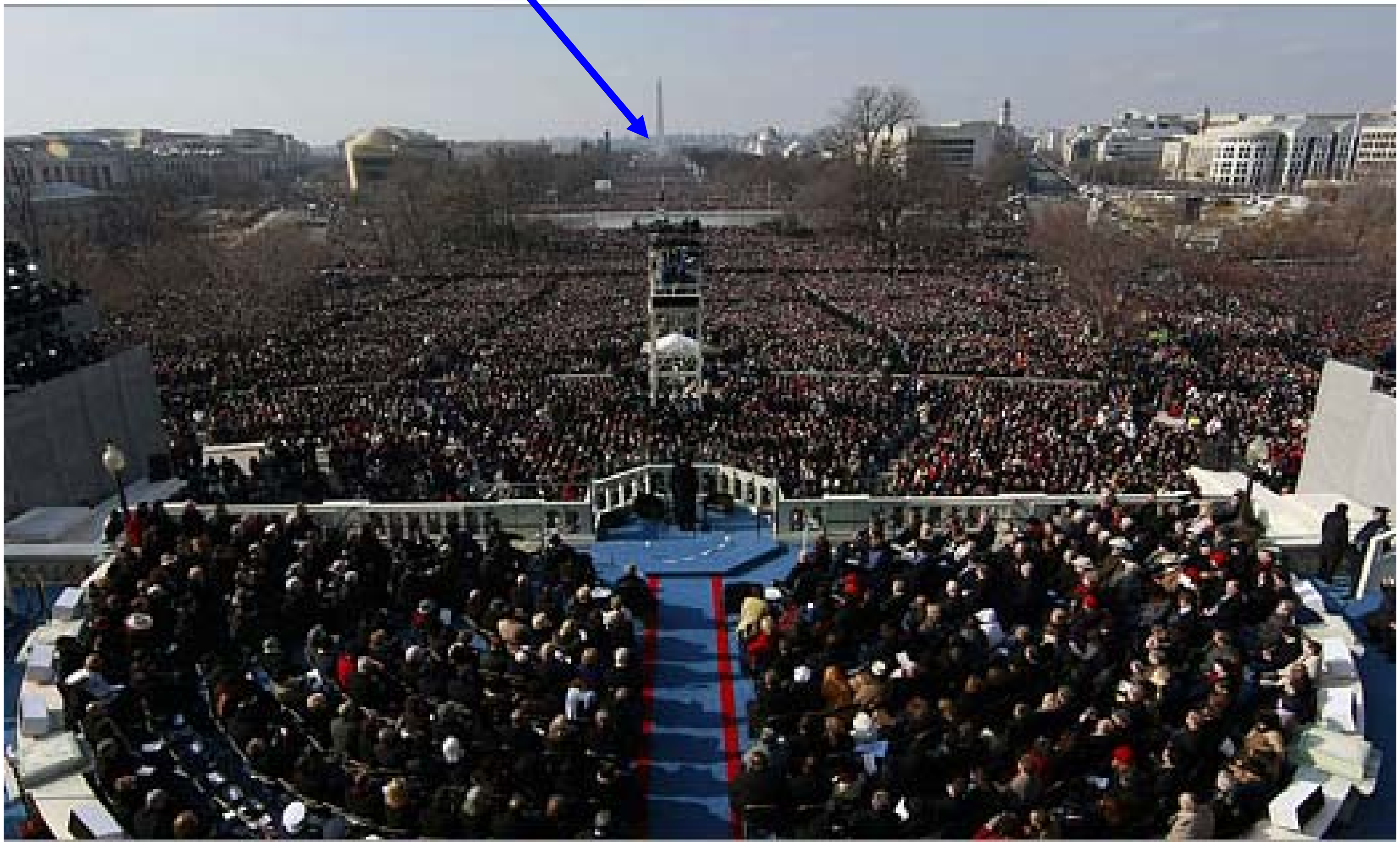
Lecture 5

Cameras, Projection, and Image Formation

Lesson from today's presidential inauguration

Vanishing point



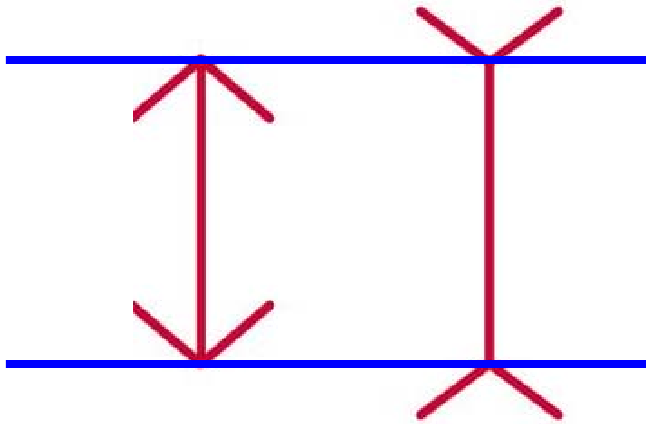
(from New York Times)

