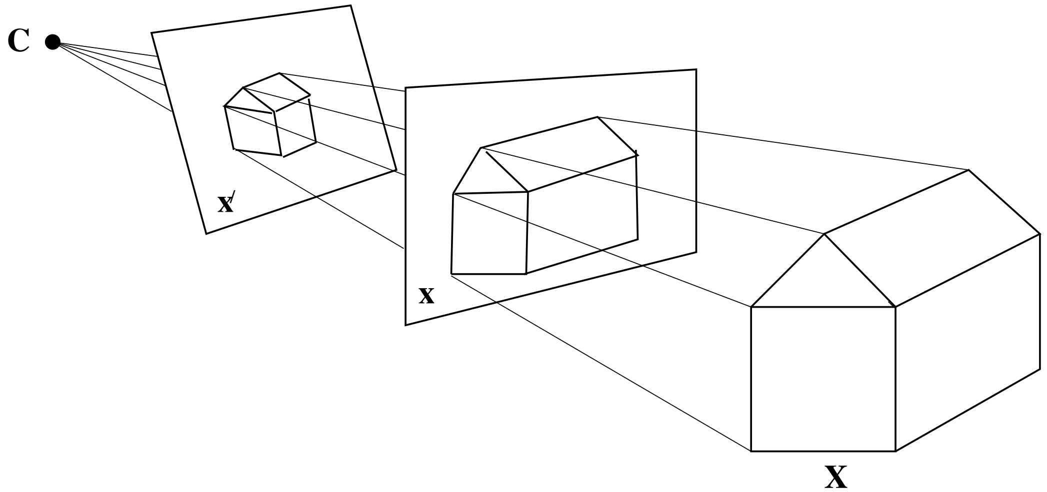


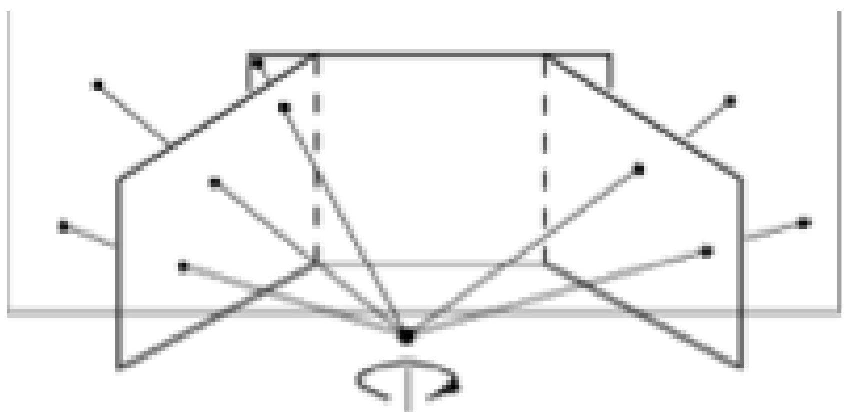
2. Models warping (perspective) effects from camera rotations



A homography is any planar transformation,
described an arbitrary 3x3 matrix applied to the image points!

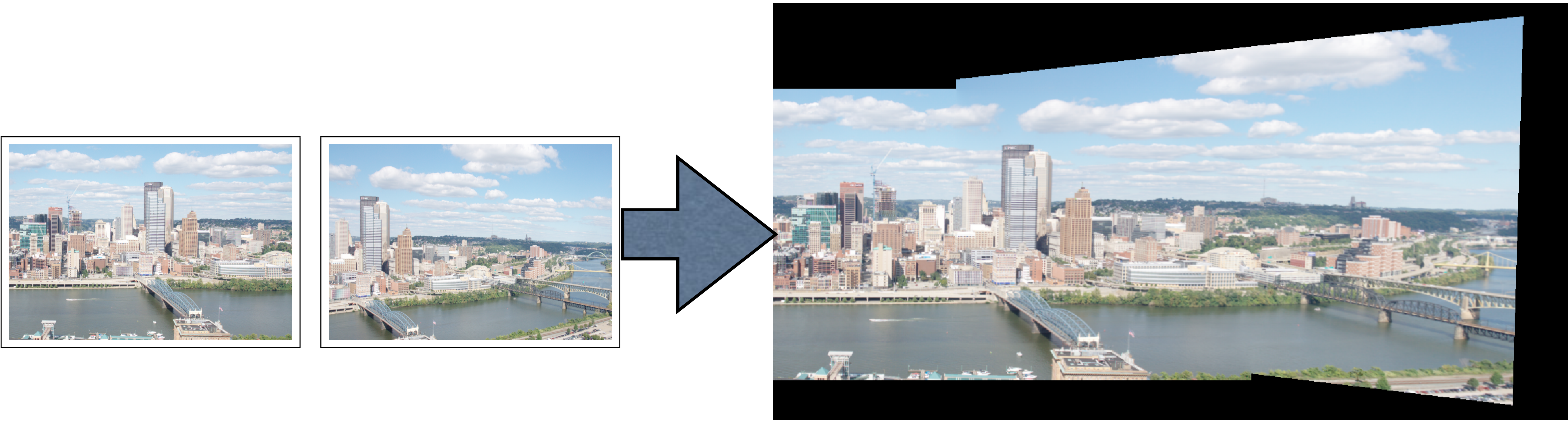
Special case of 2 views: rotations about camera center



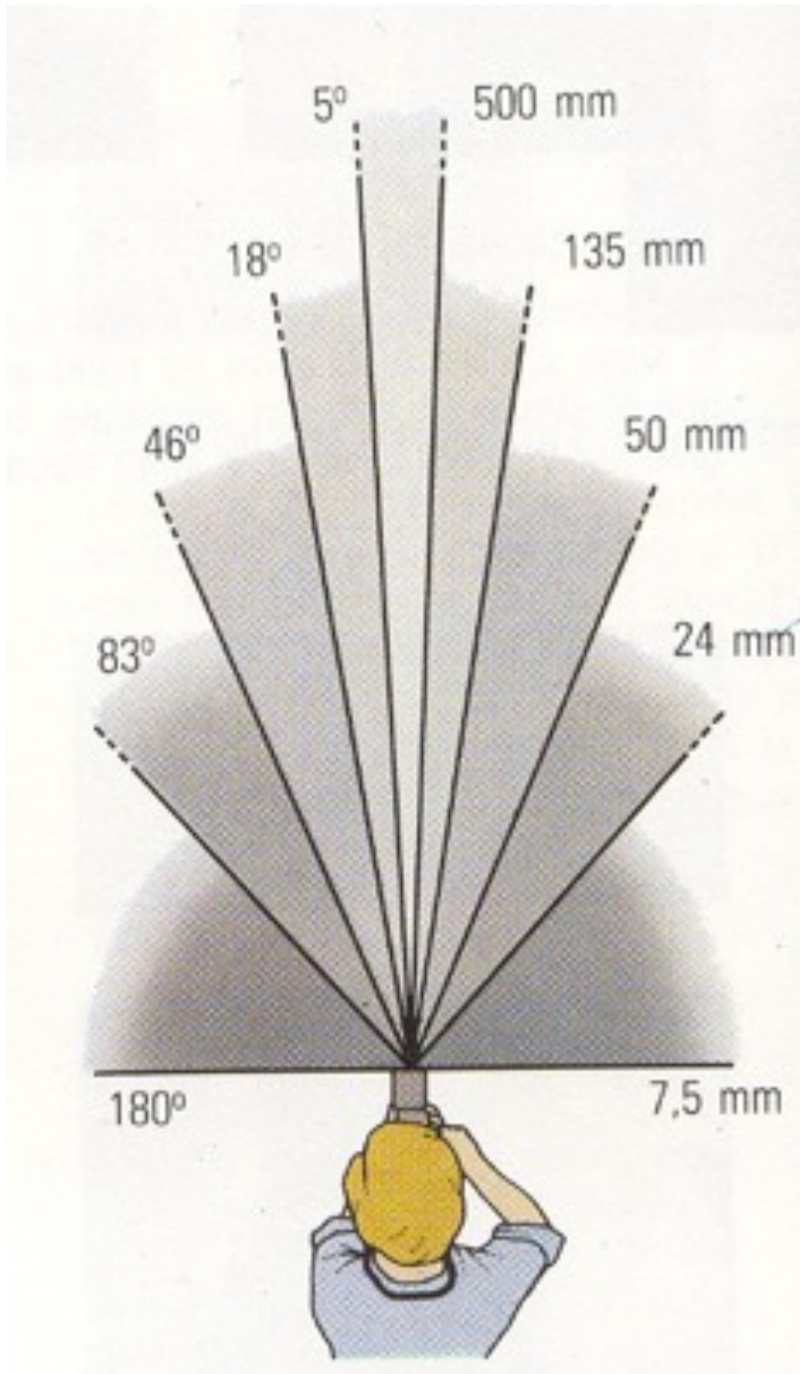


Can be modeled as planar transformations, *regardless of 3D scene geometry!*

Intuition: equivalent to looking at an image of a planar image



Field of view (FOV)



24mm

50mm

135mm

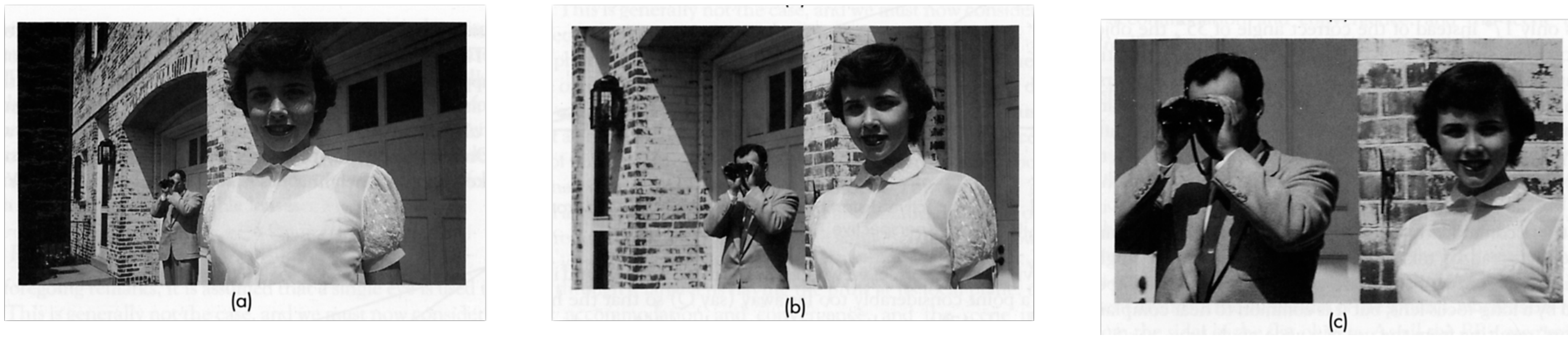
Varying the focal length

(changing the length of the shoebox)

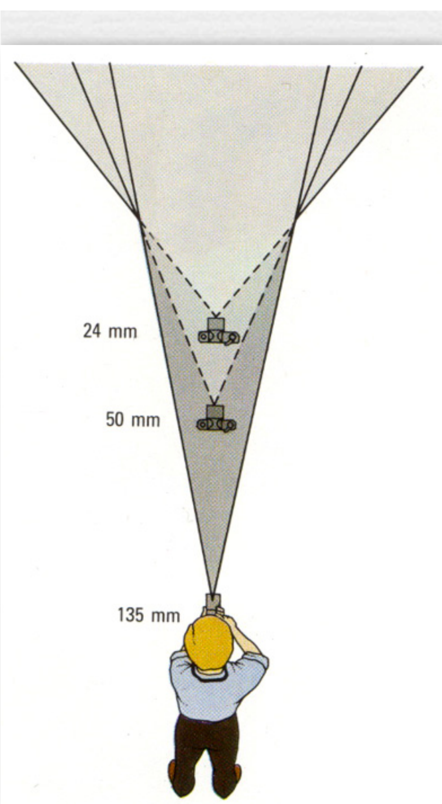
$$FOV = 2\arctan\left(\frac{\text{sensor size}}{2f}\right)$$

sensor size = size of the image

Increasing the focal length and stepping back



(a) (b) (c)



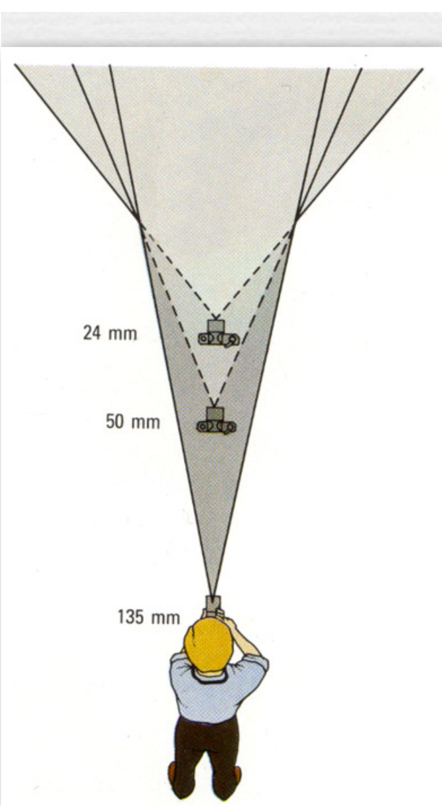
$$x = f\frac{X}{Z} \qquad y = Y\frac{f}{Z}$$

What happens to the object size if we step back (double our distance to the object) and double the focal length?

$$FOV = 2\arctan\left(\frac{\text{sensor size}}{2f}\right)$$

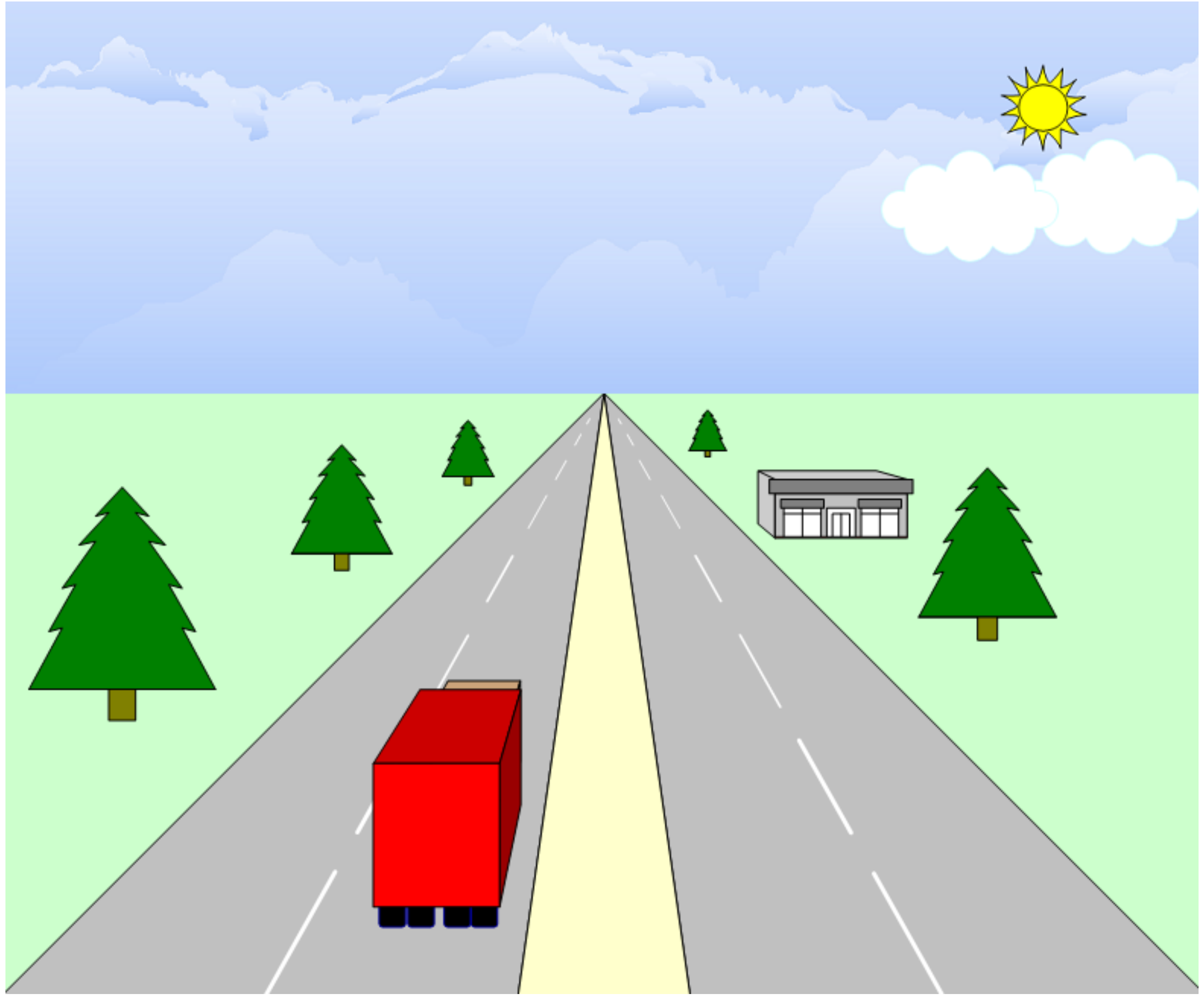
What happens to the FOV if we step back (double our distance to the object) and double the focal length?

Decreasing the focal length and moving forward



Increases the FOV but keeps object size the same!

Perspective projection



Closer objects appear larger

Closer objects are lower in the image

Parallel lines meet

All these can be simply derived with $x = X\dfrac{f}{Z} \quad y = Y\dfrac{f}{Z}$

Closer points are lower in the image

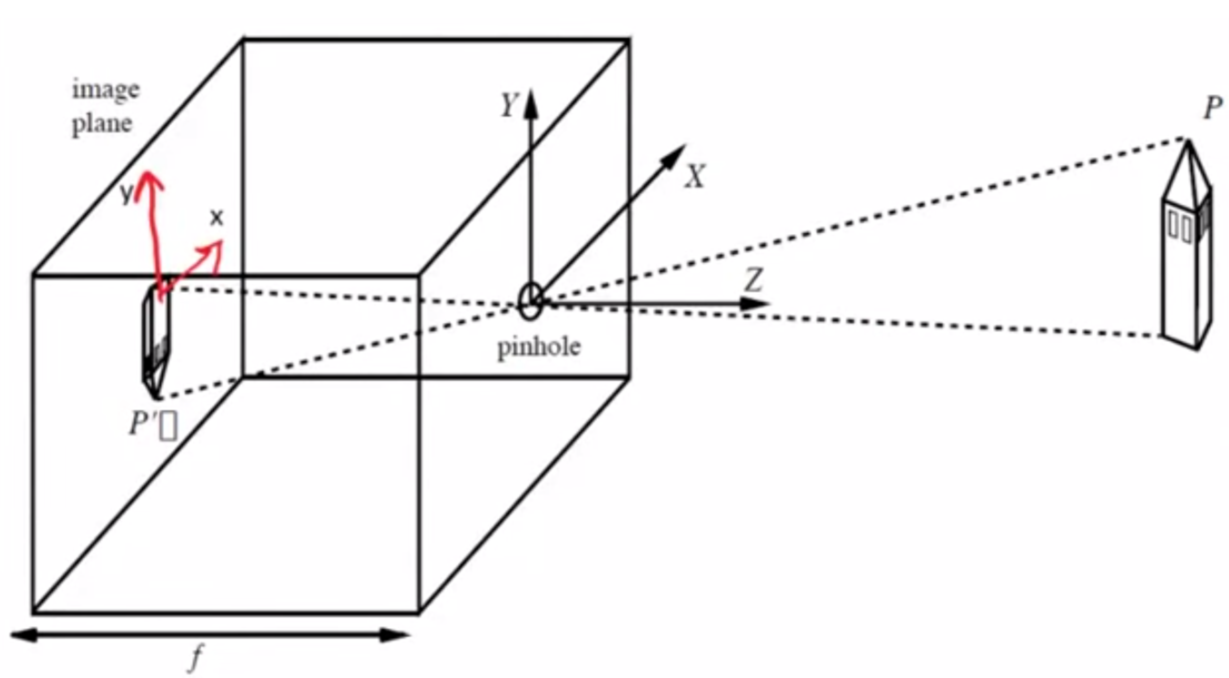
Image coordinates of a point (X,Y,Z): $x = X\frac{f}{Z}$ $y = Y\frac{f}{Z}$