The world is spherical

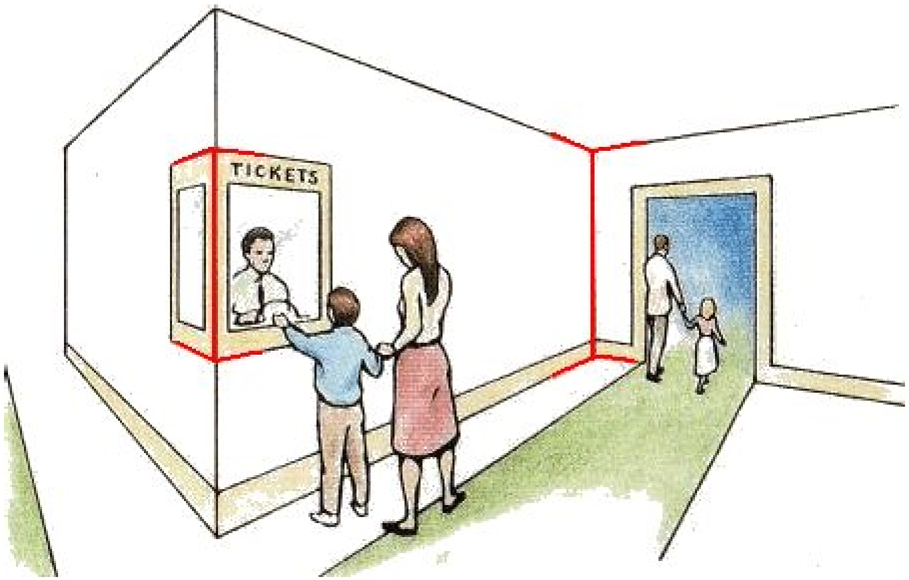
Wait…the world is flat

The brain constructs a 3D interpretation consistent with the 2D projection of the scene on your retina

Another Example: Müller-Lyer Illusion



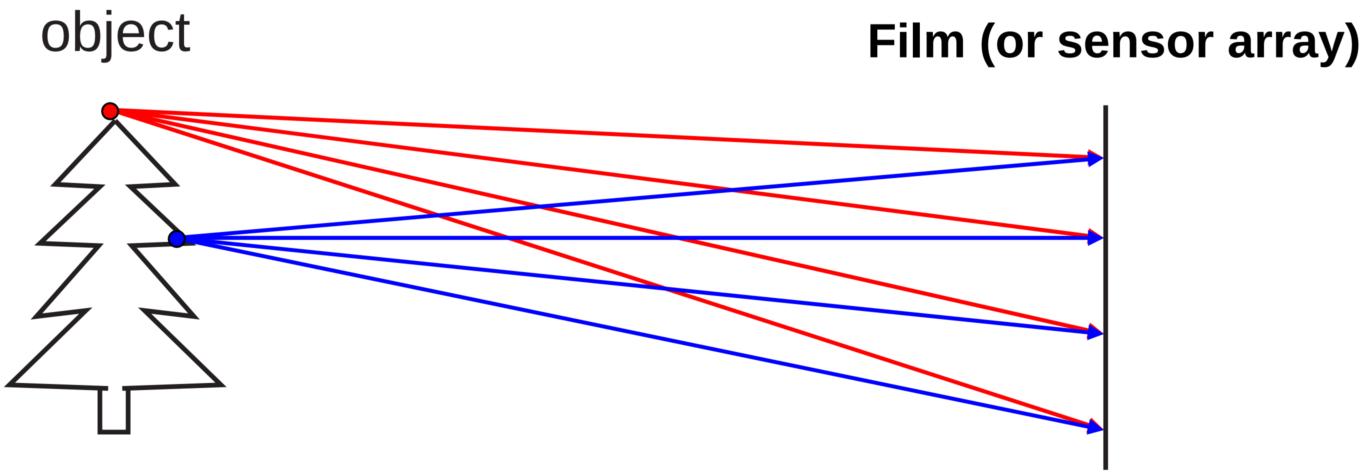
Which line is longer?



Illusion no more!
Makes sense for projection of 3D world onto 2D

http://www.michaelbach.de/ot/sze_muelue/index.html

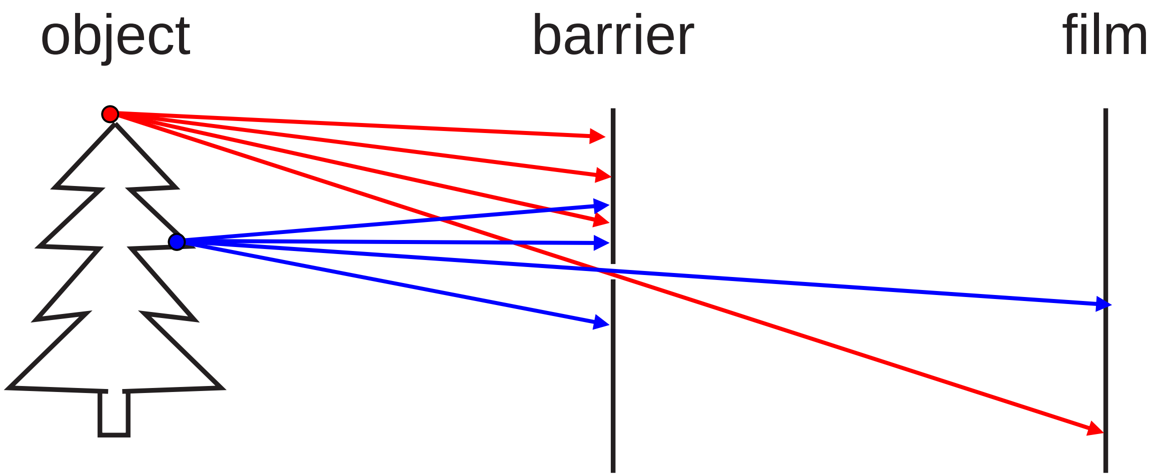
Image formation



Let's design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

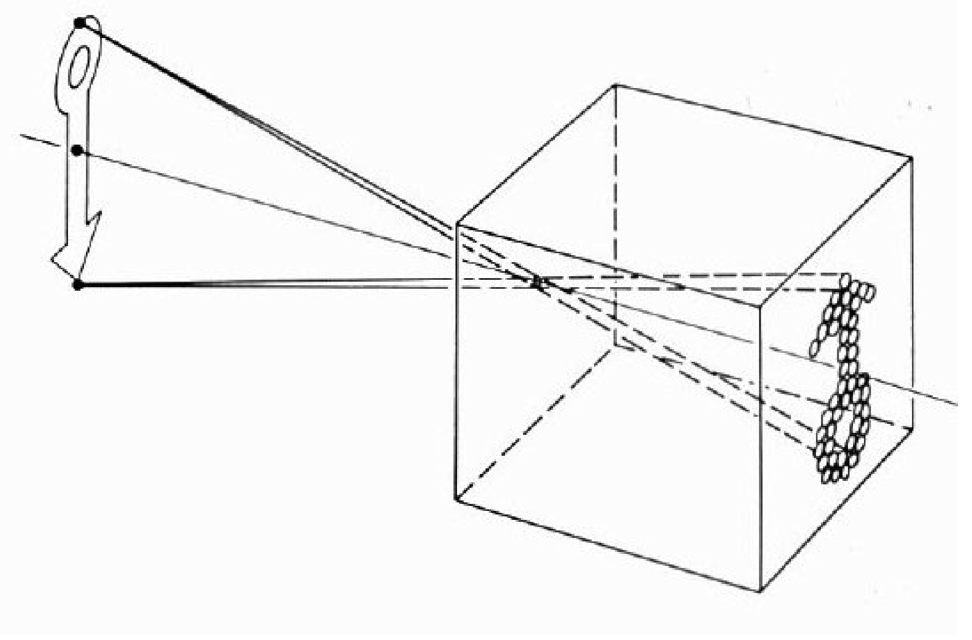
Pinhole camera



Add a barrier to block off most of the rays

- This reduces blurring
- The opening is known as the **aperture**
- How does this transform the image?