Equation of ground plane is Y = -H

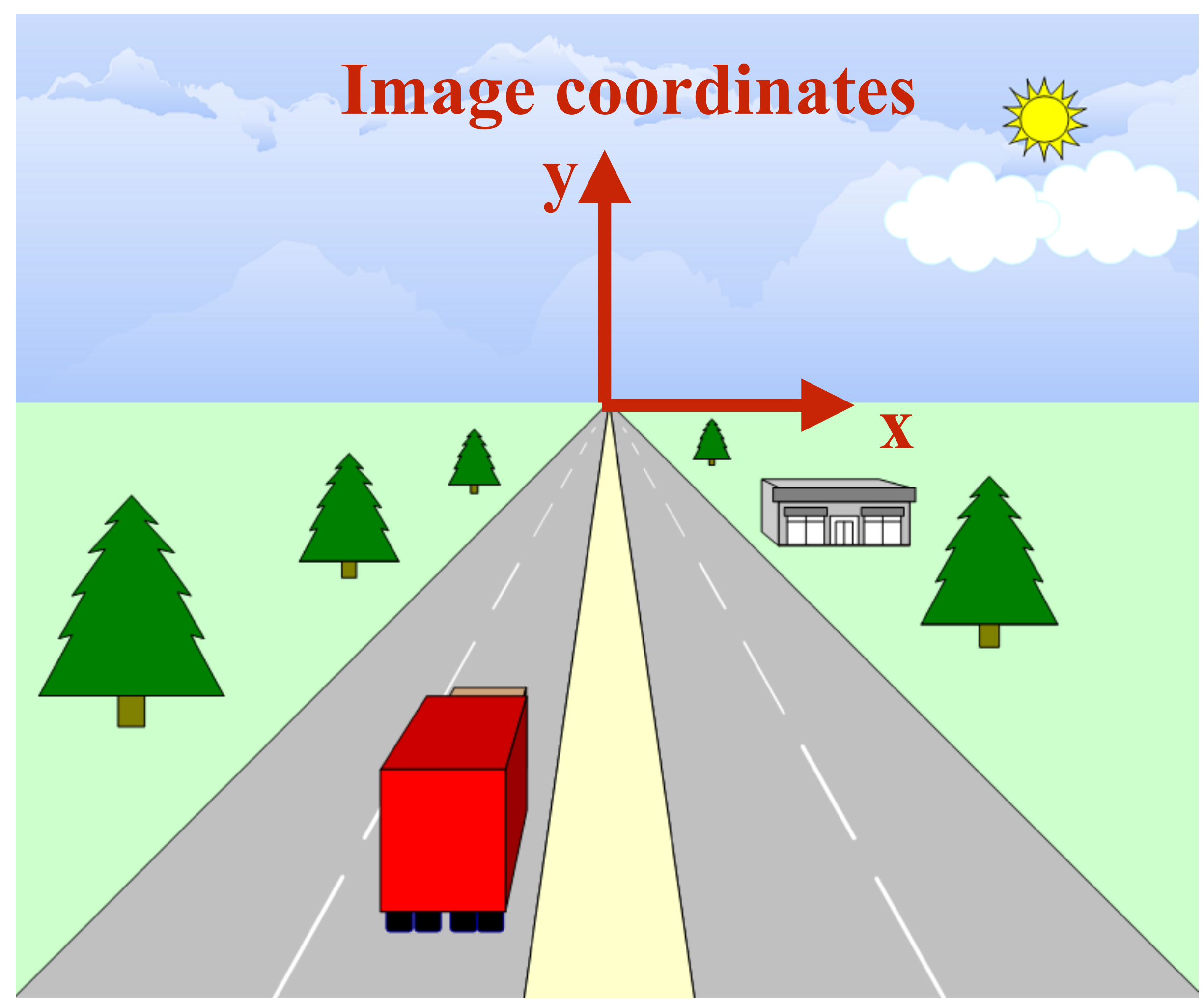
A point on ground plane will have y-coordinate = ? $-H\frac{f}{Z}$

smaller Z = larger negative y

closer to camera **lower in image**



Z = distance from the camera



Closer objects appear larger in the image

World coordinates of bottom of tree: (X, -H, Z)
(on the ground plane)

World coordinates of top of tree: (X, L-H, Z)

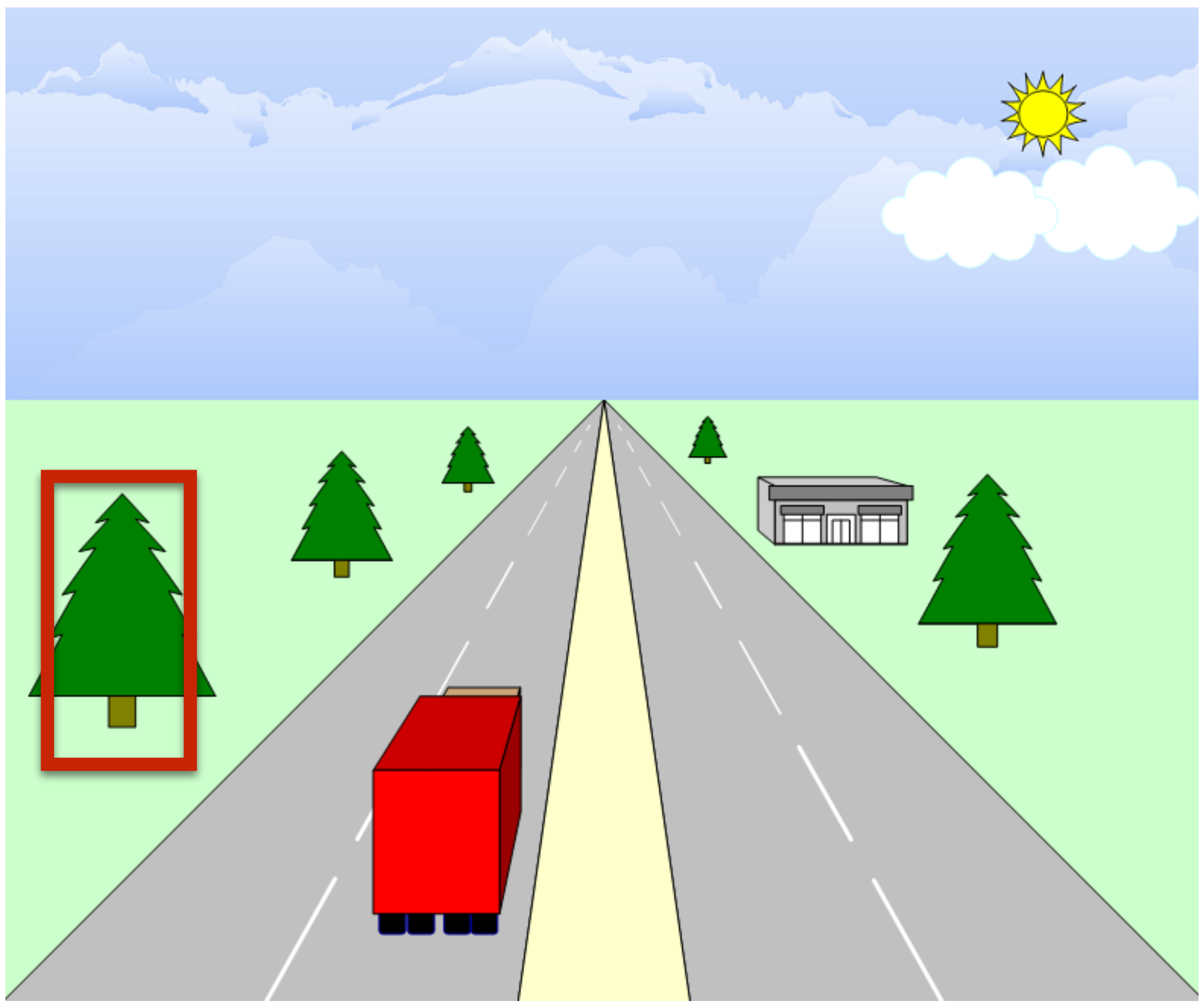


Image projection equations:

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z}$$

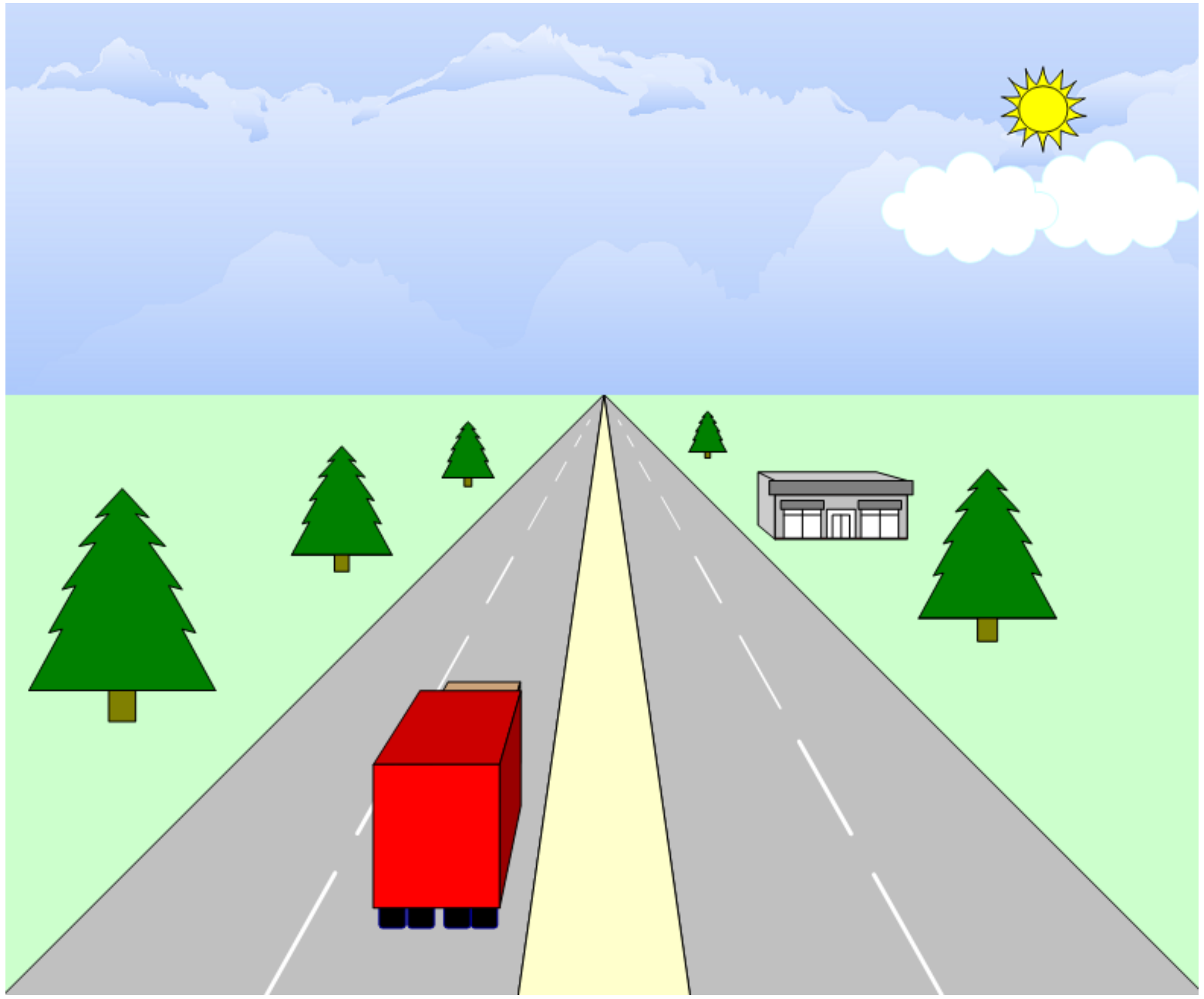
$$y_{top} = \frac{f(L-H)}{Z}$$

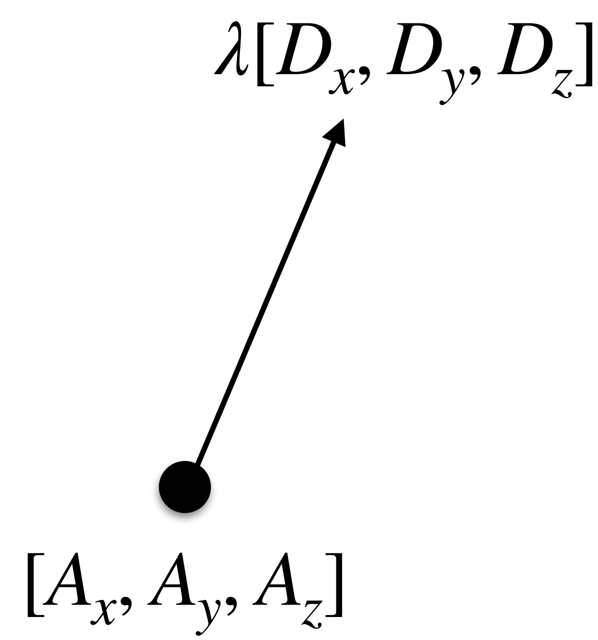
$$y_{bot} = \frac{f(-H)}{Z}$$

$$y_{top} - y_{bot} = \underbrace{\frac{f(L-H)}{Z}}_{y_{top}} - \underbrace{\frac{-fH}{Z}}_{y_{bot}} = \frac{fL}{Z}$$

Smaller Z ->
Larger image height

Parallel Lines Meet



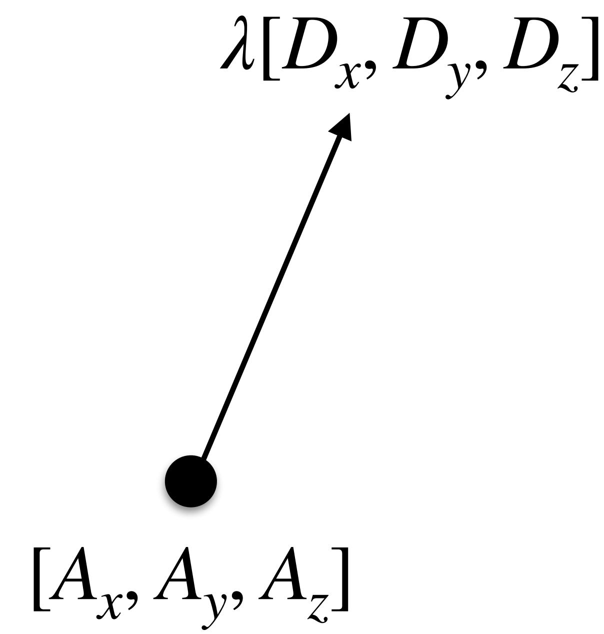


Line: $$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} A_X \\ A_Y \\ A_Z \end{bmatrix} + \lambda \begin{bmatrix} D_X \\ D_Y \\ D_Z \end{bmatrix}$$

Parallel line has the same direction $[D_x, D_y, D_z]$
but a different starting point $[A_x, A_y, A_z]$

Parallel Lines Meet

Line: $$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} A_X \\ A_Y \\ A_Z \end{bmatrix} + \lambda \begin{bmatrix} D_X \\ D_Y \\ D_Z \end{bmatrix}$$



Compute projected point (x,y) as lambda approaches infinity:

$$x = \frac{fX}{Z} = \frac{f(A_x + \lambda D_x)}{A_z + \lambda D_z} \to \frac{fD_x}{D_z} \quad \text{as} \quad \lambda \to \infty$$

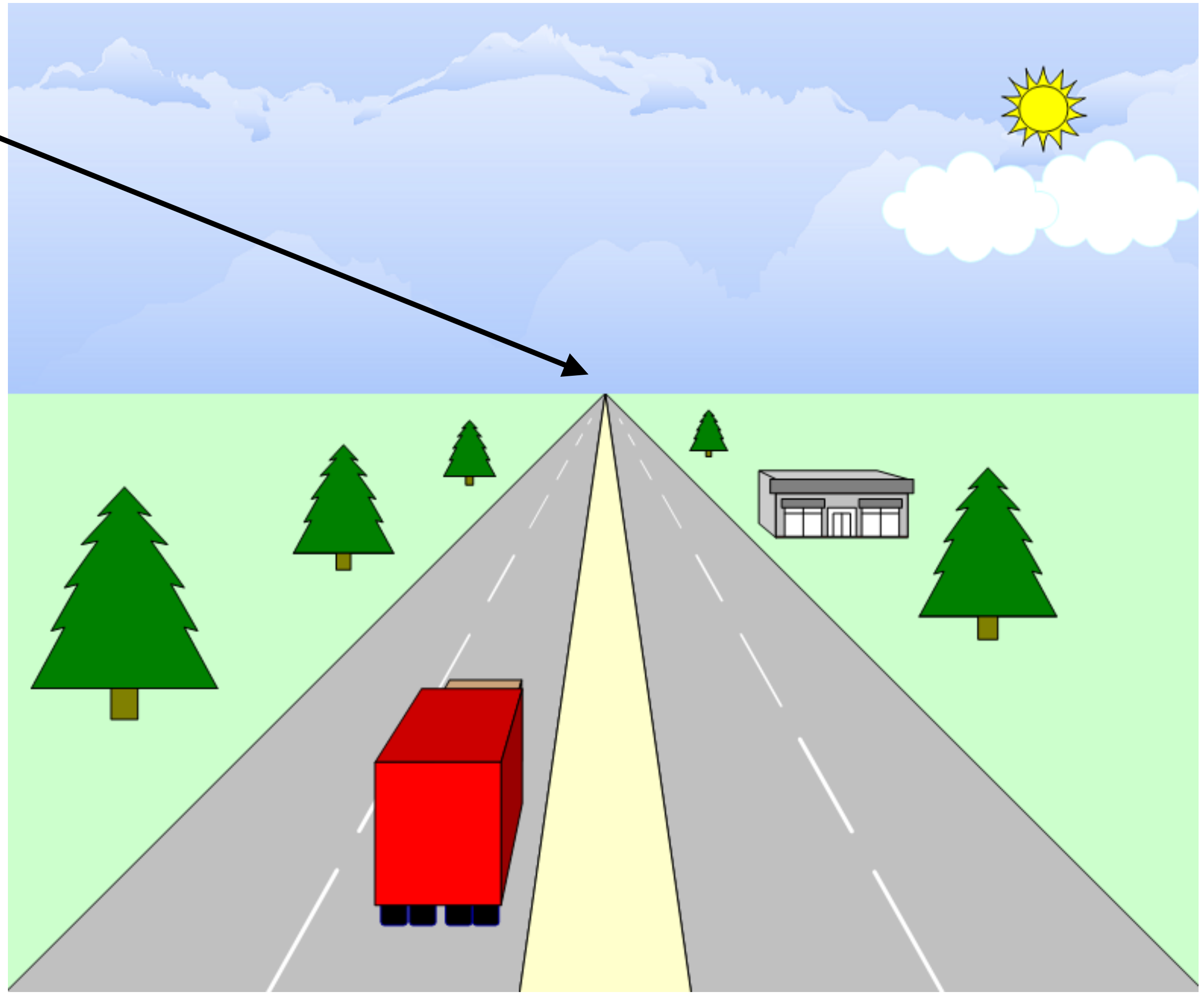
$$y = \frac{fY}{Z} = \frac{f(A_y + \lambda D_y)}{A_z + \lambda D_z} \to \frac{fD_y}{D_z} \quad \text{as} \quad \lambda \to \infty$$

- Parallel 3D lines (with identical direction vectors) meet at same 2D image location or “vanishing point”
- Specifically they meet at $[x, y] = [f\frac{D_x}{D_z}, \ \ f\frac{D_y}{D_z}]$ (vanishing point)

Perspective projection

Parallel lines meet at $[x, y] = [f\dfrac{D_x}{D_z},\ \ f\dfrac{D_y}{D_z}]$ (vanishing point)

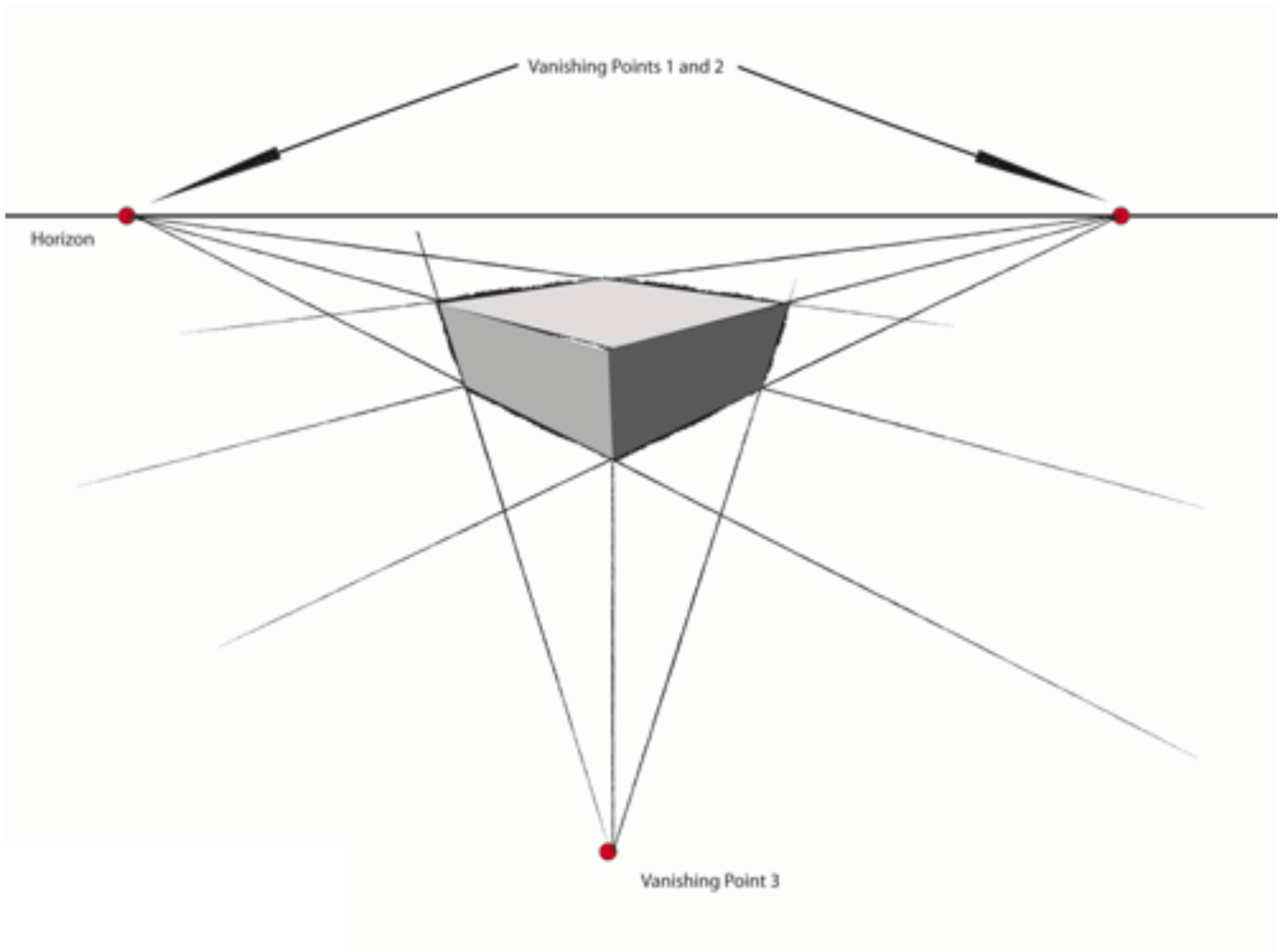
vanishing point



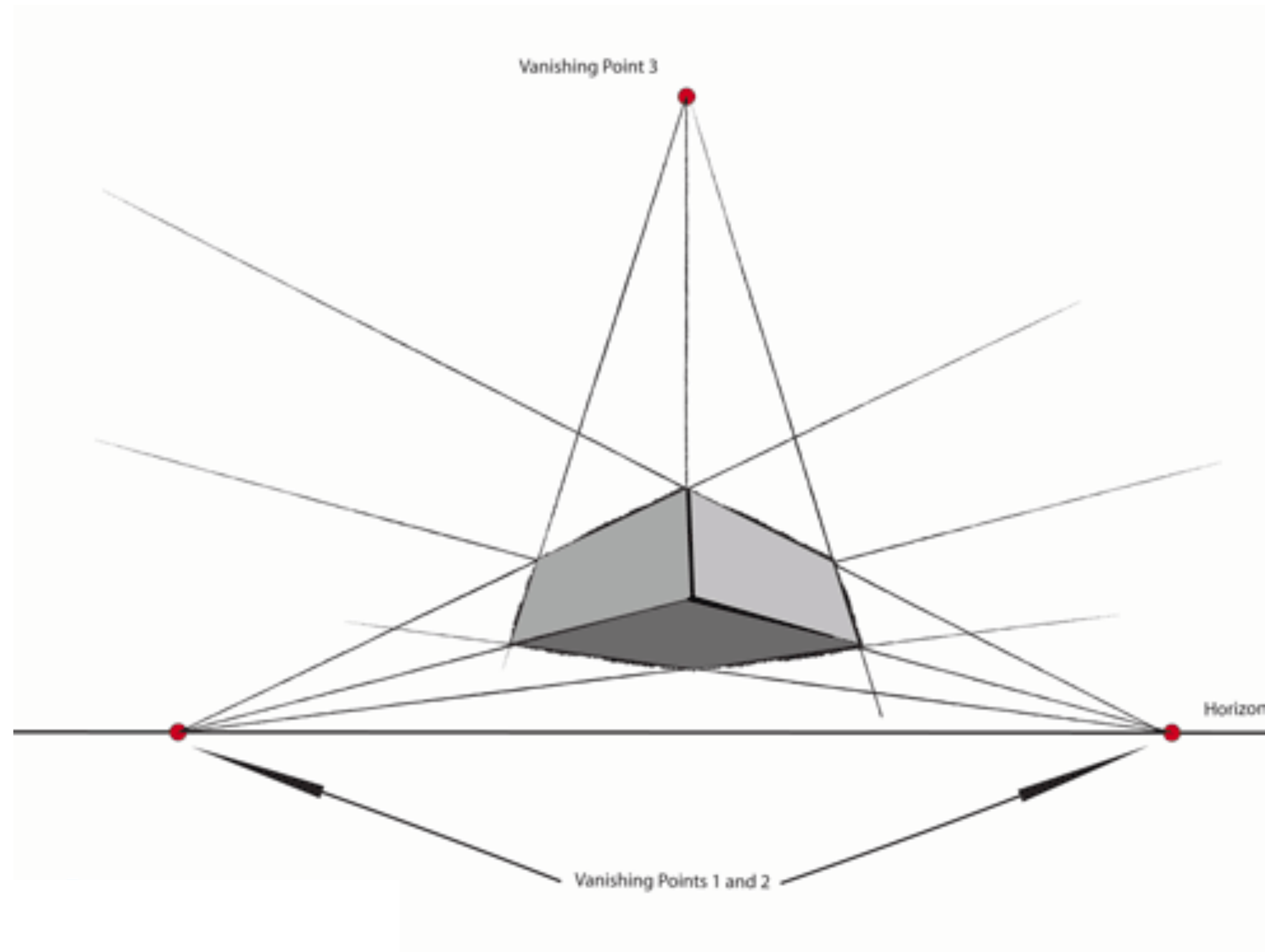
Parallel lines meet

A different set of parallel lines (in a different direction) will meet at a different vanishing point

Cube: 3 vanishing points



Cube from above

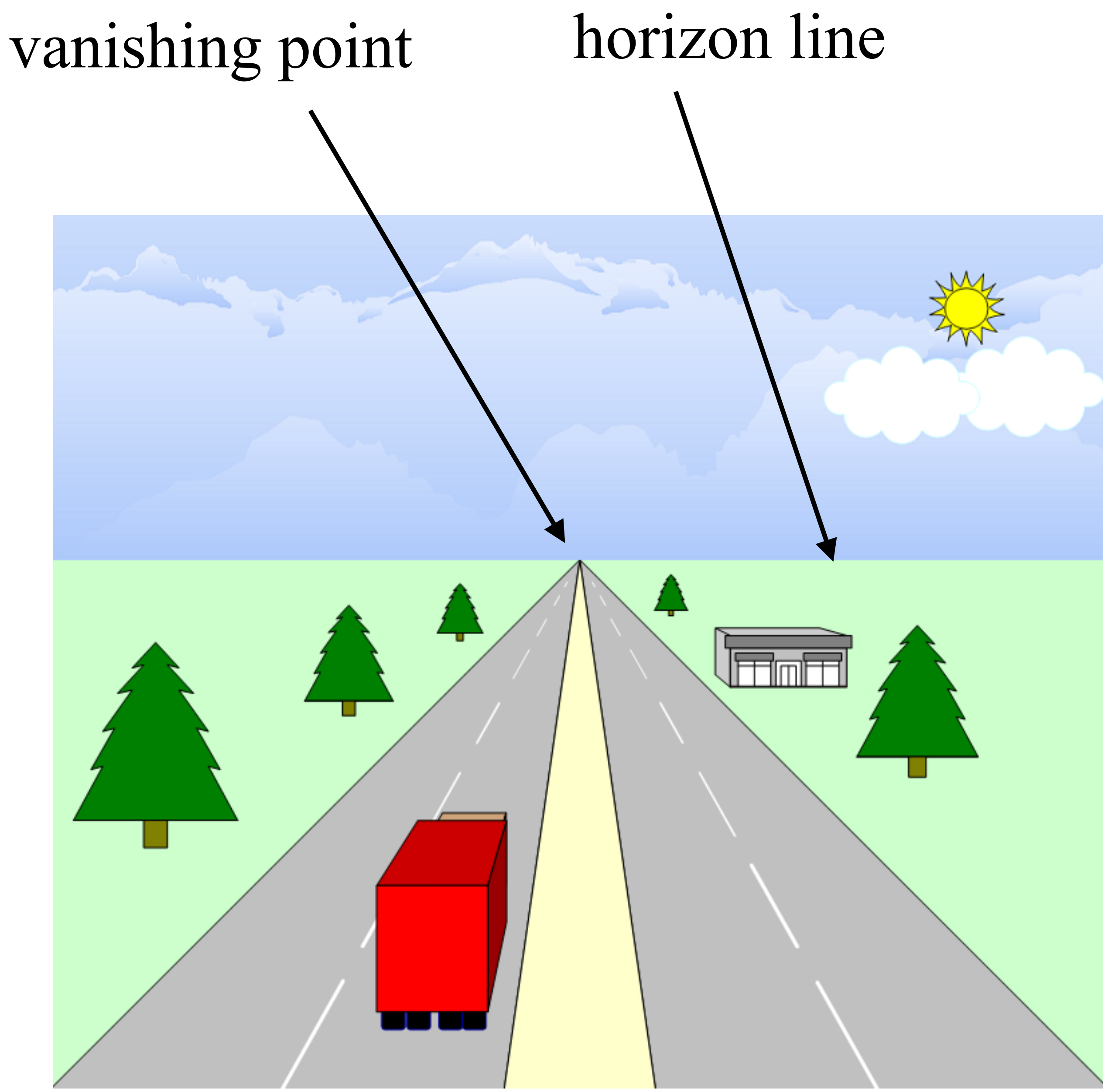


Cube from below

What happens if we look at a cube “head-on”?

Think about this after class:
what if the direction is perfectly vertical: $[0, D_y, 0]$

Special case: horizon line

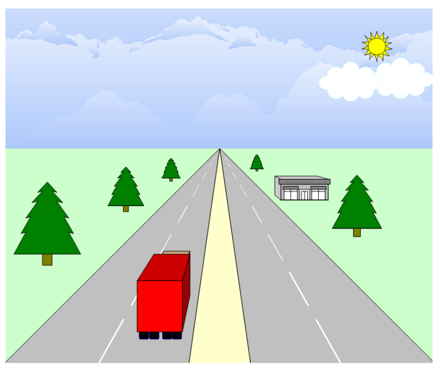


Claim: all 3D lines on ground plane meet at a *horizon line*

Horizon line: proof

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} A_X \\ A_Y \\ A_Z \end{bmatrix} + \lambda \begin{bmatrix} D_X \\ D_Y \\ D_Z \end{bmatrix}$$

$$(x, y) \rightarrow \left(\frac{fD_x}{D_z}, \frac{fD_y}{D_z}\right) \quad \text{as} \quad \lambda \rightarrow \infty$$

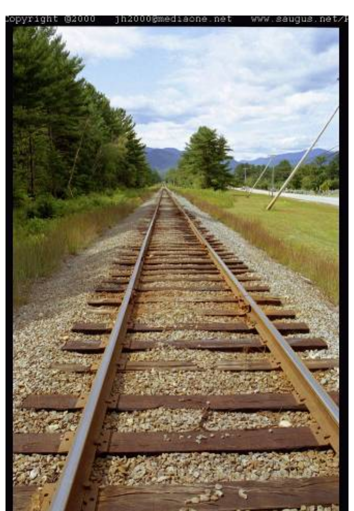


Equation of ground plane is Y = -H

(a distance of H below the camera)

Points A on the ground plane: (A_x, -H, A_z)

Direction D along ground plane (D_x, 0, D_z)



Where will the vanishing points converge to?

$$(x, y) \rightarrow \left(\frac{fD_x}{D_z}, \frac{fD_y}{D_z}\right) \quad = \left(\frac{fD_x}{D_z}, 0\right)$$

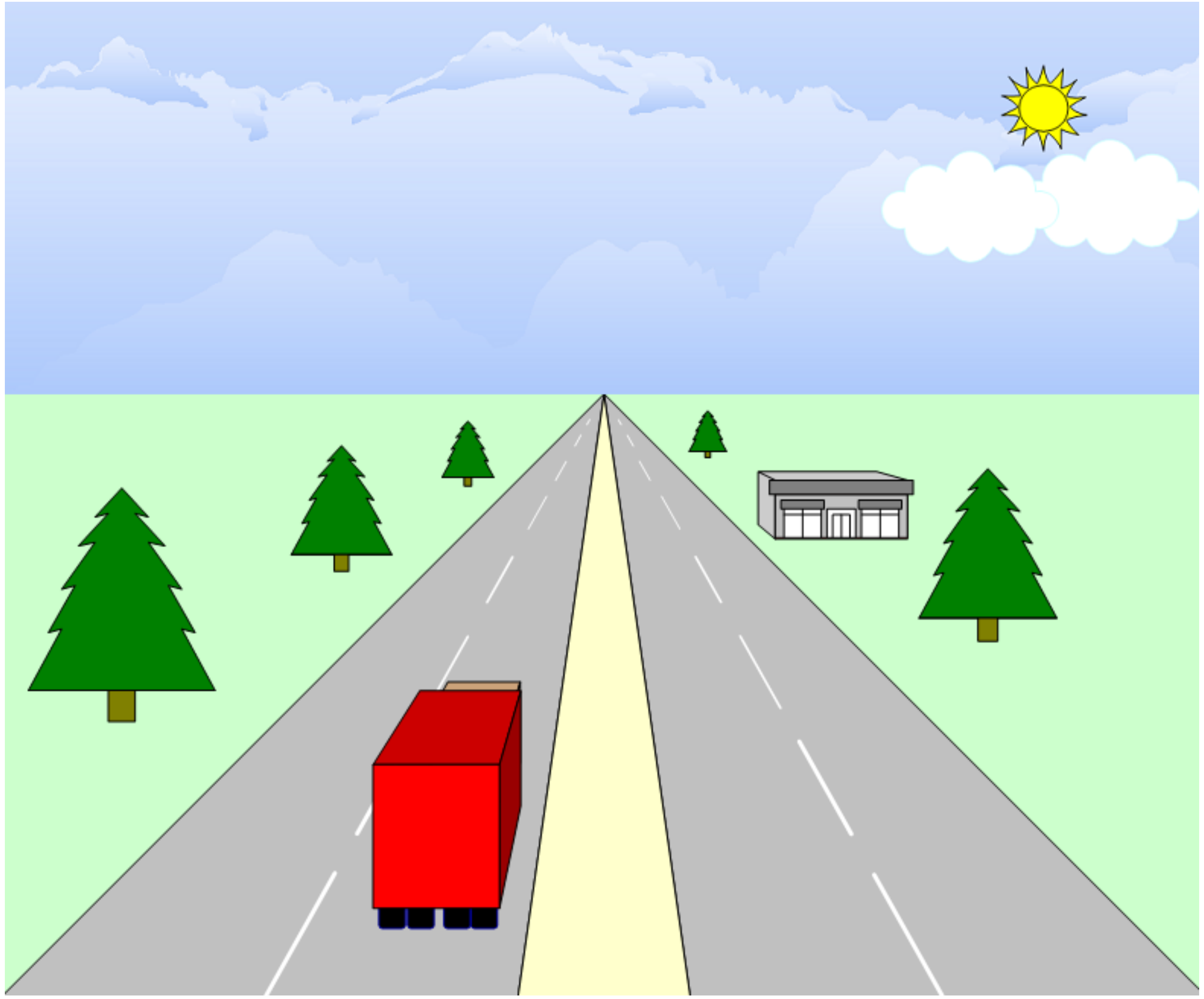
Line with y = 0
(This is the horizon line)

Question: why is the horizon line not always at center of image?

The camera is not always pointing parallel to the ground plane;
lines on ground not exactly (D_x, 0, D_z)

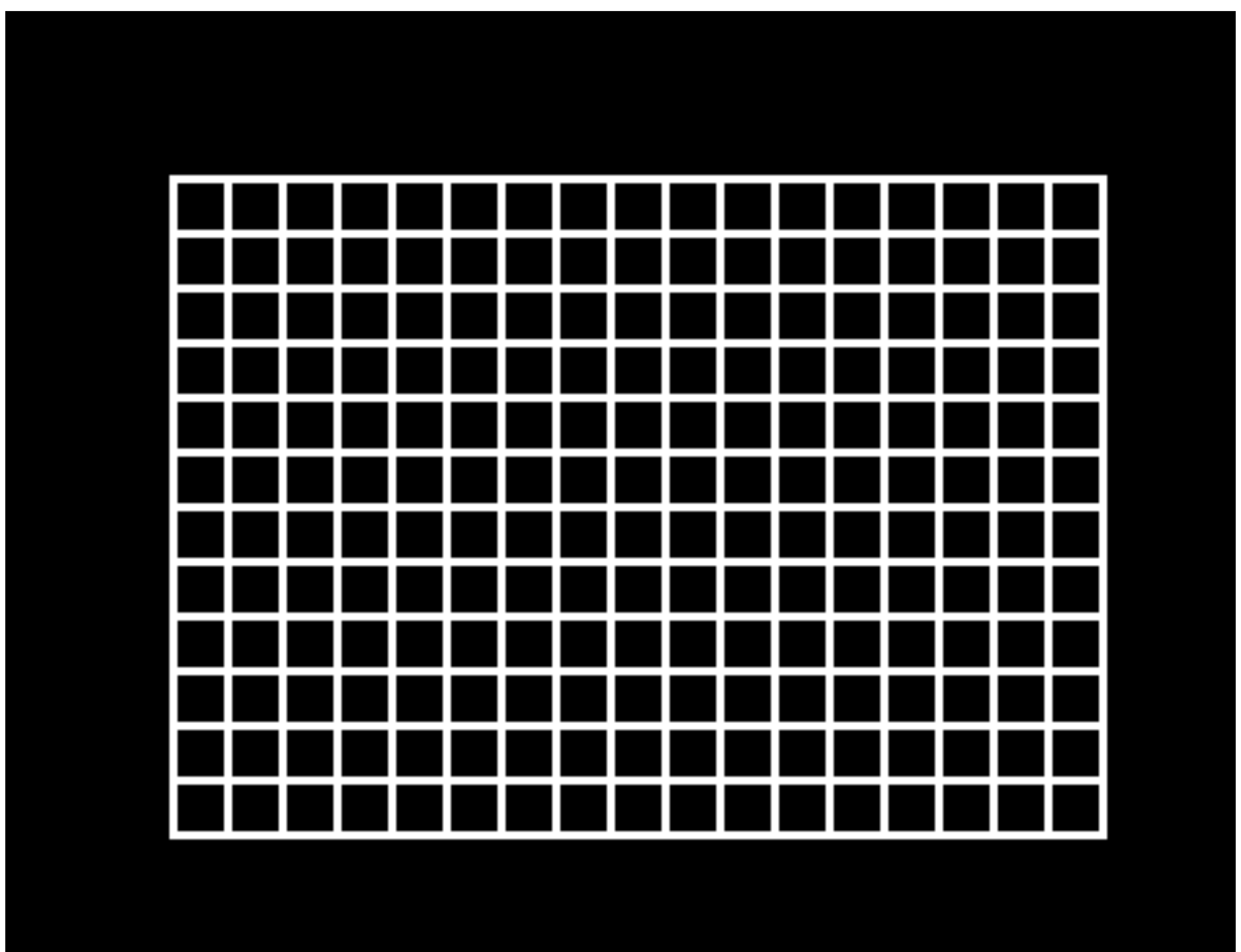
Consequence: Distances and angles aren't preserved in camera projection