Camera Obscura

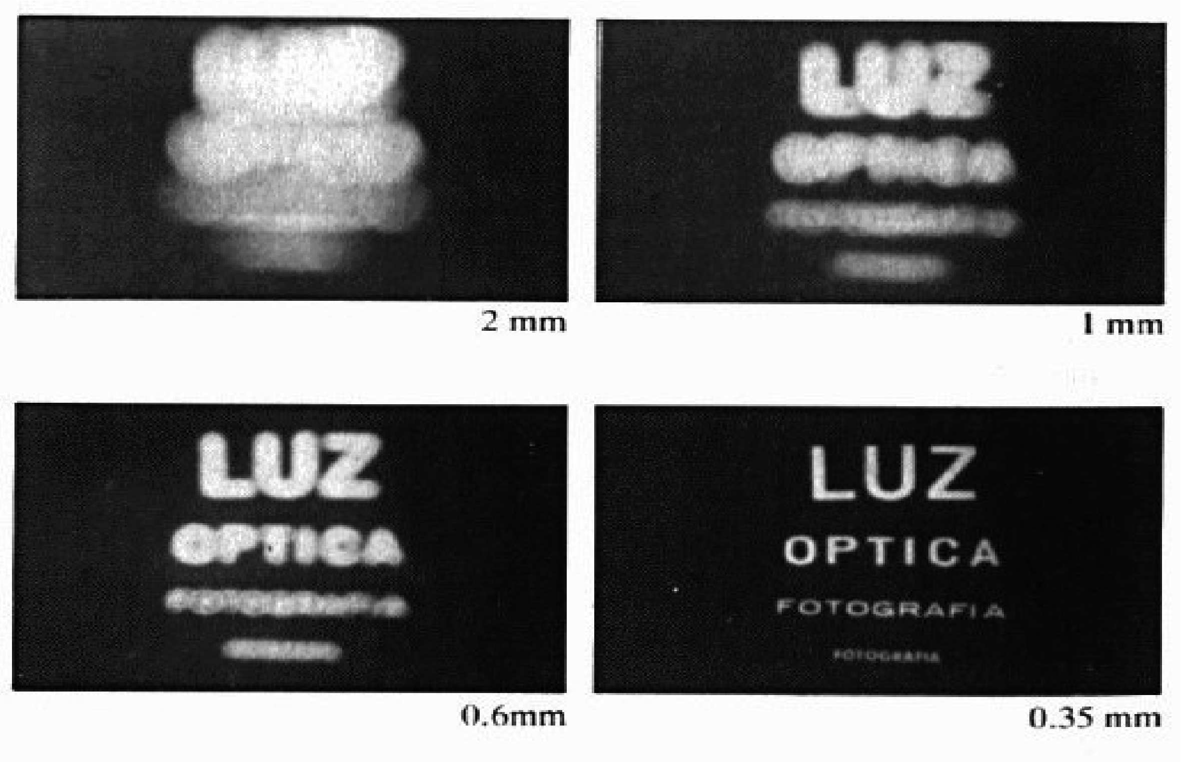


The first camera

- Known to Aristotle
- Analyzed by Ibn al-Haytham (Alhazen, 965-1039 AD) in Iraq

How does the aperture size affect the image?