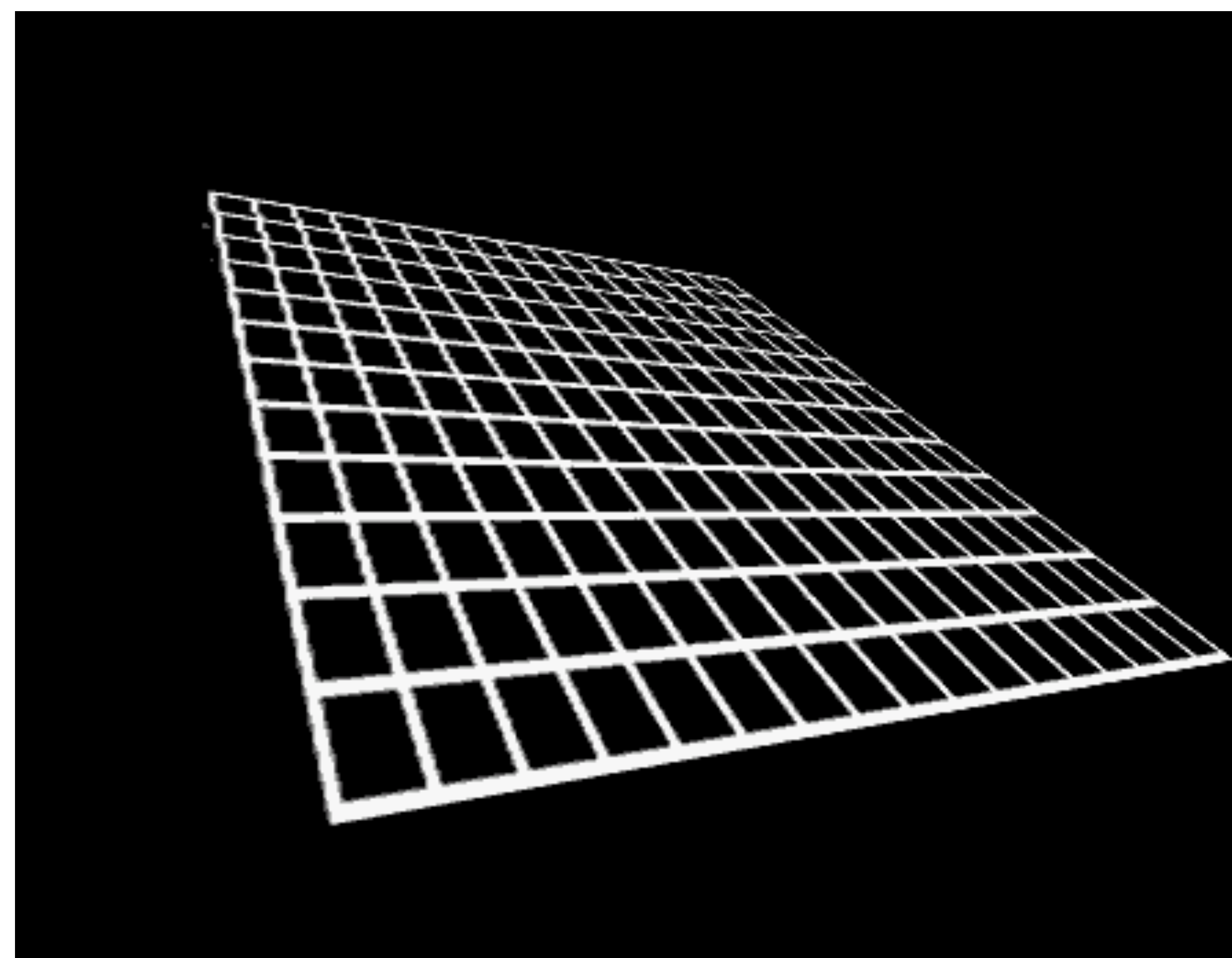
If 2 points are the same distance d apart in the world,
they might not be the same number of pixels apart in the image!



Consequence: Distances and angles aren't preserved in camera projection

If 3 points have an angle θ between them in the world,
they might have a different angle between them in the image

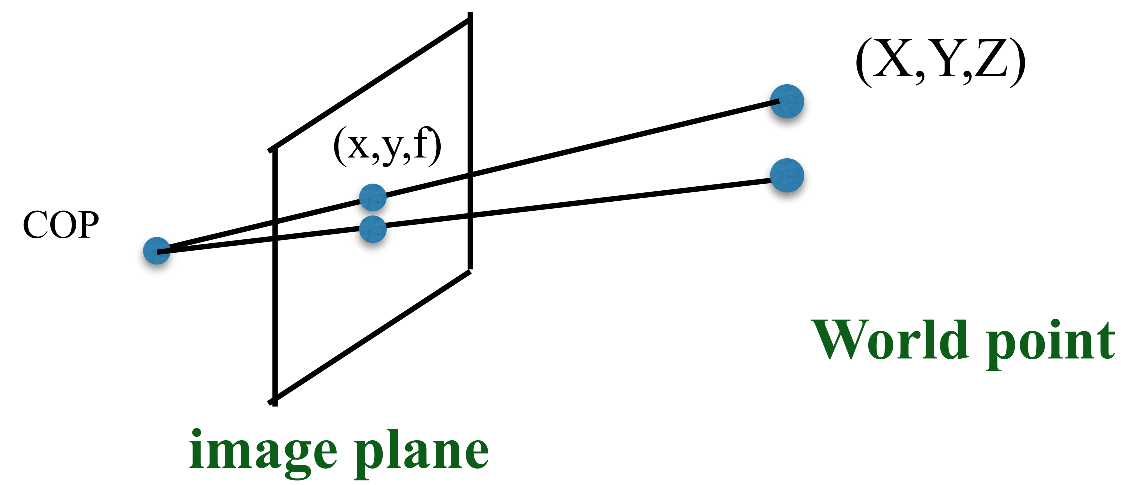




Different Camera Models

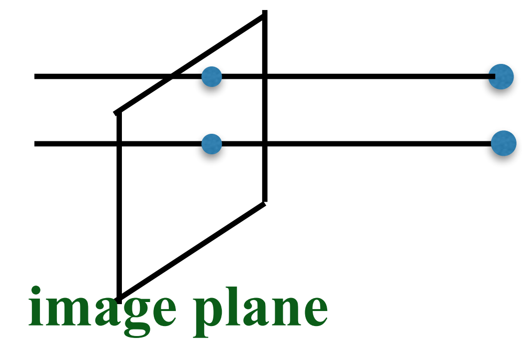
Perspective Projection
(most realistic)

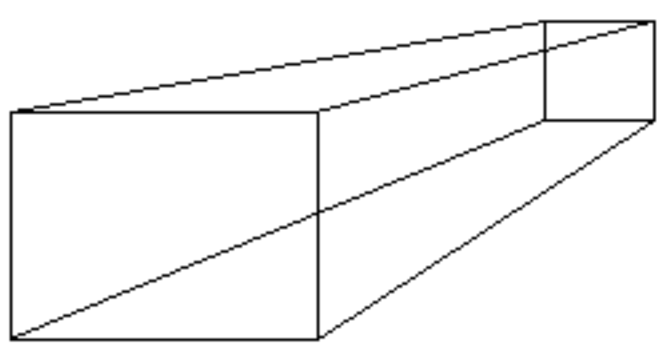
$x = fX/Z$
$y = fY/Z$



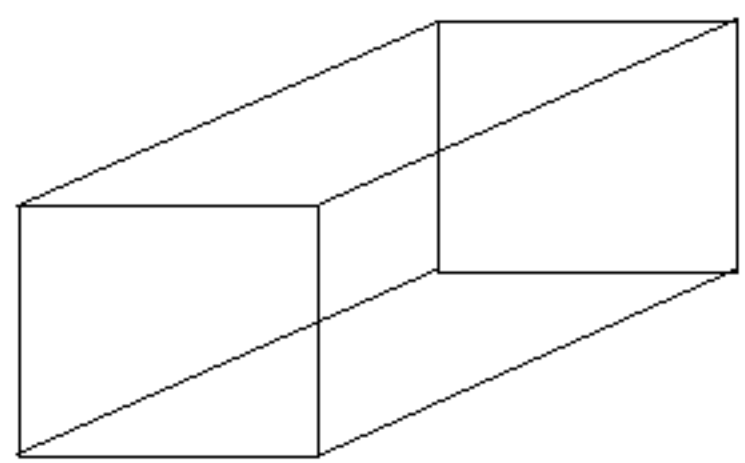
Orthographic Projection
(simplified, not realistic)

$x = X$
$y = Y$





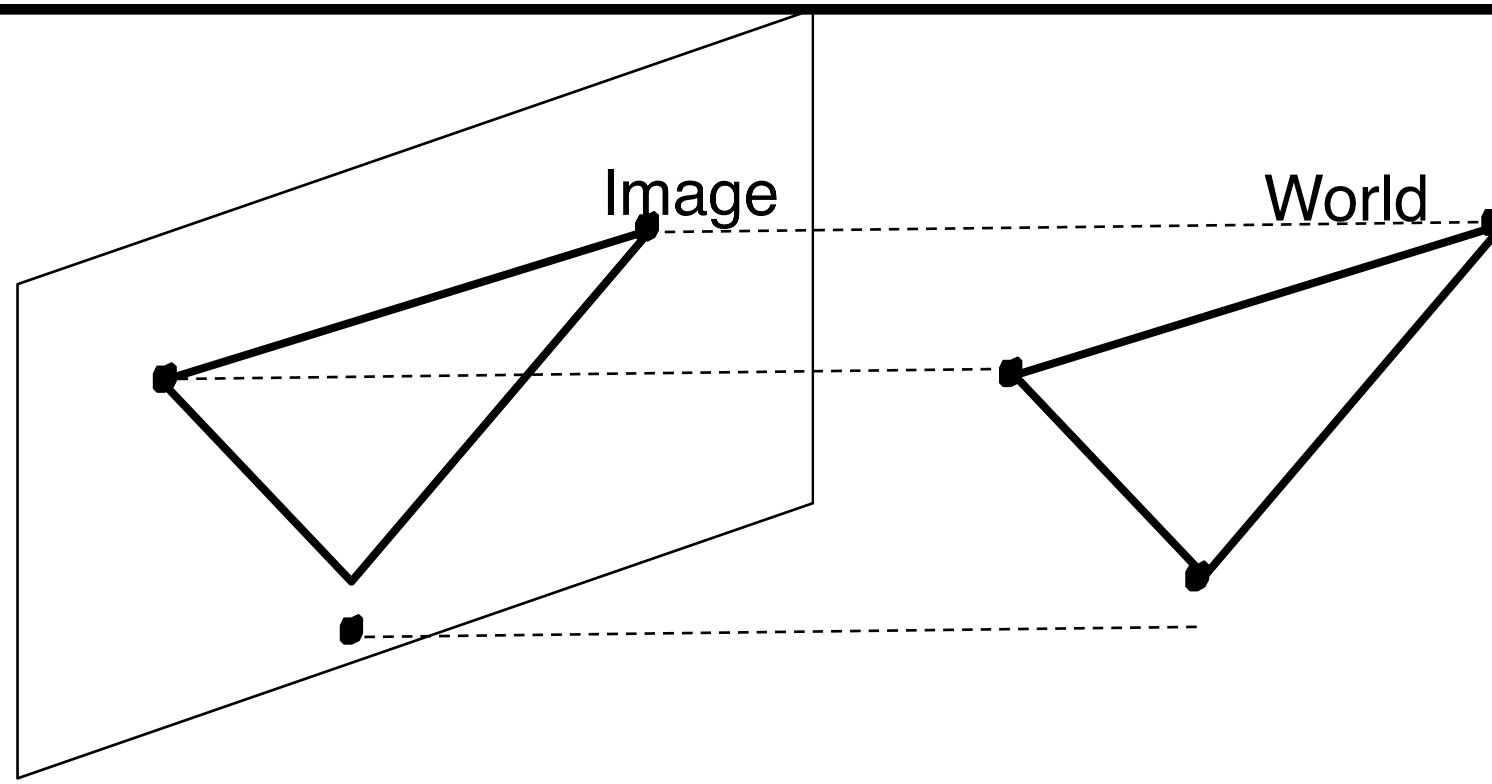
Perspective projection



Orthographic projection

Orthographic Projection: Angles and distances are preserved!

Orthographic Projection



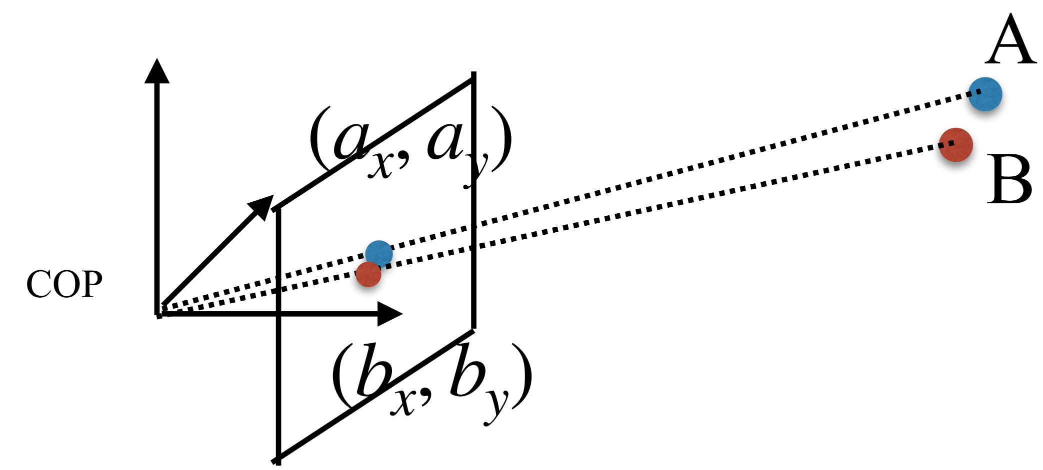
What does the camera projection matrix M_{3x4} look like for orthographic projection?

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow x = X, y = Y$$

Scaled orthographic / "weak perspective" projection

Consider two points (A,B) at different depths that are far away from camera:

$$\begin{bmatrix} A_x \\ A_y \\ Z \end{bmatrix} \qquad \begin{bmatrix} B_x \\ B_y \\ Z + \Delta Z \end{bmatrix}$$



if $\Delta Z \ll Z$, what happens to their image projections?

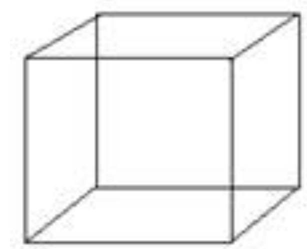
$$a_x = f\frac{A_x}{Z} \quad = \alpha A_x$$

$$b_x = f\frac{B_x}{Z + \Delta Z} \quad \approx f\frac{B_x}{Z} \quad = \alpha B_x$$

We can approximate projecting points at similar depths as multiplying by a scale factor $\frac{f}{Z}$

Scaled orthographic / “weak perspective” projection

Orthographic / scaled orthographic projection **is** reasonable if points are far away ($\Delta Z \ll Z$)



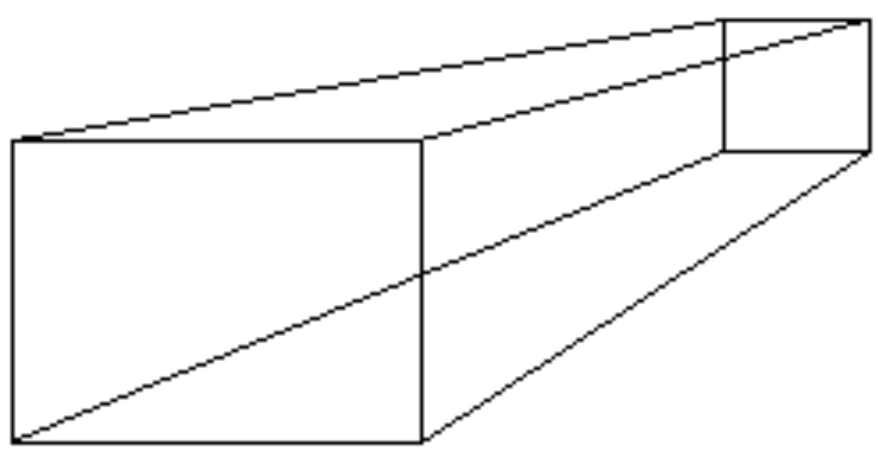
far

far

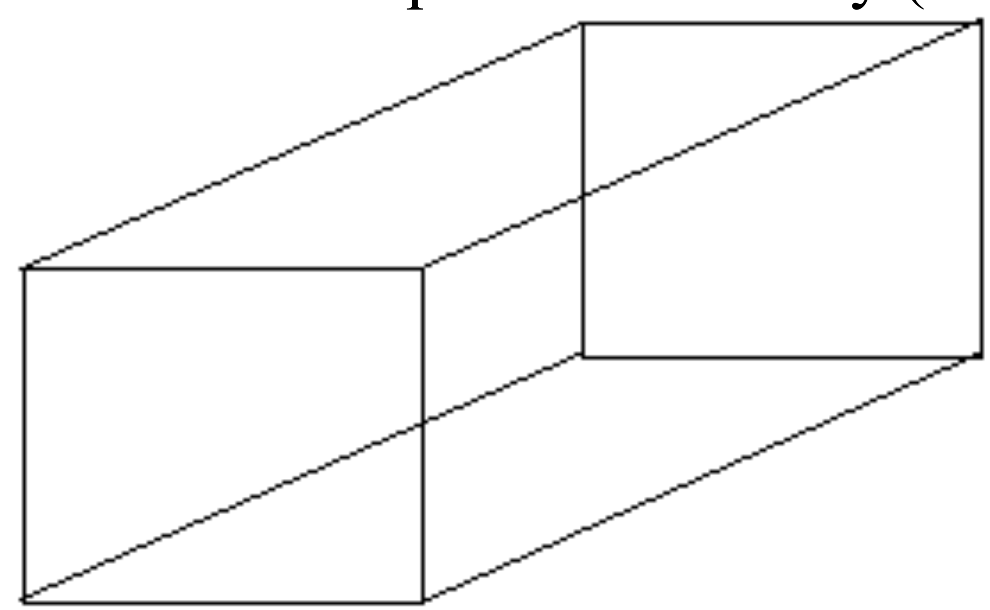
Orthographic / scaled orthographic projection **is not** reasonable if points are nearby ($\Delta Z \cancel{\ll} Z$)

far

close



Perspective projection

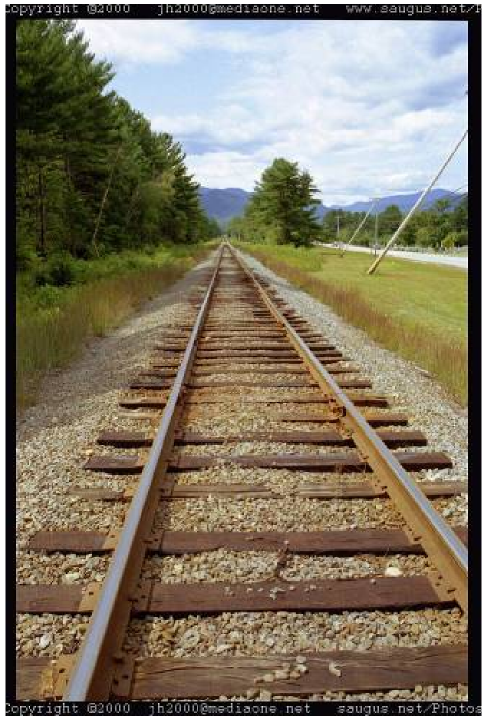


Orthographic projection

Scaled orthographic / "weak perspective" projection

Orthographic / scaled orthographic projection **is** reasonable if points are far away ($\Delta Z \ll Z$)

Orthographic / scaled orthographic projection **is not** reasonable if $\Delta Z \cancel{\ll} Z$



Perspective vs Orthogrpahic

Wide angle | Standard | Telephoto

(Subject is close to camera - perspective effects)

(Subject is farther from camera - orthographic effects)

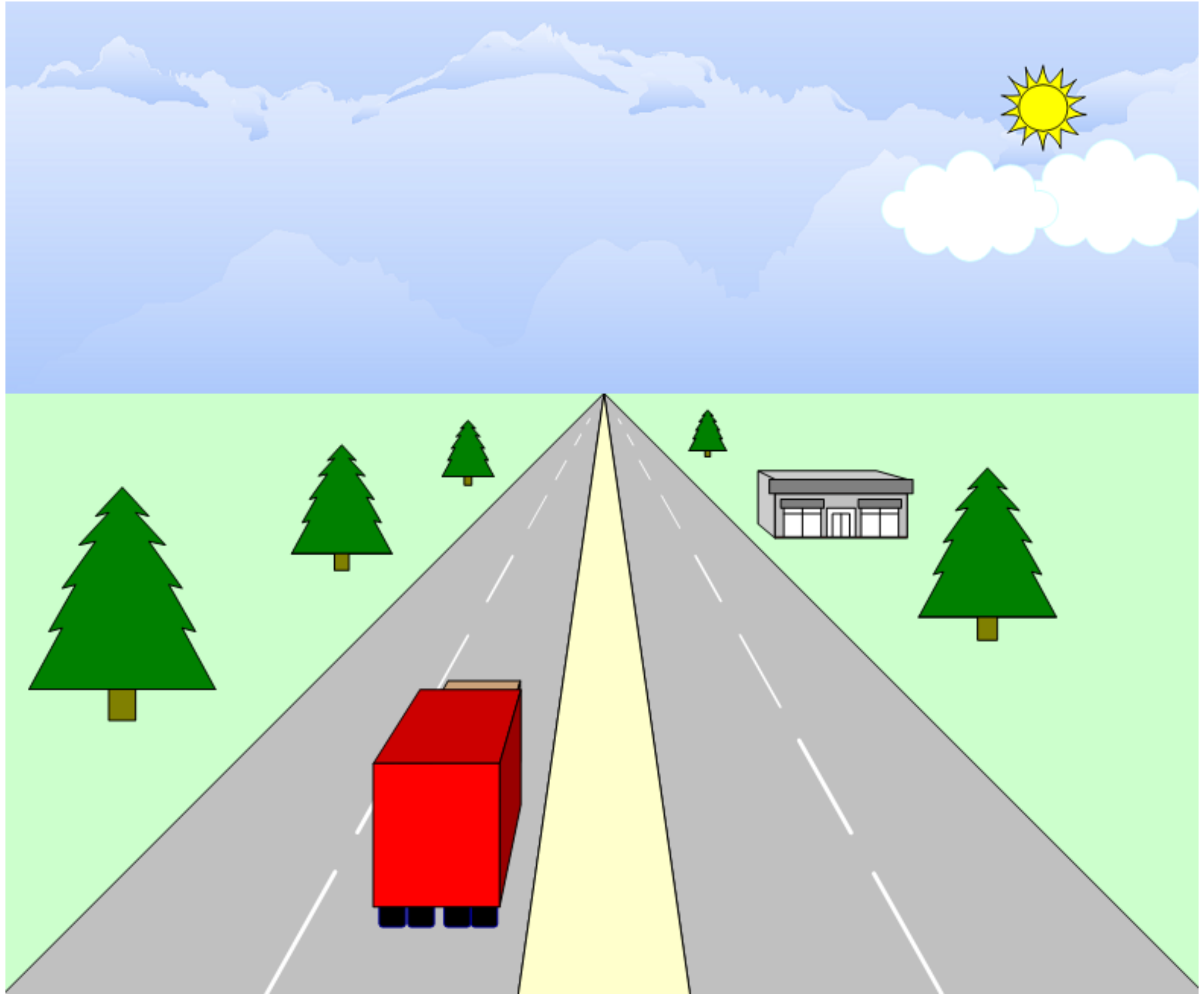
Scaled orthographic is a reasonable approximation for many realistic photos

Objects at image border are viewed from a different angle
… but scaled orthographic model doesn't capture this! (just scales the object)

Shocking implication:
we don't actually want our object detectors to be translation invariant!