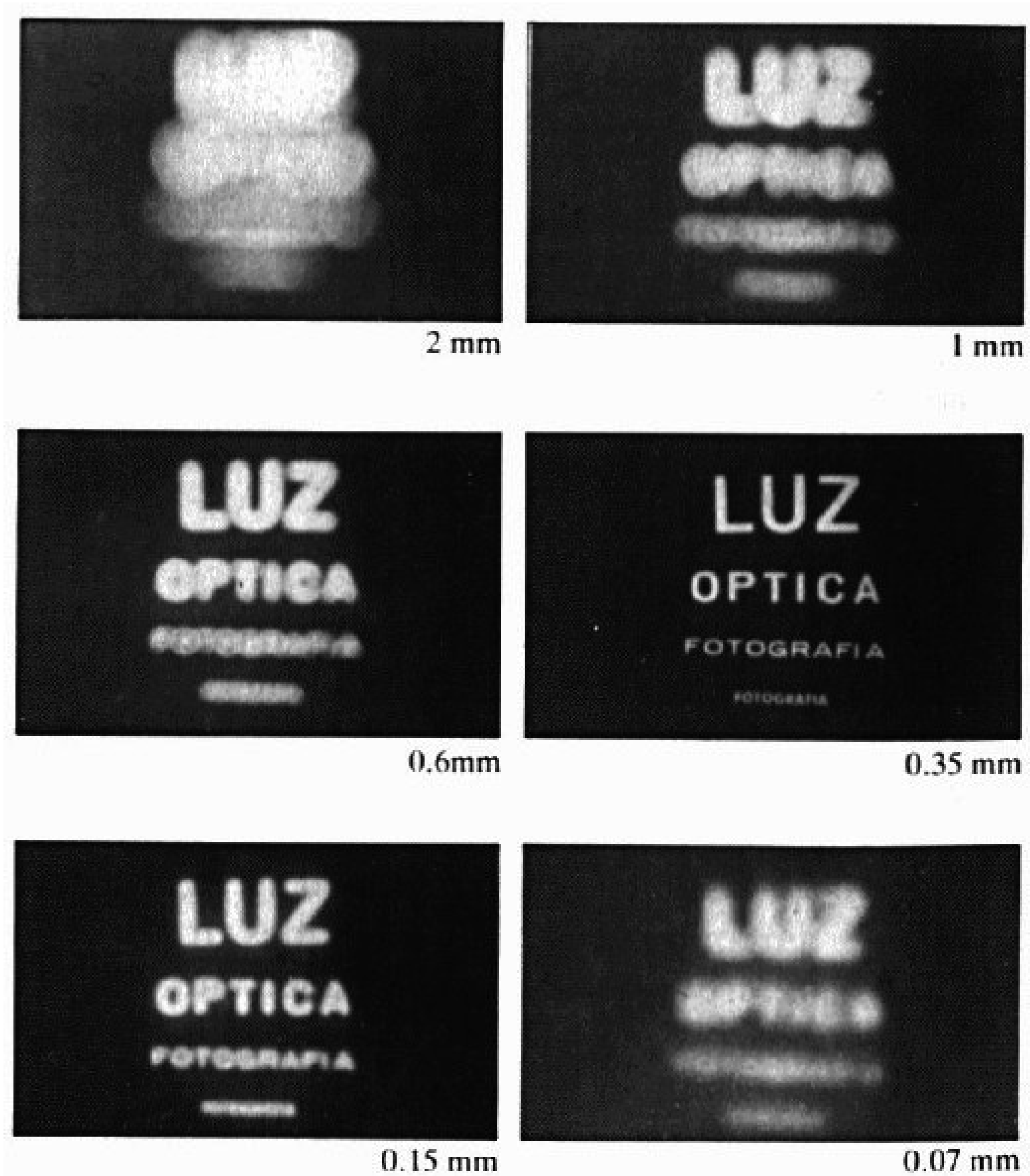
Shrinking the aperture



Why not make the aperture as small as possible?

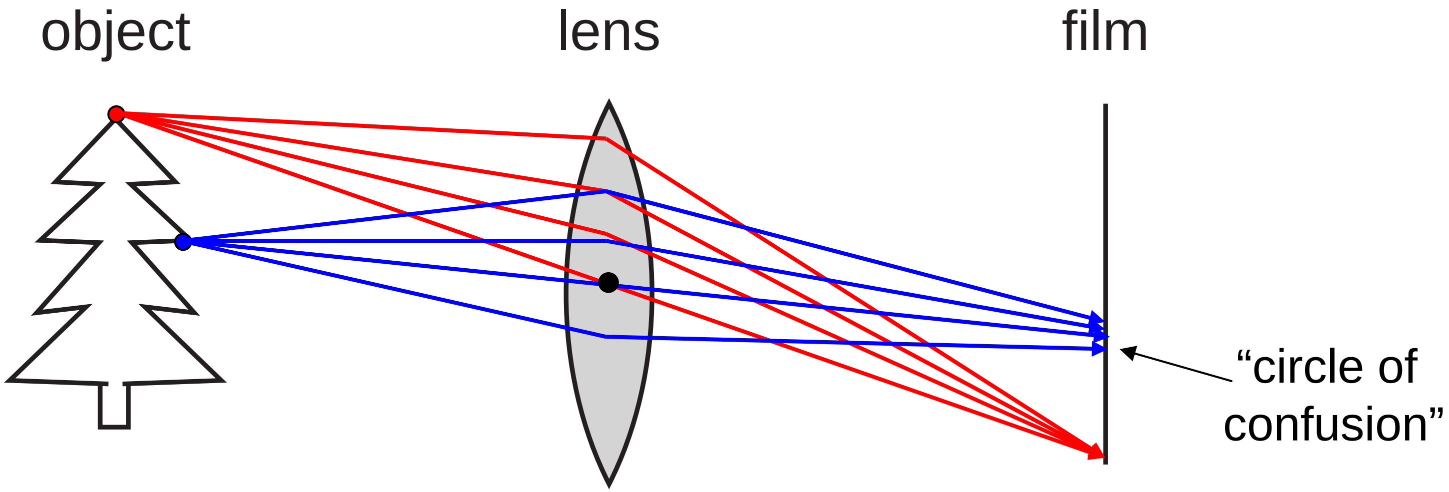
- Less light gets through
- *Diffraction* effects...

Shrinking the aperture



2 mm | 1 mm | 0.6mm | 0.35 mm | 0.15 mm | 0.07 mm

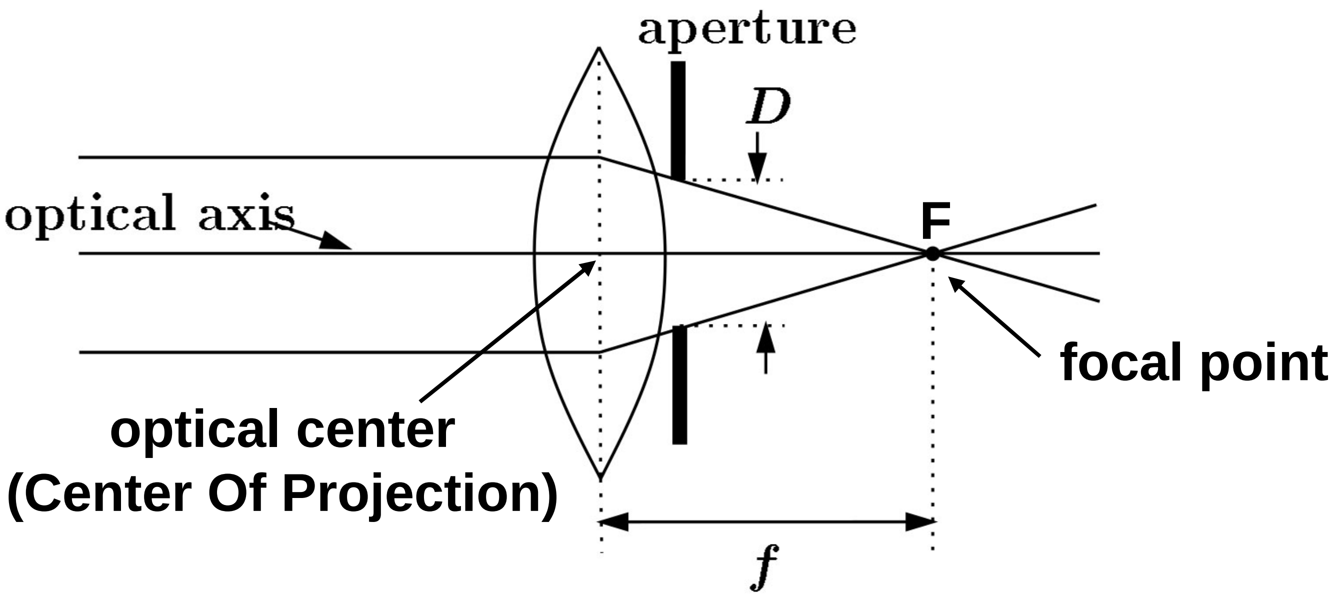
Adding a lens



A lens focuses light onto the film

- There is a specific distance at which objects are “in focus”
  - other points project to a “circle of confusion” in the image
- Changing the shape of the lens changes this distance

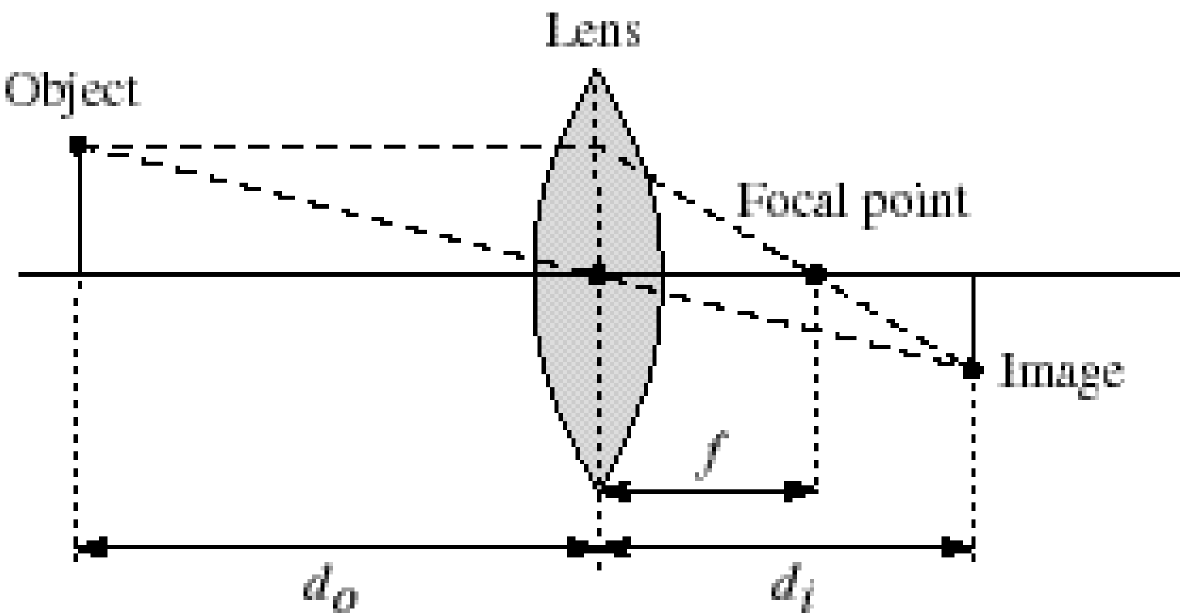
Lenses



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)

Thin lenses



Thin lens equation (derived using similar triangles):

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

- Any object point satisfying this equation is in focus (assuming $d_0 > f$)
  - What happens when $d_o < f$? (e.g, f = 2 and $d_o = 1$)