(Computer vision methods do not typically take this into account)

Review: Perspective projection



Closer objects appear larger

Closer objects are lower in the image

Parallel lines meet

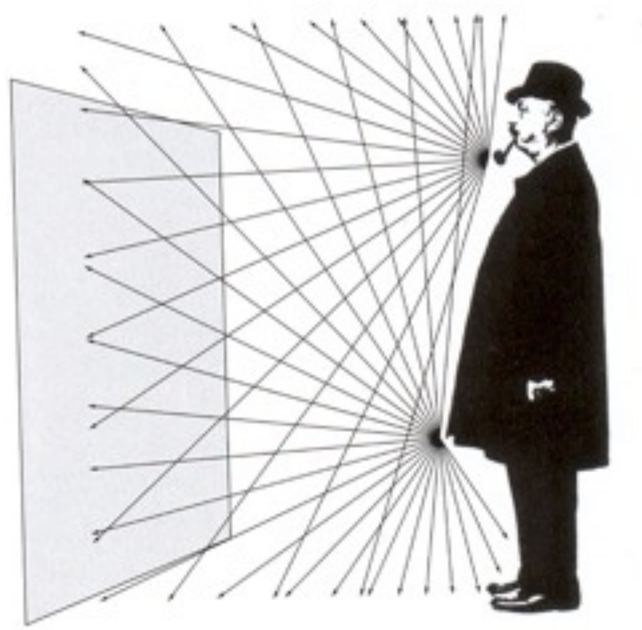
All these can be simply derived with $x = X\frac{f}{Z} \quad y = Y\frac{f}{Z}$

Summary

- Pinhole optics
  - Perspective projection (vanishing points, horizon, object height)
  - Projection models (orthographic, scaled orthographic, paraperspective)

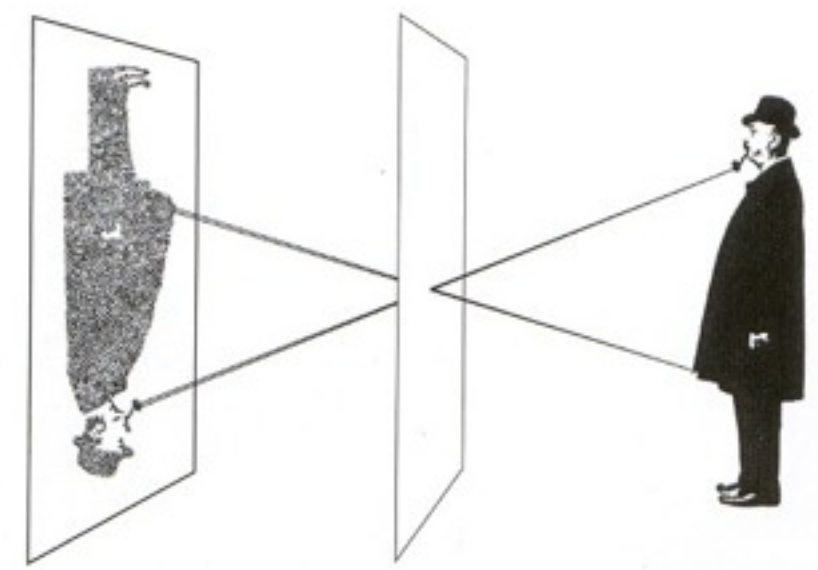
Pinhole optics

A bad camera



With a large pinhole, images will be blurry

A better camera



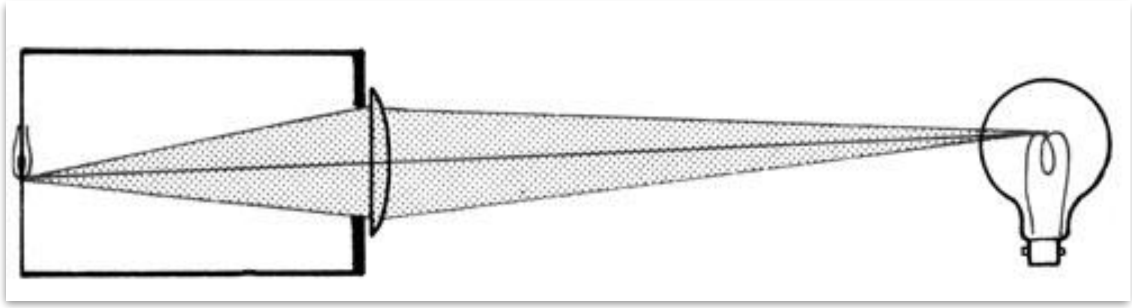
pinhole

- With a very small pinhole, not enough light will get through
- We will need to keep the pinhole open a long time to get enough light for the image (camera exposure)

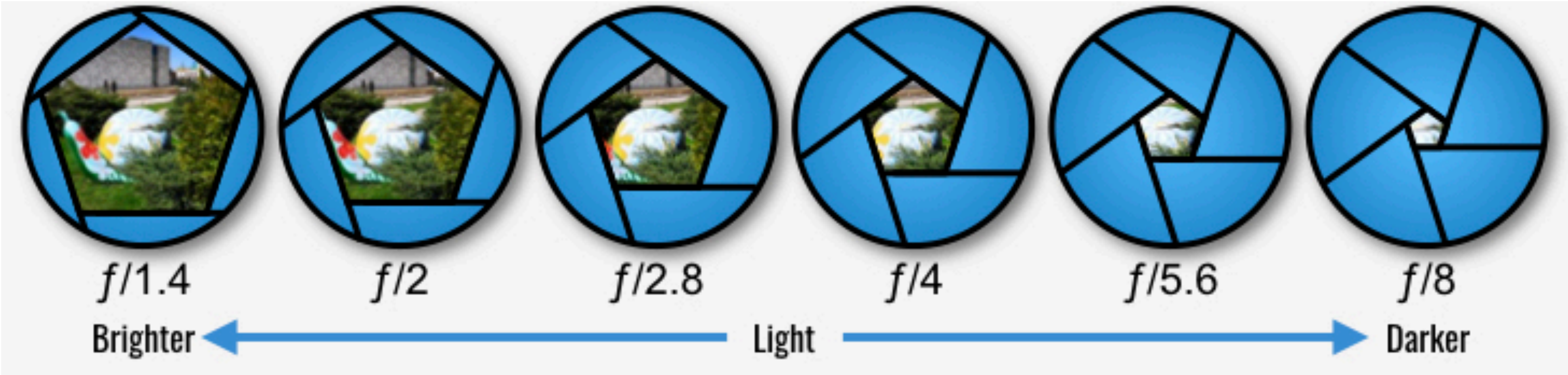
How do we get a sharp image with a small exposure time?

Depth of Field

Use a lens to grab more photons
Allows for a larger pinhole; lens focuses the light
from a particular depth



Most cameras can adjust the size of the pinhole ("aperture"):

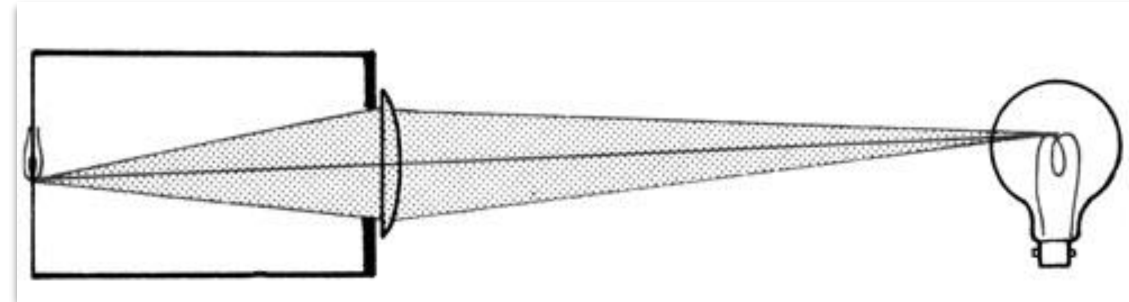


But why would we ever want a smaller pinhole if we can just use a larger pinhole and a lens?

Cons of a larger pinhole:
- Depth of field
- Radial Distortion

Depth of Field

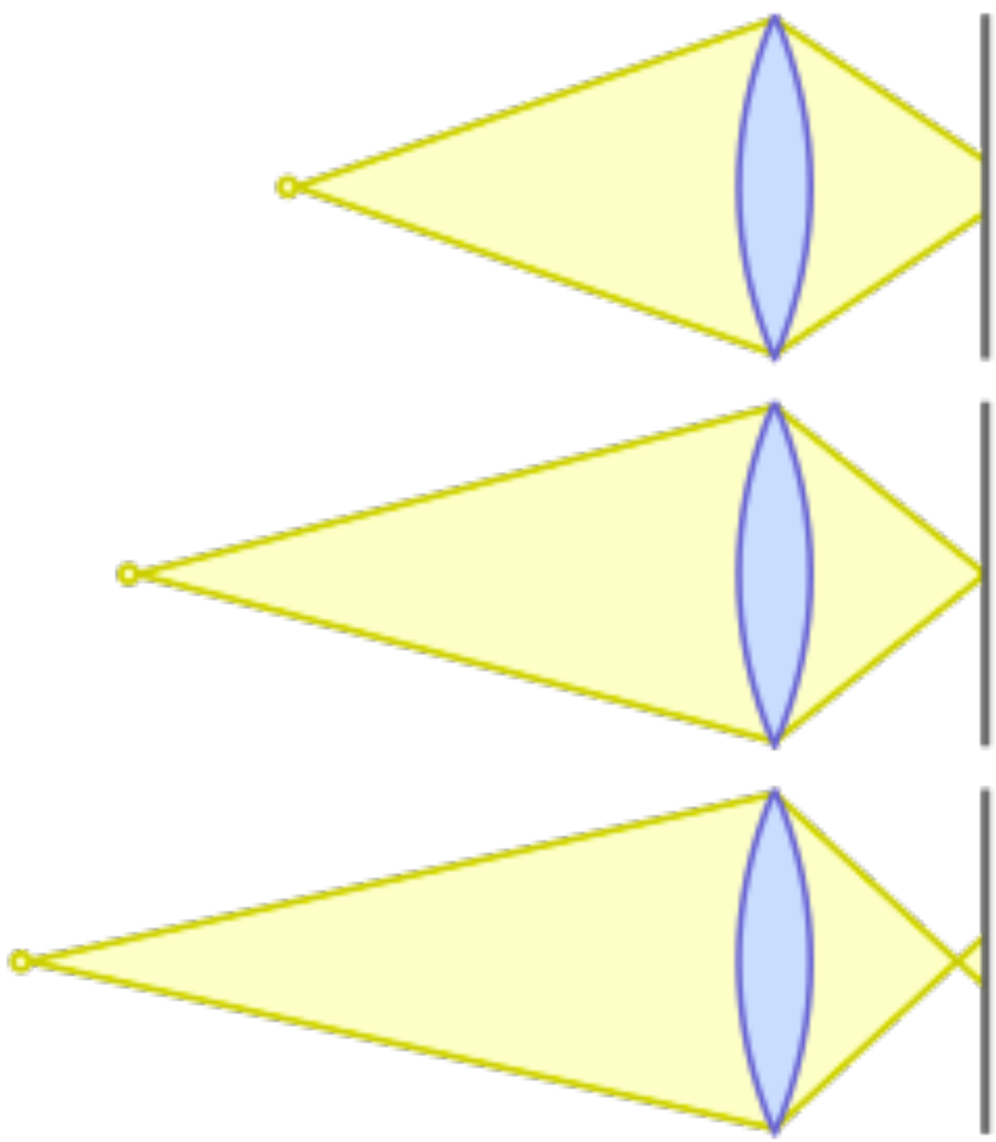
Use a lens to grab more photons
Allows for a larger pinhole; lens focuses the light
from a particular depth



With a lens, objects outside of a particular depth will be blurred:

World point Lens Image plane

Camera image



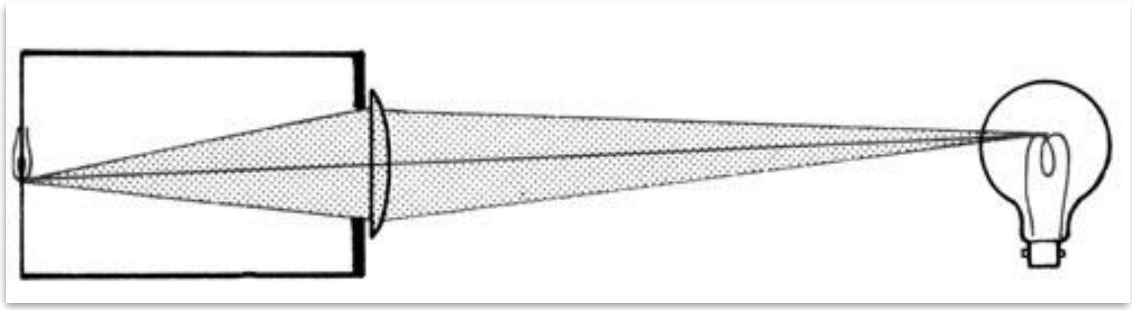
Focus at a
“circle of confusion”

Focus at a
single point

Focus at a
“circle of confusion”

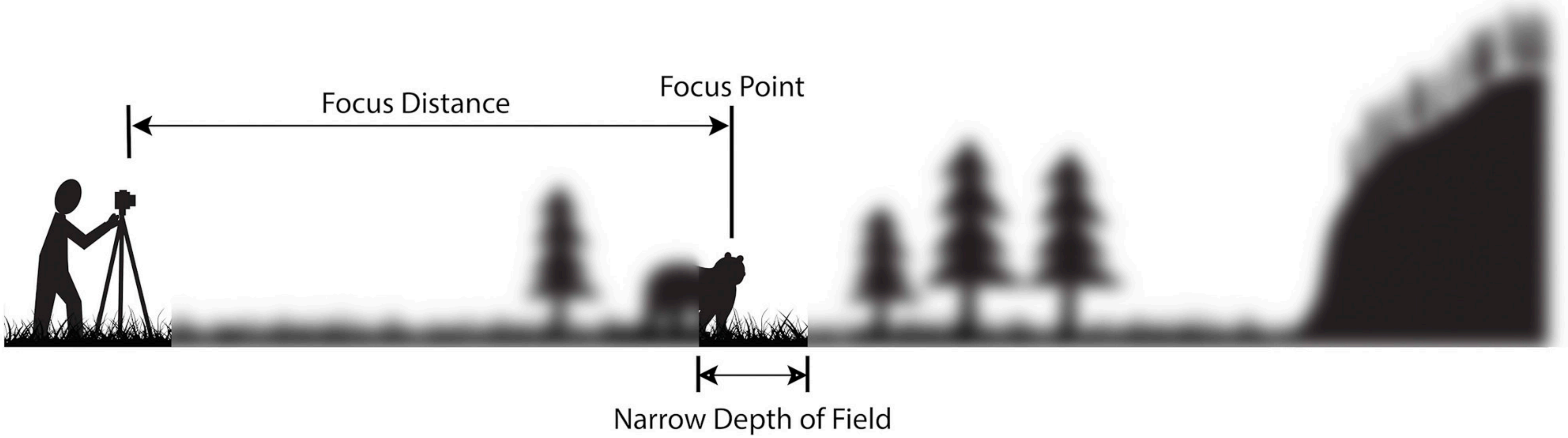
Depth of Field

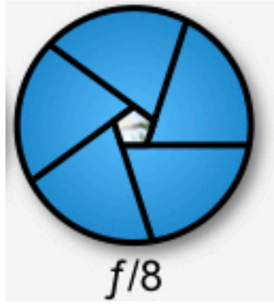
Use a lens to grab more photons
Allows for a larger pinhole; lens focuses the light from a particular depth



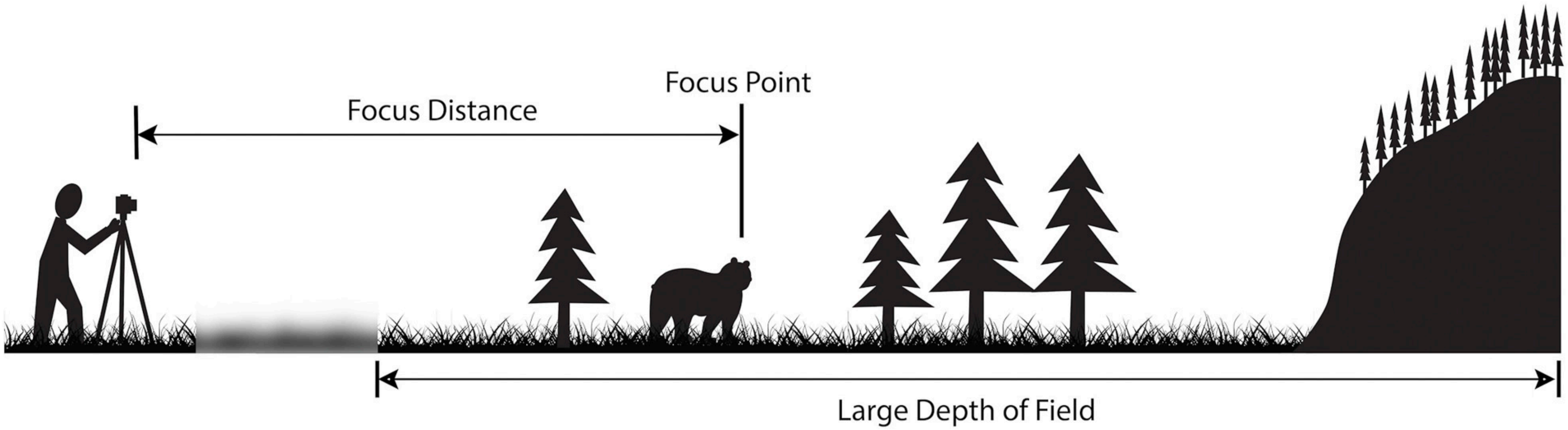
Objects outside the particular depth will be blurred (limited "depth-of-field")