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

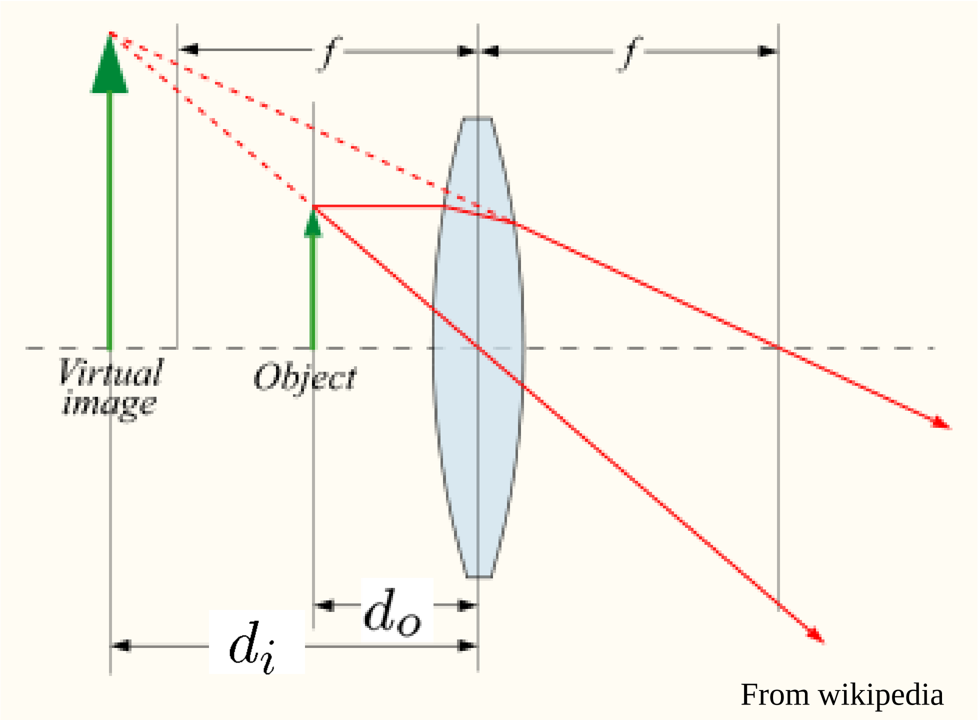
Magnification

When d_o < f, d_i becomes *negative*

We get a *virtual image* that an observer looking through the lens can see (as with a magnifying glass)

Magnification by the lens is defined as:

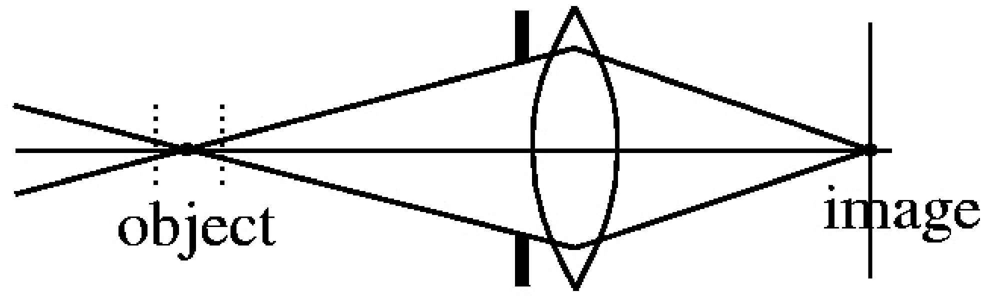
$$M = -\frac{d_i}{d_o} = \frac{f}{f - d_o}$$



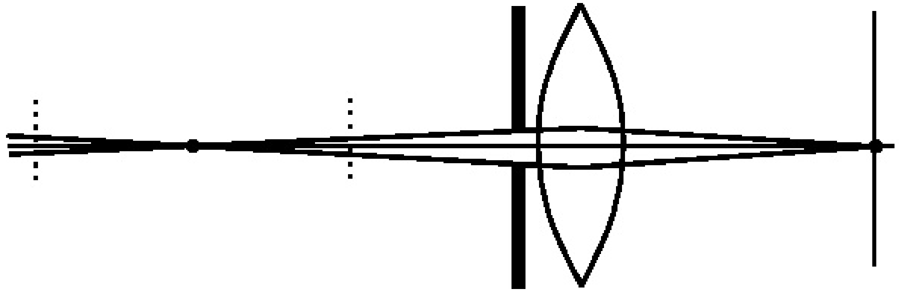
(M positive for upright (virtual) images, negative for real images)