Larger aperture

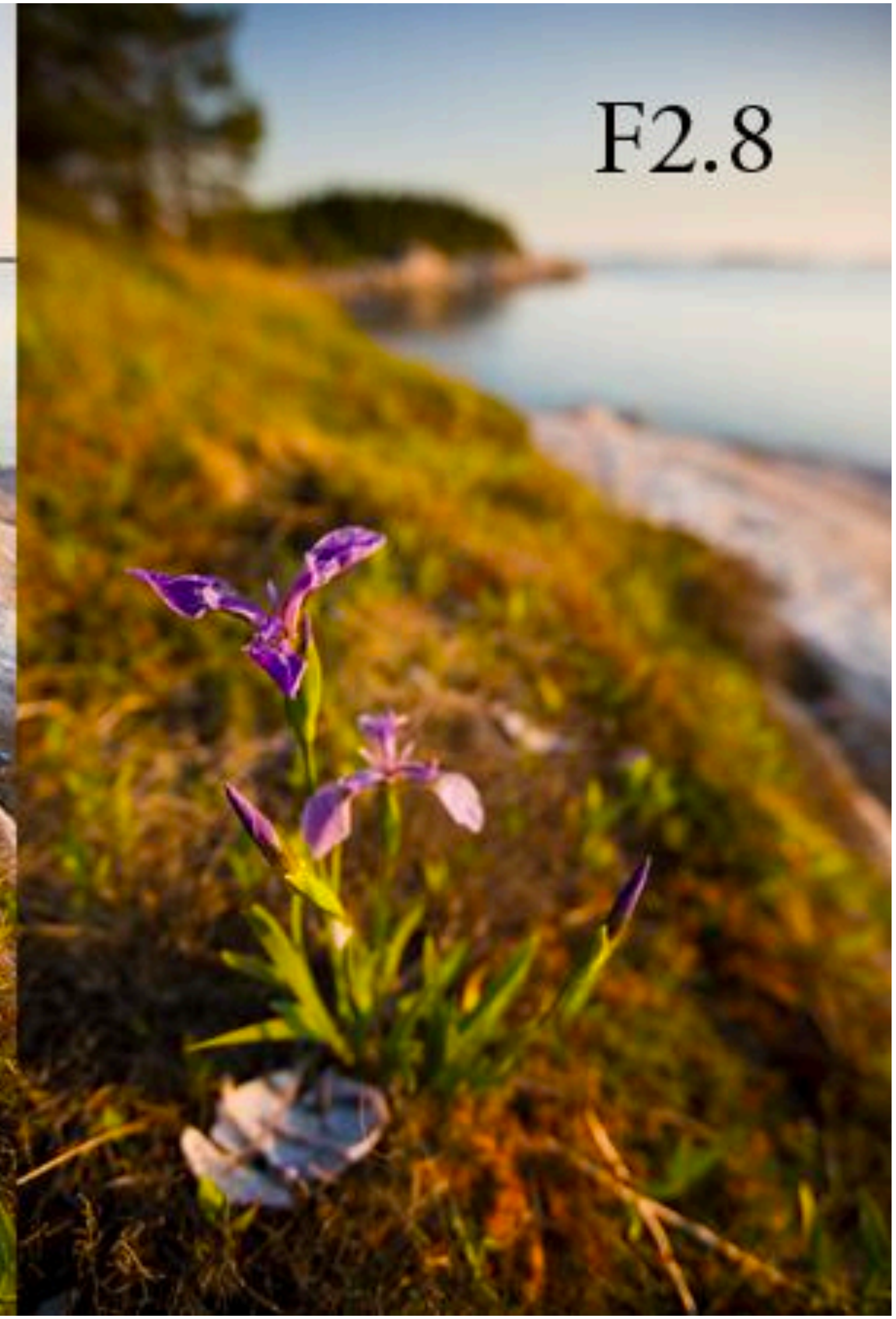




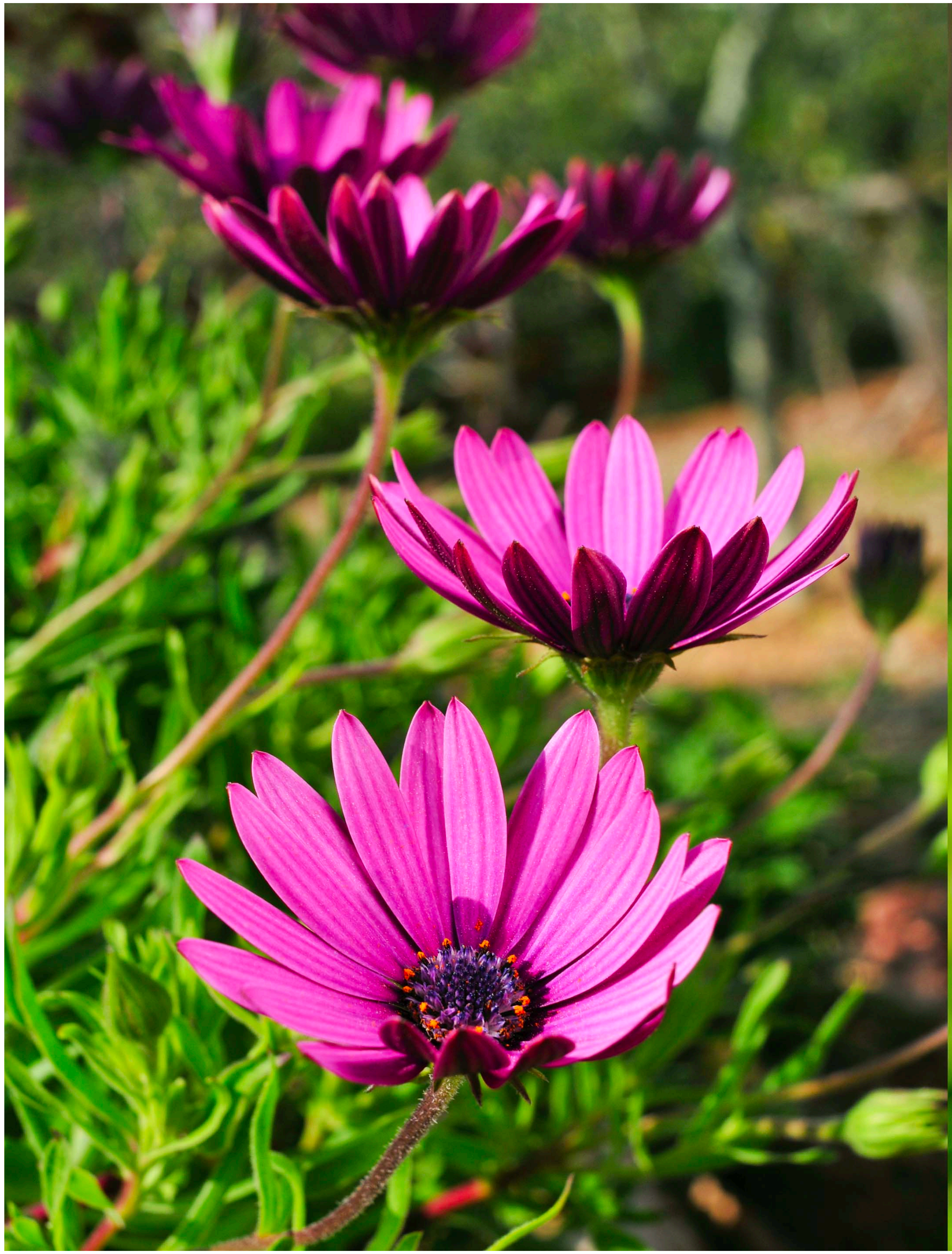
Smaller aperture



Small Aperture
Large Depth of Field



Large Aperture
Narrow Depth of Field



Small Aperture
Large Depth of Field

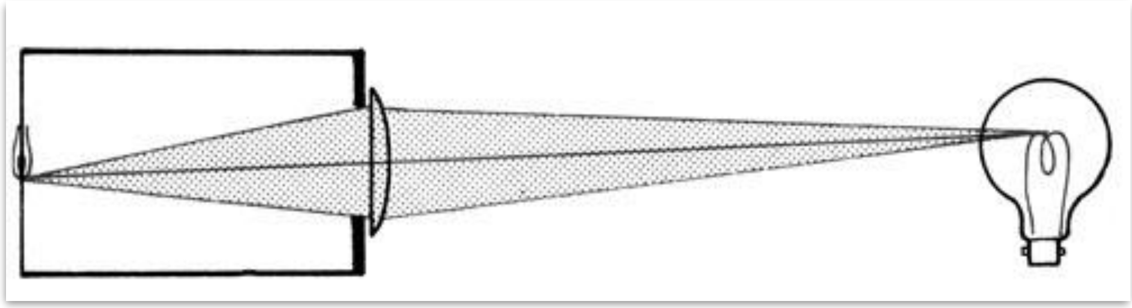
Large Aperture
Narrow Depth of Field

Small Aperture
Large Depth of Field

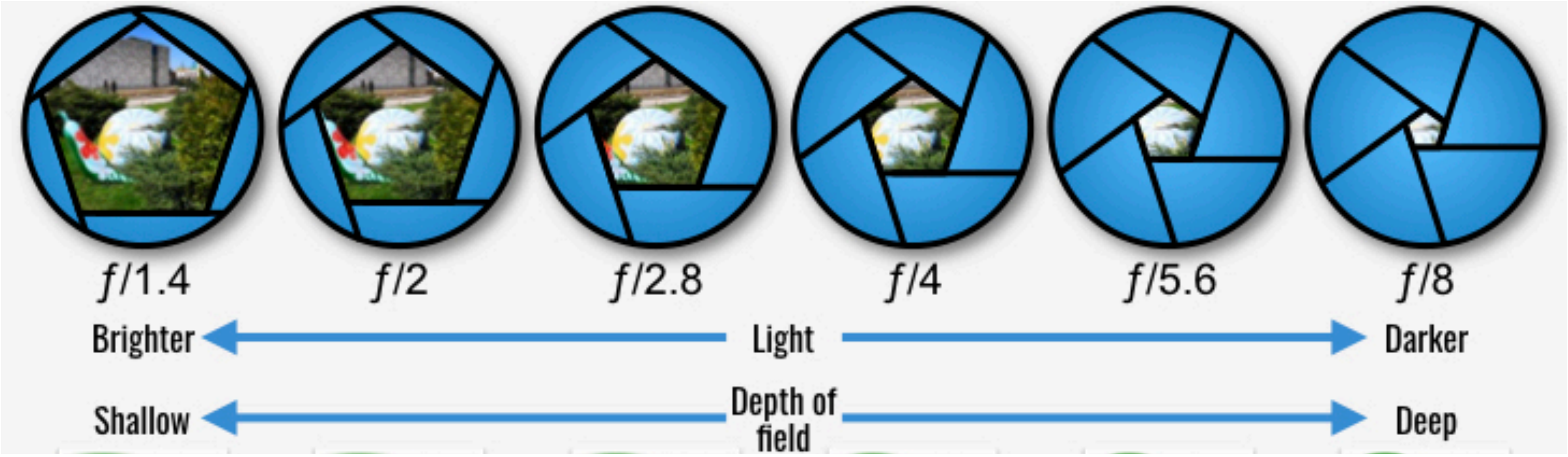
Large Aperture
Narrow Depth of Field

Depth of Field

Use a lens to grab more photons
Allows for a larger pinhole; lens focuses the light from a particular depth

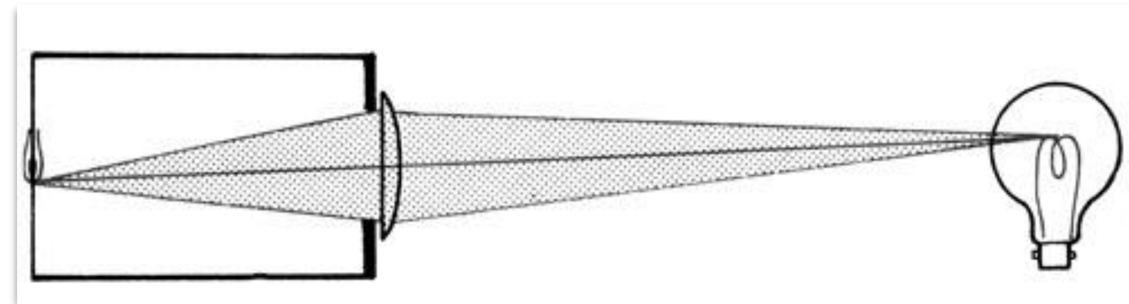


Larger aperture -> larger depth of field



Depth of Field

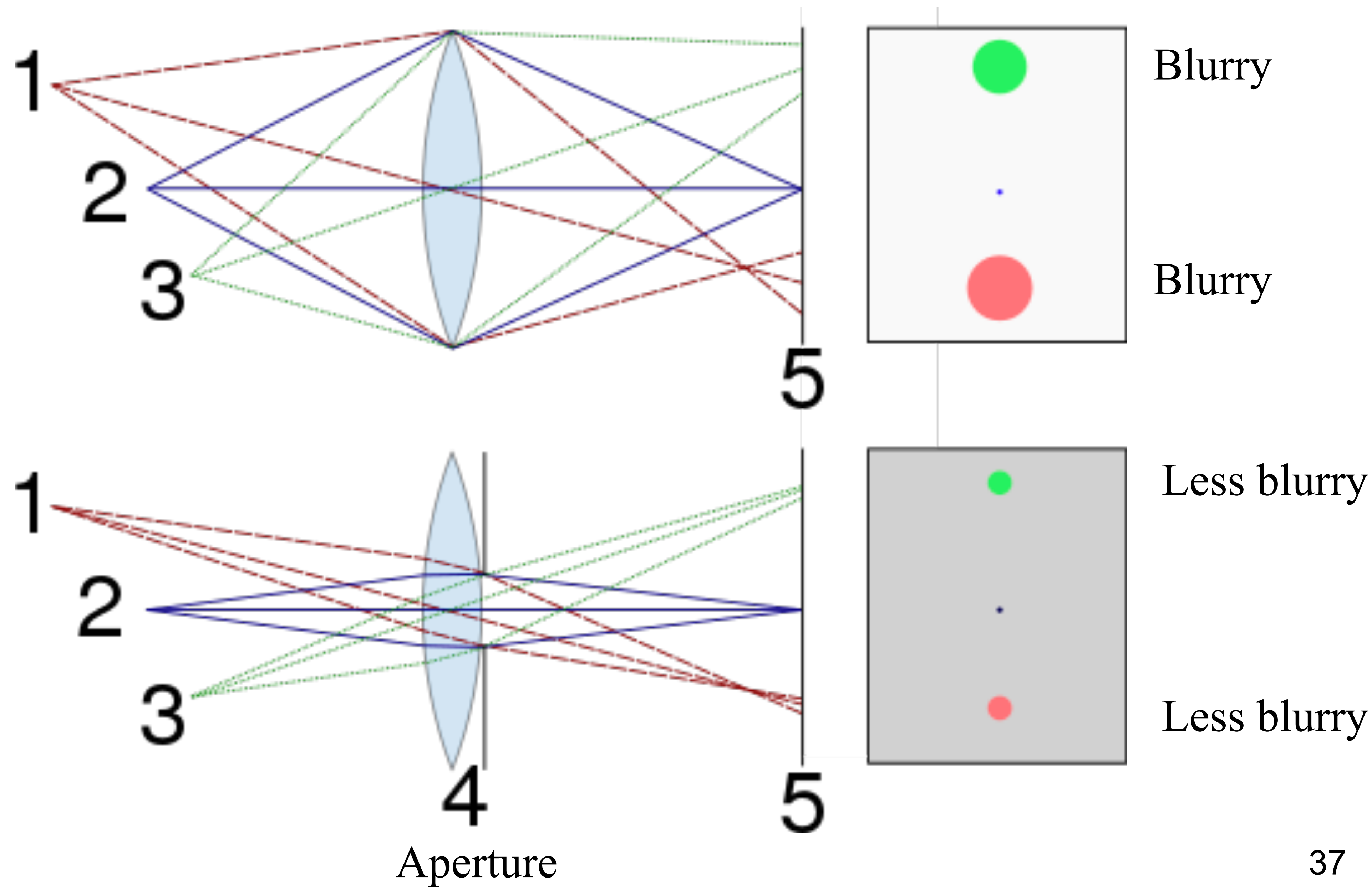
Use a lens to grab more photons
Allows for a larger pinhole; lens focuses the light
from a particular depth



Larger aperture -> larger depth of field

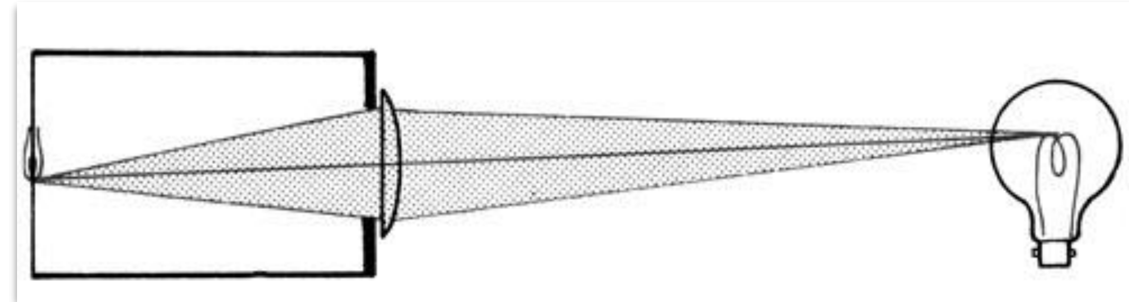
Larger aperture
Small Depth of field

Small aperture
Large Depth of field

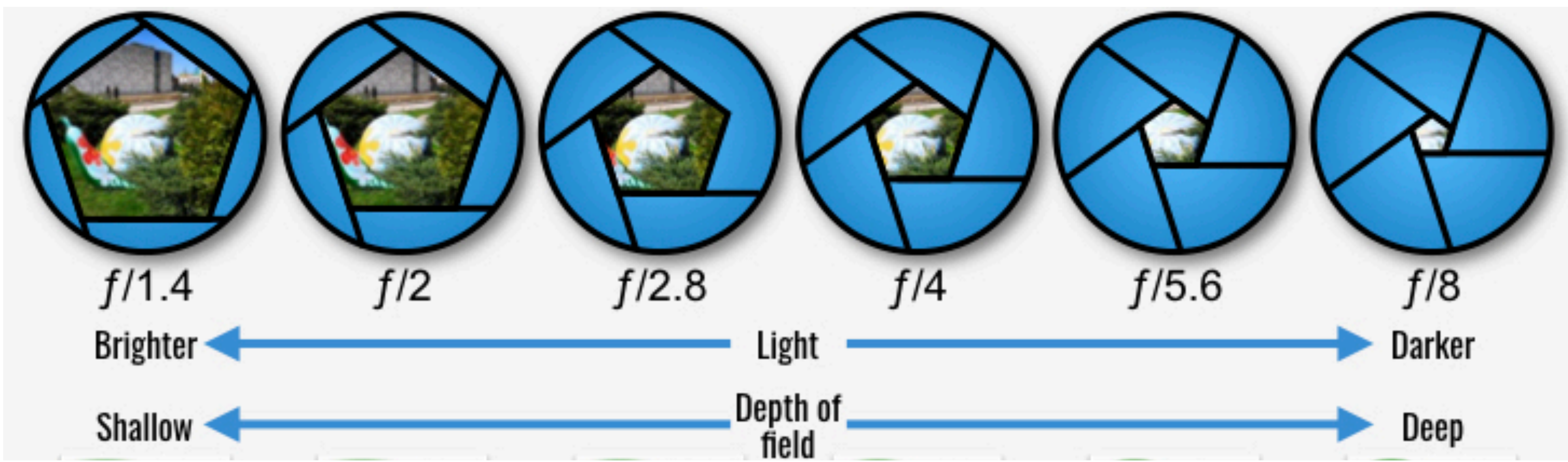


Depth of Field

Use a lens to grab more photons
Allows for a larger pinhole; lens focuses the light
from a particular depth



Larger aperture -> larger depth of field



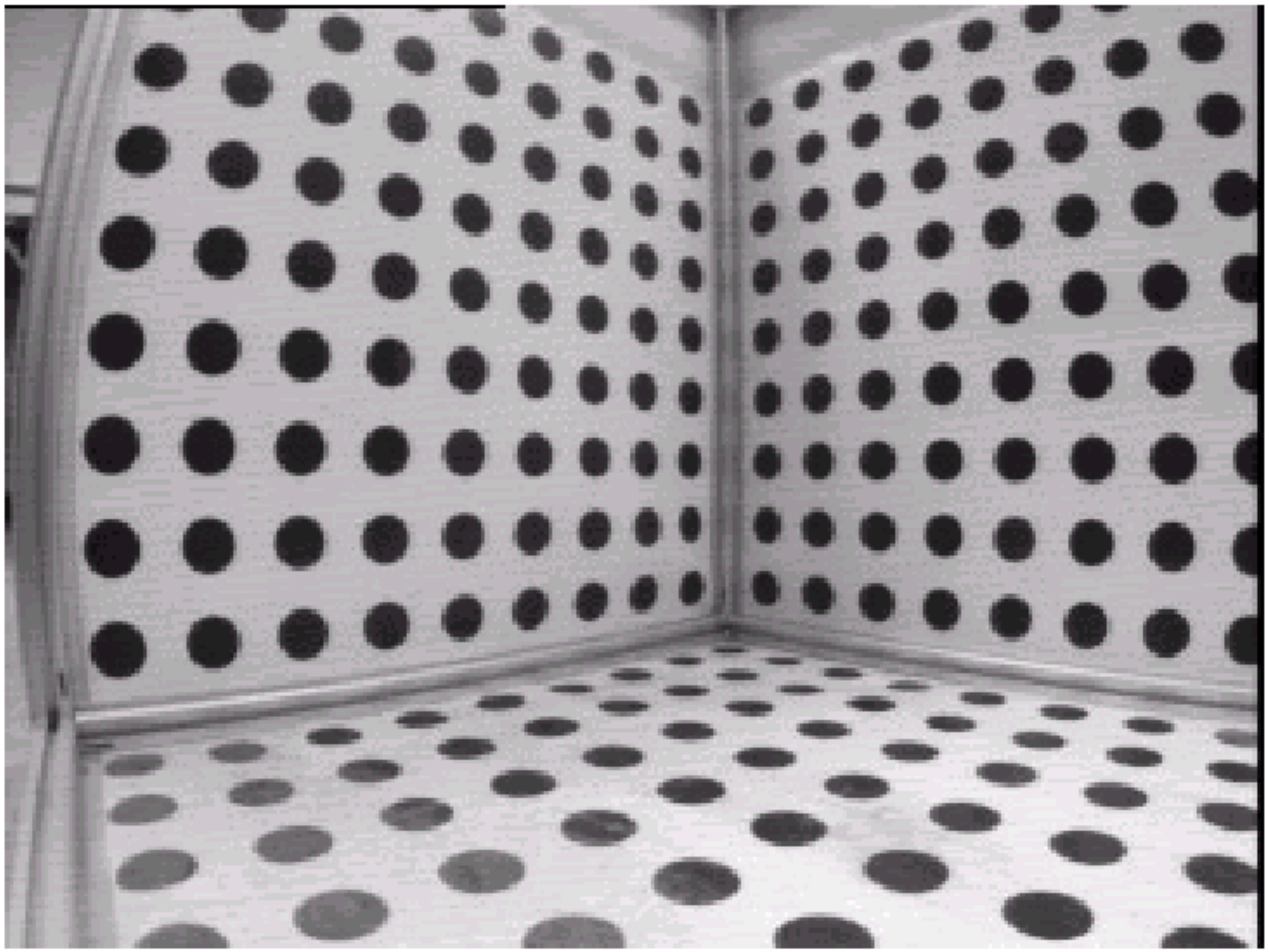
Less in focus -> Bad for computer vision

Brightest -> Good for computer vision

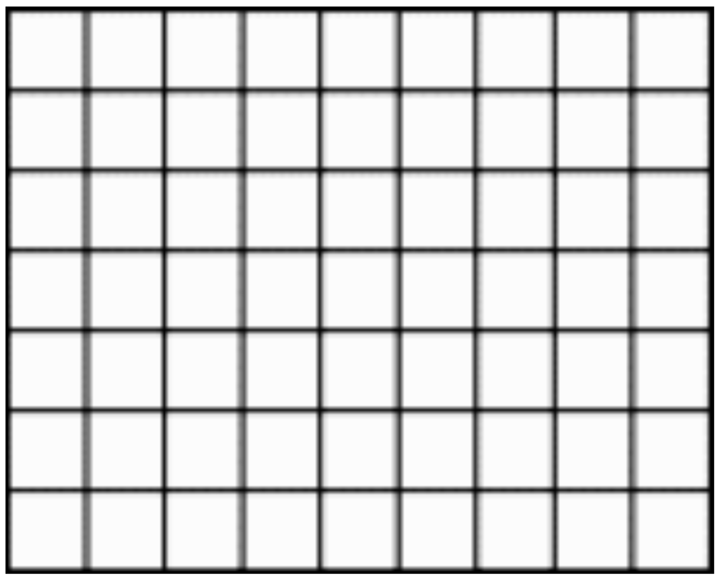
More in focus -> Good for computer vision

Darkest -> Bad for computer vision
(Low Signal-to-Noise Ratio)

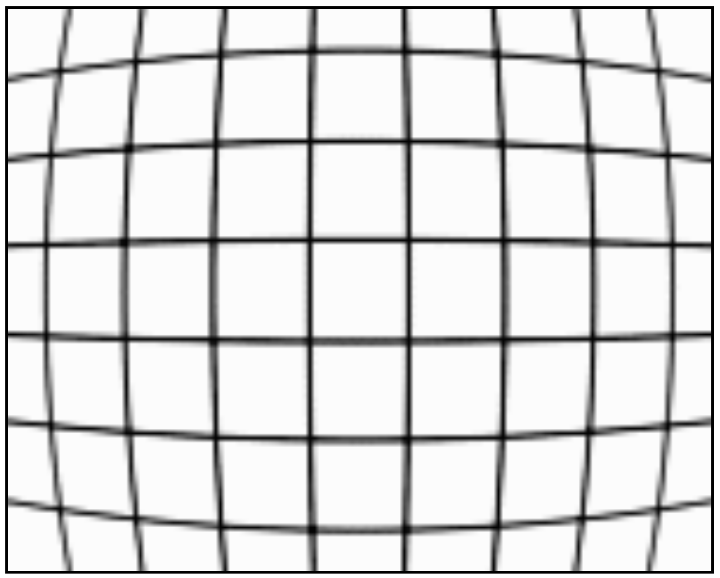
Lenses also create Radial Distortions



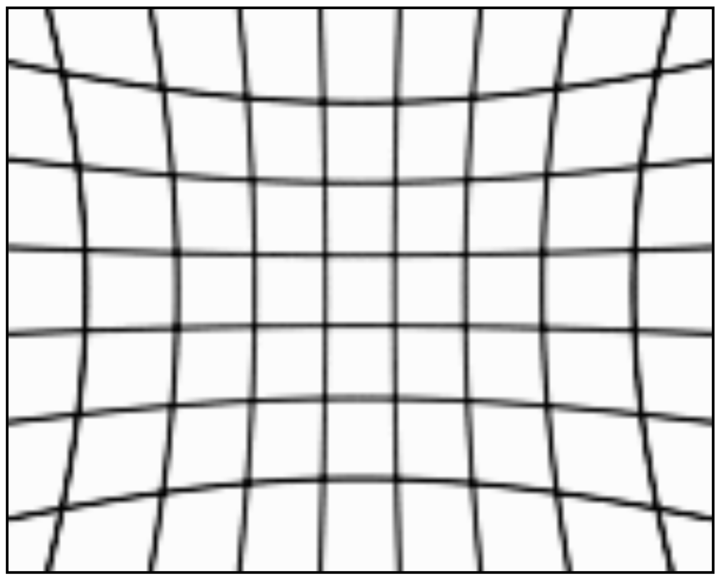
Radial Lens Distortions



No Distortion

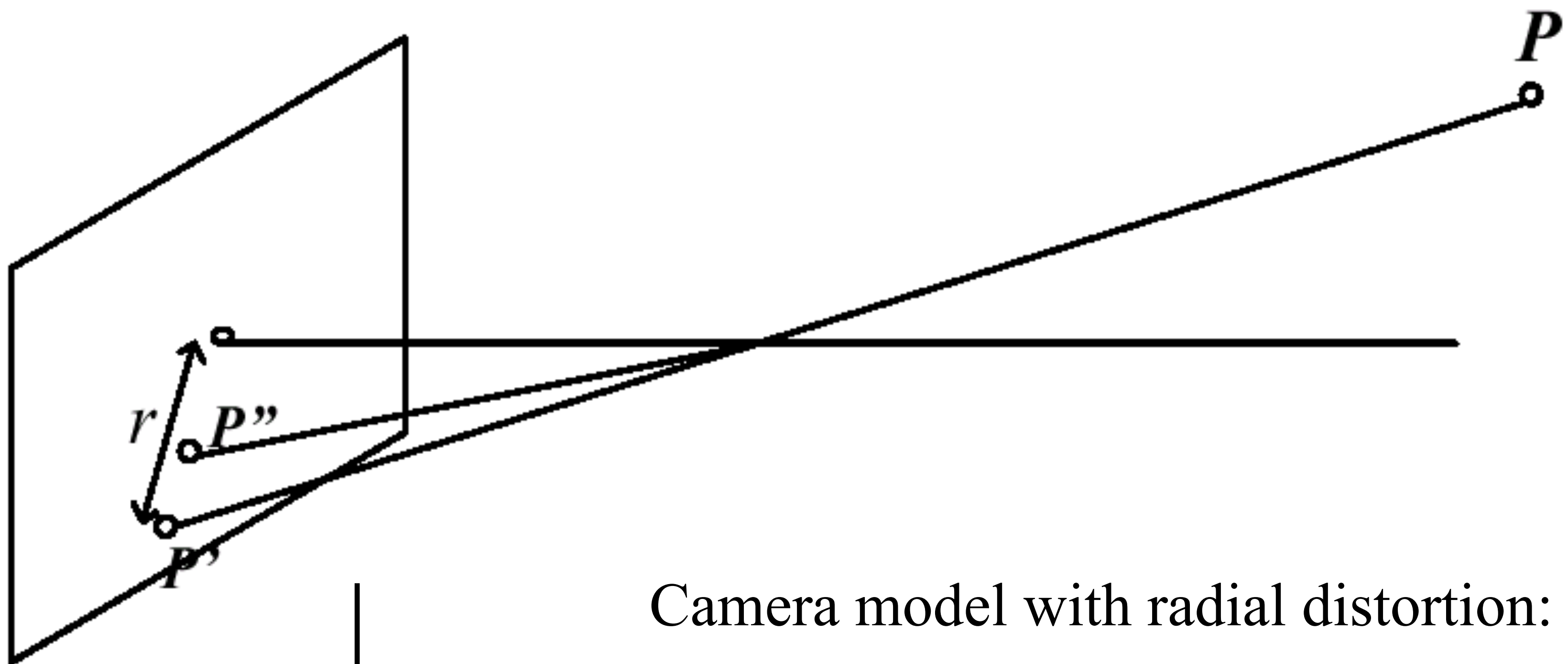


Barrel Distortion



Pincushion Distortion

Radial Distortion Model



Camera model with radial distortion:

Regular camera projection

$$x' = f\,\frac{x}{z}$$

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

Scale image coordinates by distortion factor:

$$x'' = \frac{1}{\lambda}x'$$

$$y'' = \frac{1}{\lambda}y'$$

$$\lambda = 1 + k_1 r^2 + k_2 r^4 + \cdots$$

r = distance of point from center of image

Use corresponding pairs of points (real world, image coordinates) to estimate k (with other camera parameters)

We can then invert these distortion scaling to “undistort” the image

Correcting Radial Lens Distortions

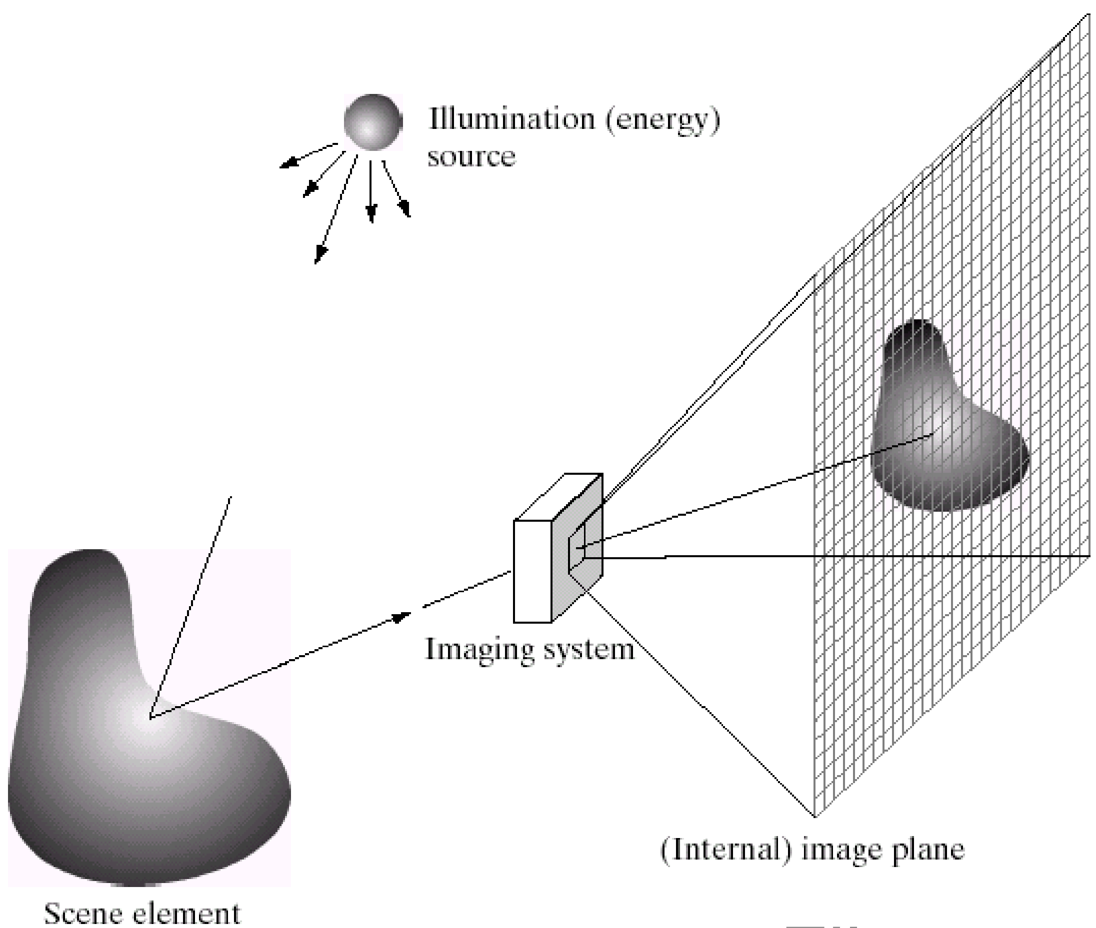
Before

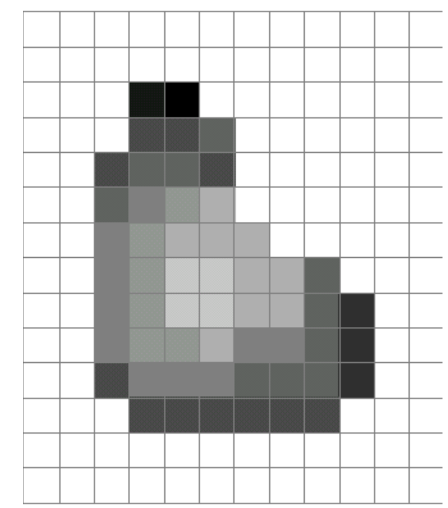
After

http://www.grasshopperonline.com/barrel_distortion_correction_software.html

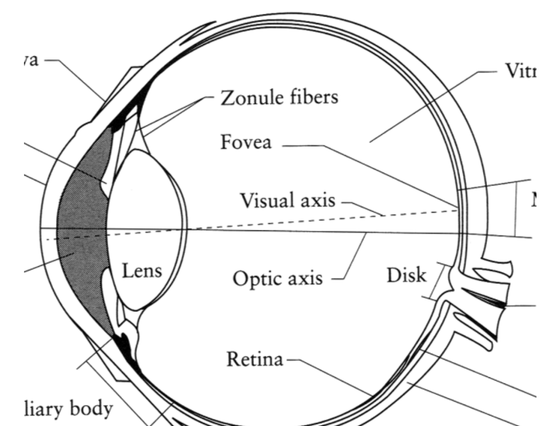
Misc

Image formation



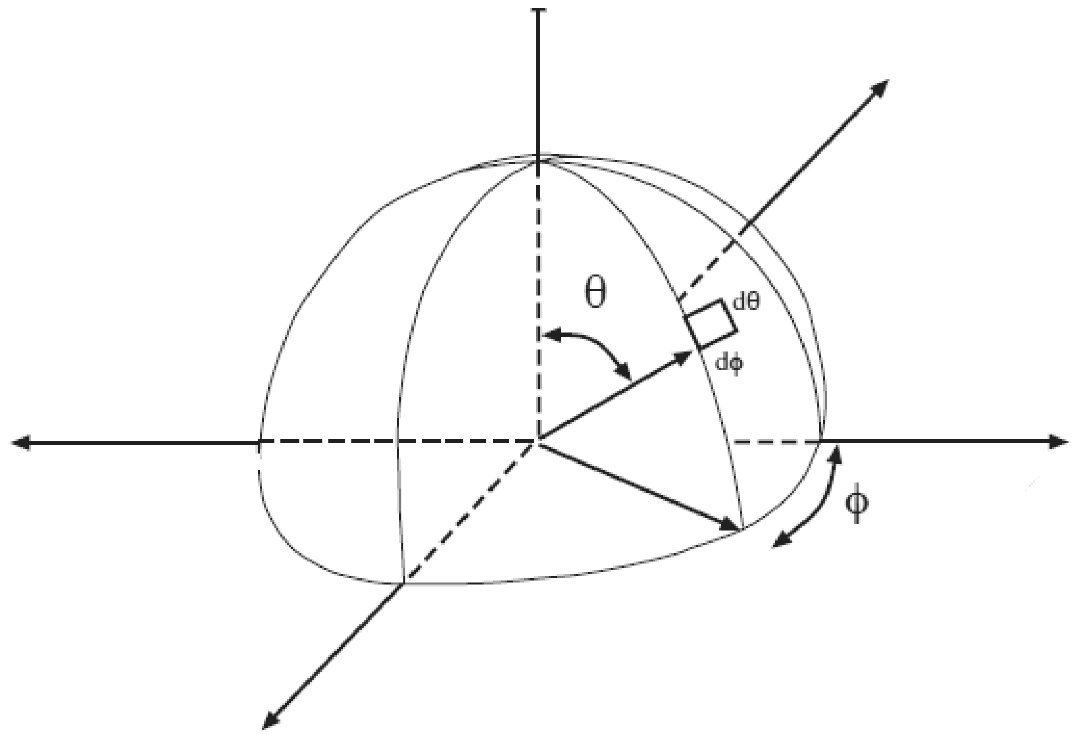


Digital Image



Human eye

Pixel brightness



Depends on the physical properties of the surface that light is reflecting off of (BRDF)

Extreme examples:

- Mirror
- Perfectly black surface
  - Actual material: "**Vantablack**" absorbs 99.96% of all light that falls on it

(More on "light as physics" at end of semester)

Pinhole optics

A bad camera

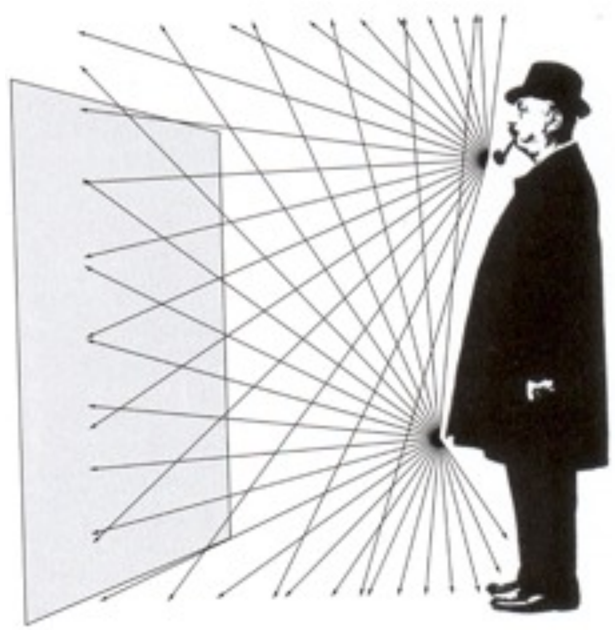


Image
plane
(pixels)

A better camera

Image is
upside down

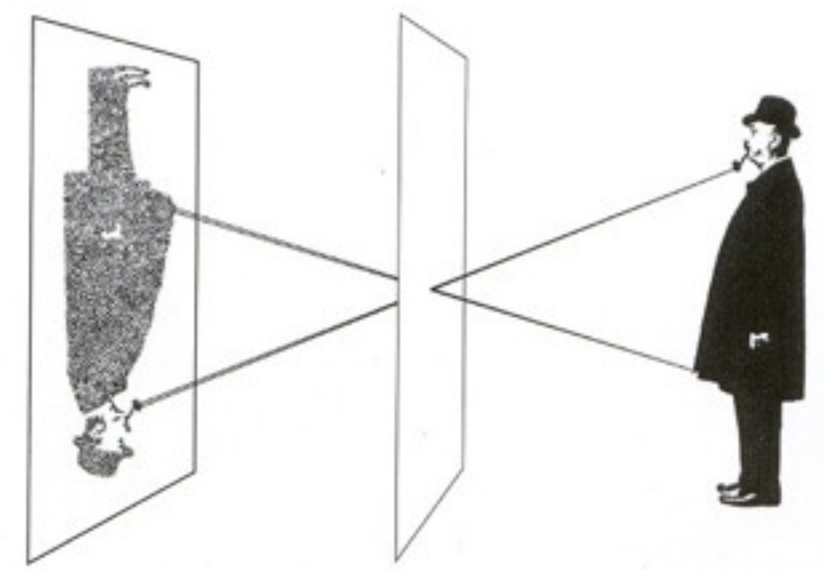
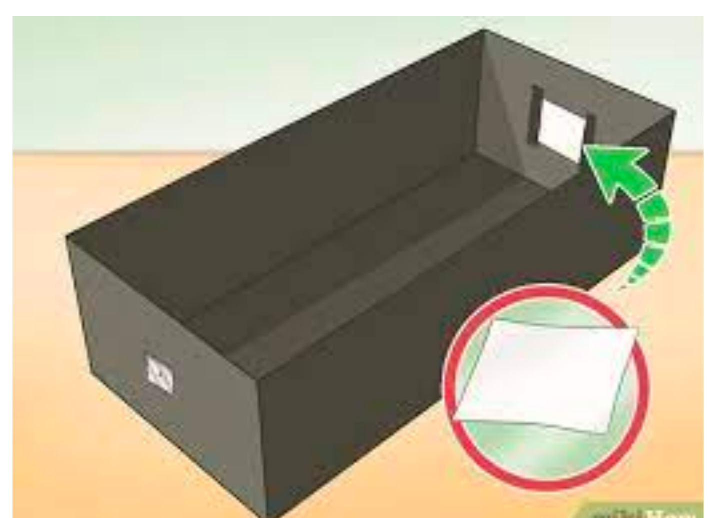


Image
plane
(pixels)

pinhole

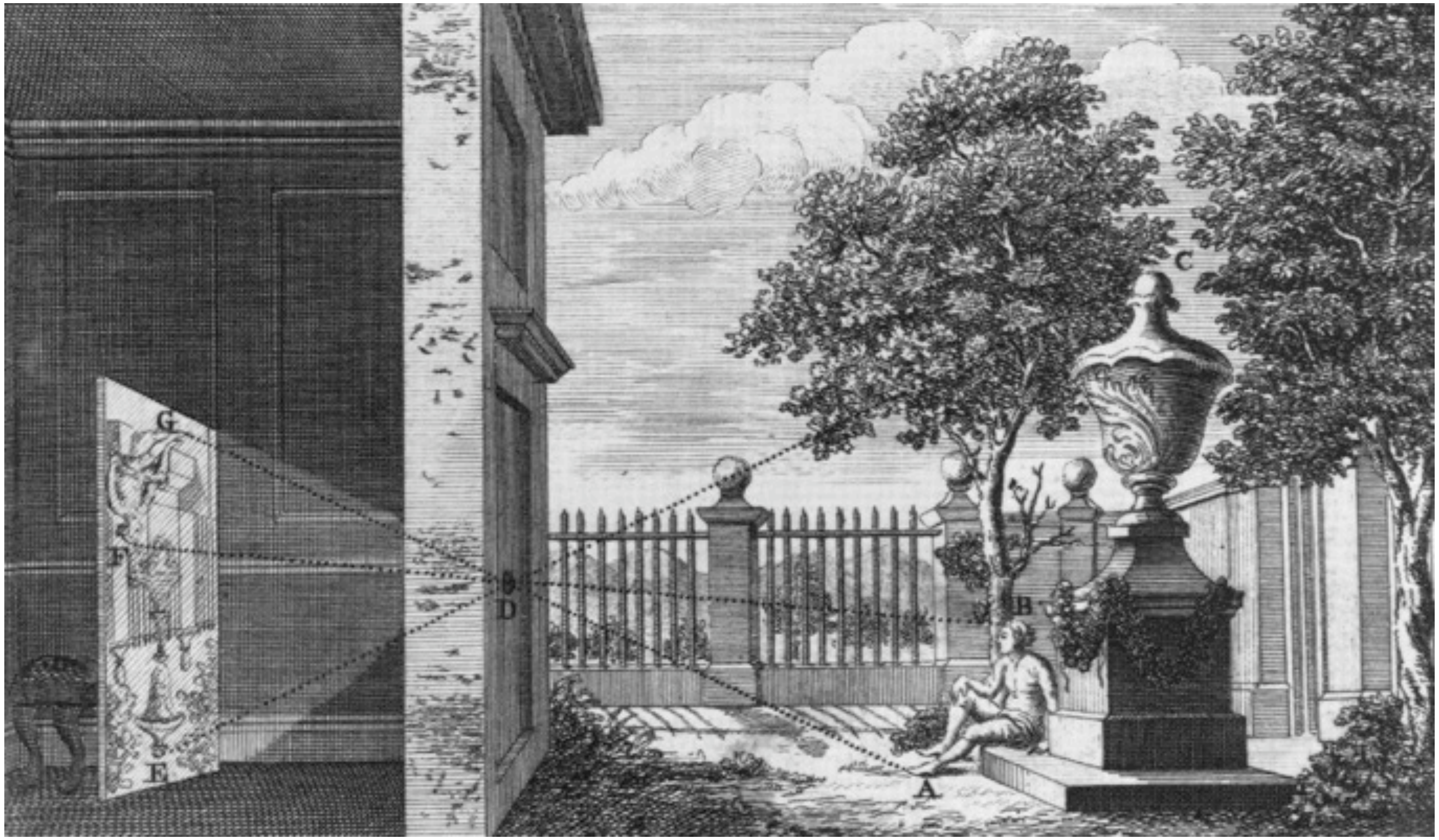


You can try this at home with a shoebox!

World's largest photograph

El Toro Marine Corps, Irvine CA 2006

Camera Obscura

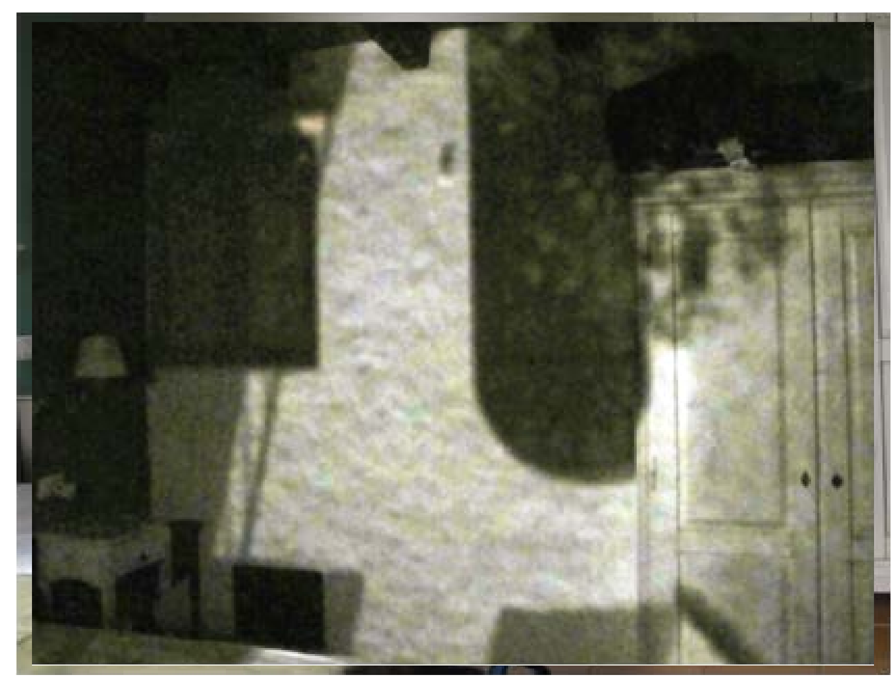


So why don't we see pinhole images in our daily lives?

Accidental pinholes

what's the dark stuff?

(the view from Antonio Torralba's hotel room)



Accidental pinhole and pinspeck cameras: revealing the scene outside the picture

CVPR 2012

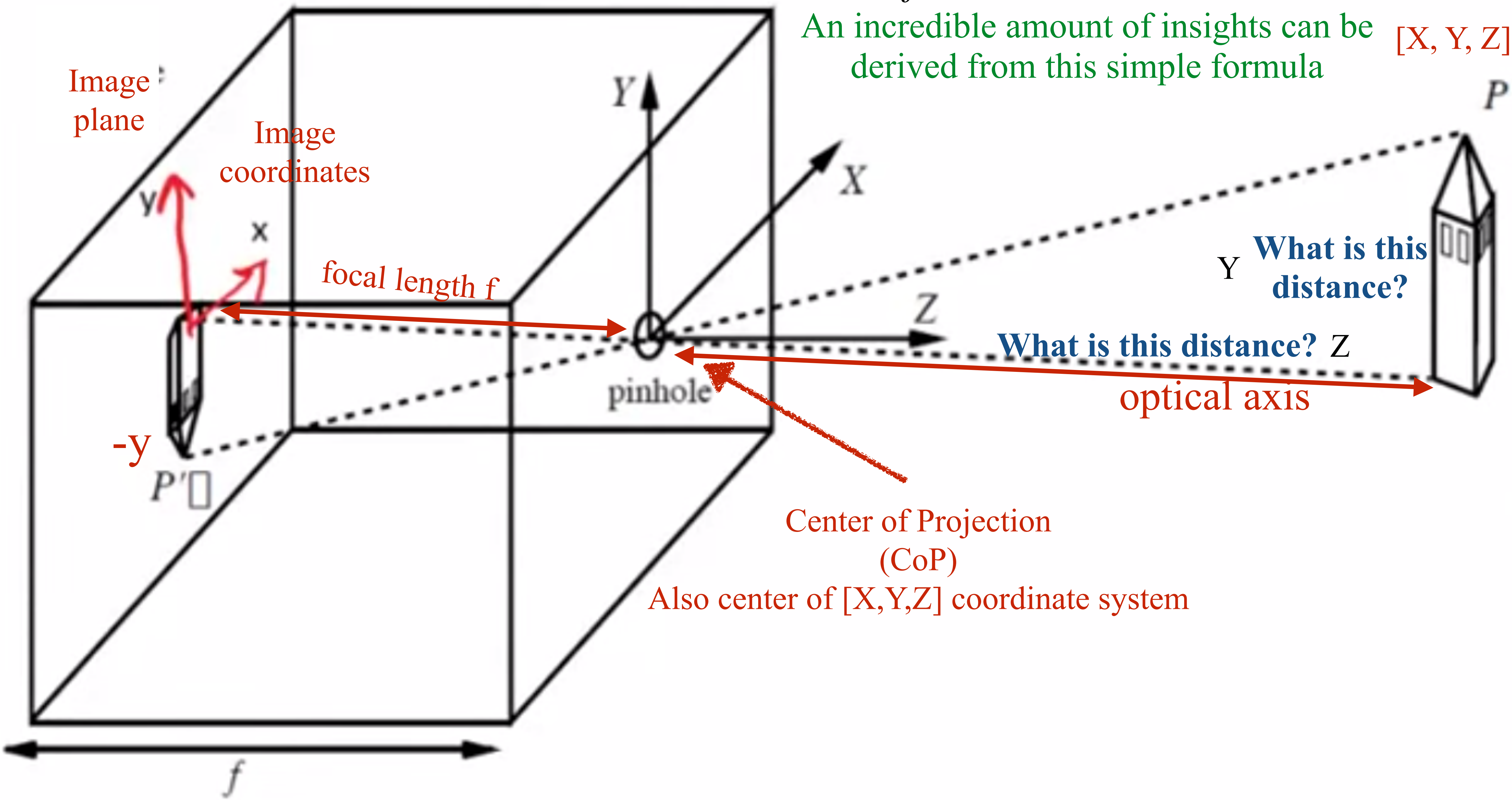
Antonio Torralba, William T. Freeman
Computer Science and Artificial Intelligence Laboratory (CSAIL)
MIT
torralba@mit.edu, billf@mit.edu

Pinhole Camera

Similar triangles:

$$\frac{-y}{f} = \frac{Y}{Z} \qquad y = -f\frac{Y}{Z} \qquad x = -f\frac{X}{Z}$$

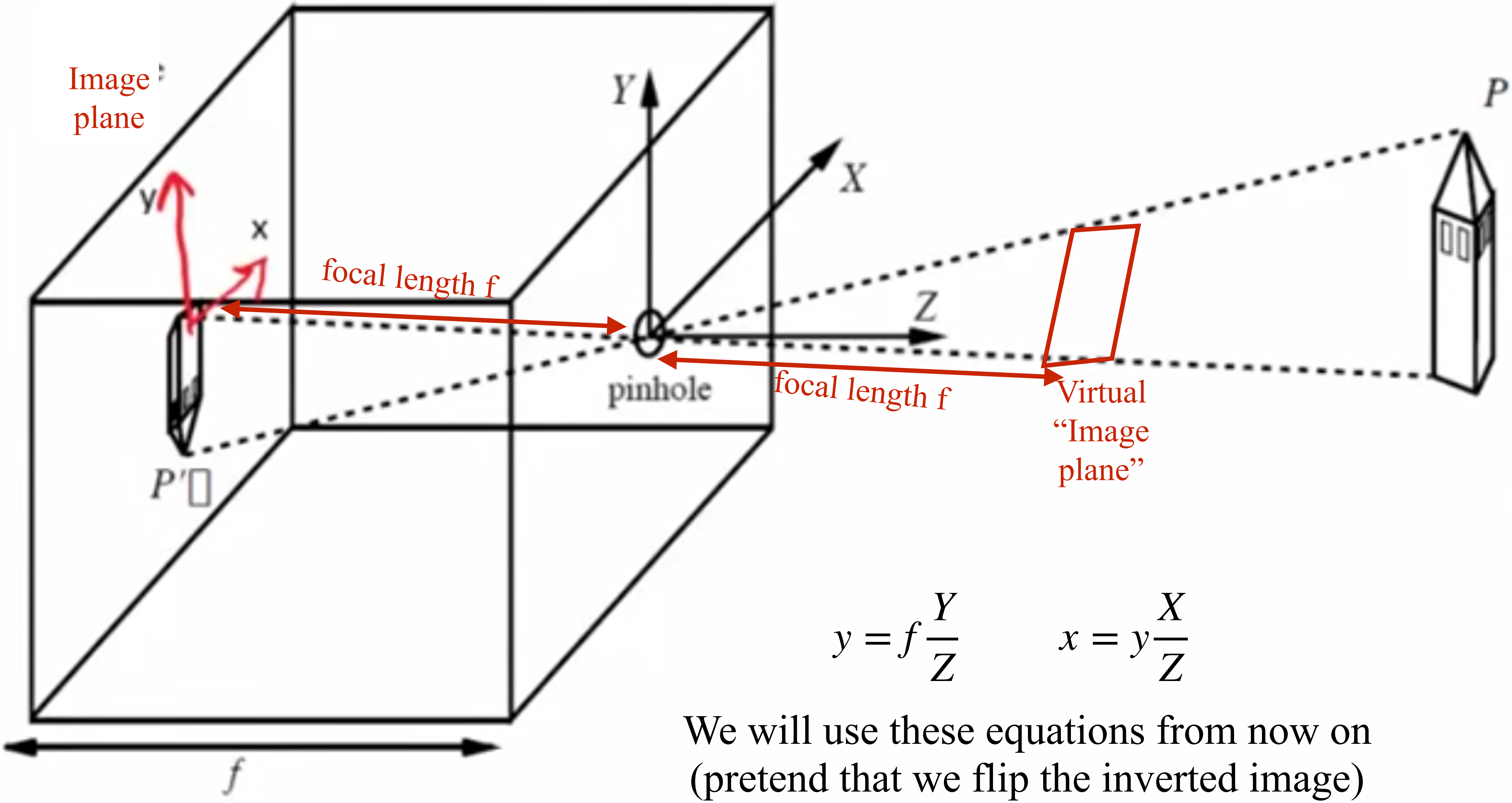
An incredible amount of insights can be derived from this simple formula



Given the world coordinates for some point [X, Y, Z]
we want to compute the image coordinates for this point [x, y]

(right-handed coordinate system)

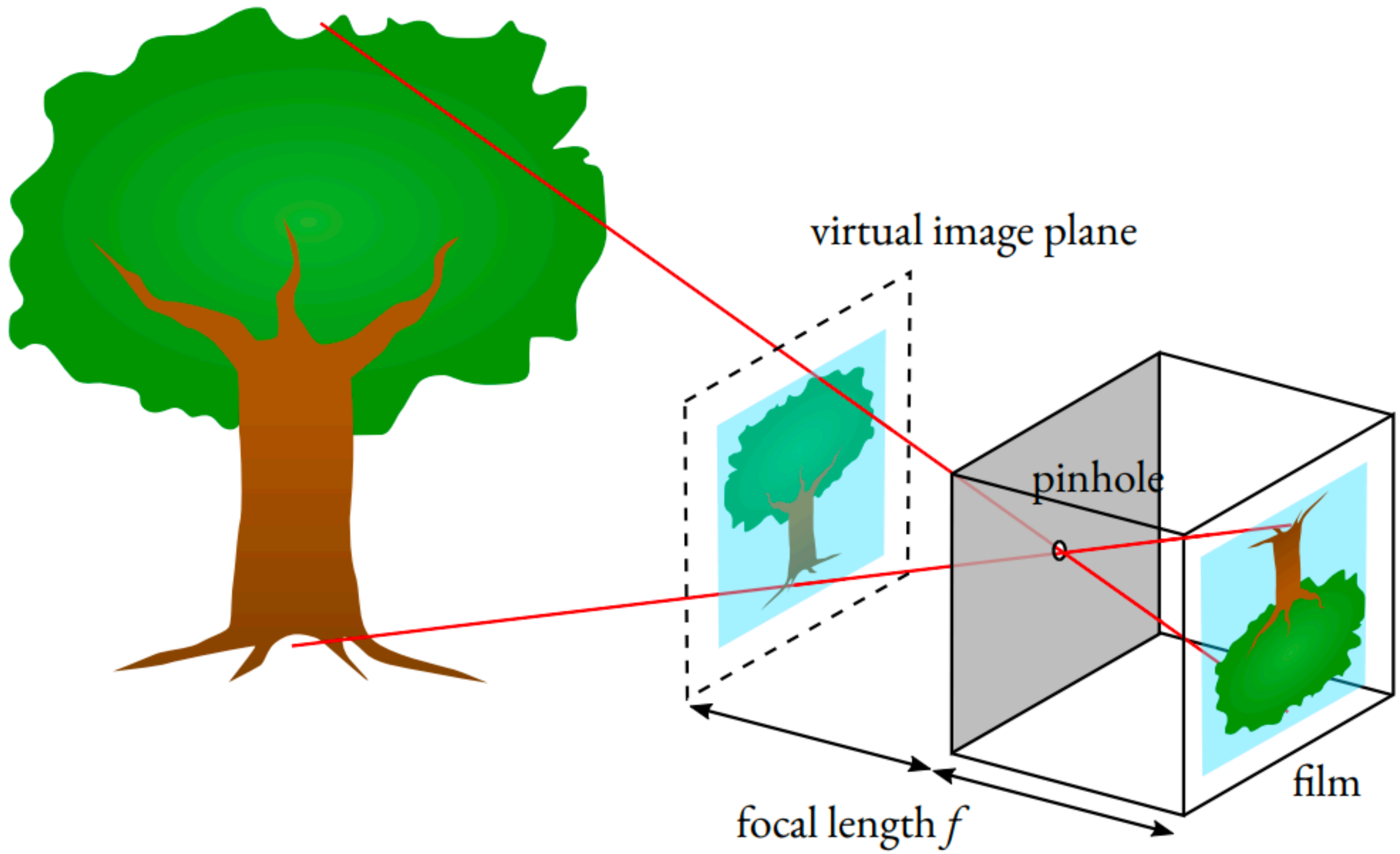
Trick to avoid inversion:



$$y = f\frac{Y}{Z} \qquad x = y\frac{X}{Z}$$

We will use these equations from now on
(pretend that we flip the inverted image)

Trick to avoid inversion:



$$y = f\frac{Y}{Z} \qquad x = y\frac{X}{Z}$$

We will use these equations from now on
(pretend that we flip the inverted image)

Physical model that avoids inversion