|M| > 1 indicates magnification

Depth of field



f / 32



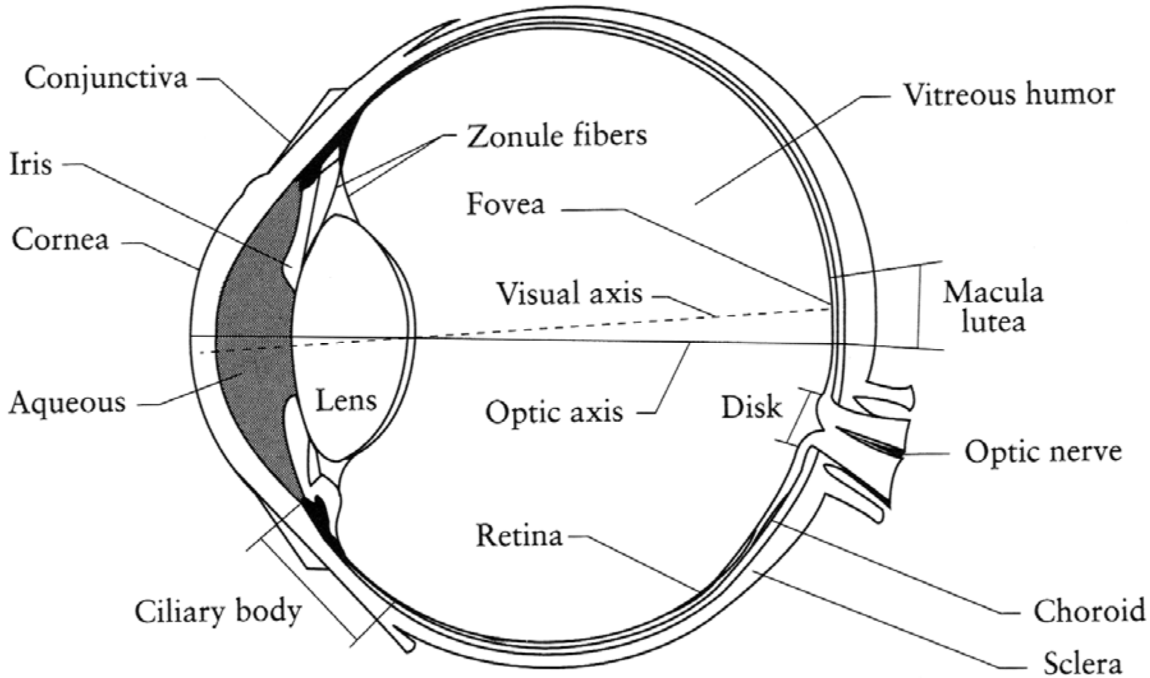
f / 5.6

Changing the aperture size affects depth of field

- A smaller aperture increases the range in which the object is approximately in focus

Flower images from Wikipedia http://en.wikipedia.org/wiki/Depth_of_field

The eye



The human eye is a camera

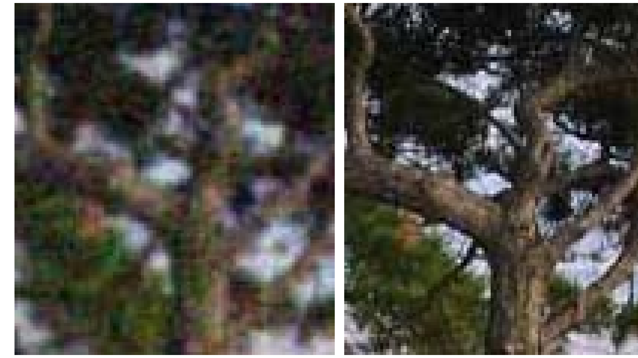
- **Iris** - colored disc with radial muscles and hole in the center
- **Pupil** - the hole (aperture) in iris whose size is controlled by iris muscles
- What's the "film"?
  - photoreceptor cells (rods and cones) in the **retina**

Digital camera

A digital camera replaces film with a sensor array

- Each cell in the array is a **C**harge **C**oupled **D**evice
  - light-sensitive diode that converts photons to electrons
  - other variants exist: CMOS is becoming more popular
  - http://electronics.howstuffworks.com/digital-camera.htm

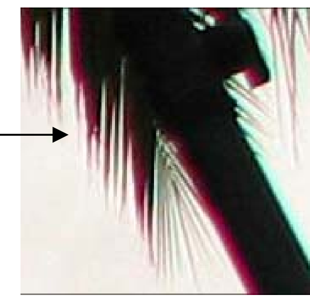
Issues with digital cameras



Noise
- big difference between consumer vs. SLR-style cameras
- low light is where you most notice noise

Compression
- creates artifacts except in uncompressed formats (tiff, raw)

Color
- color fringing artifacts

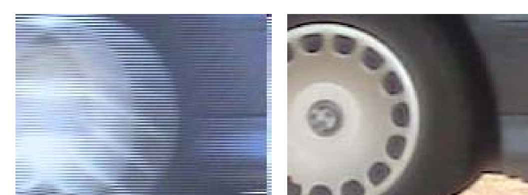


Blooming
- charge overflowing into neighboring pixels

In-camera processing
- oversharpening can produce halos

Stabilization
- compensate for camera shake (mechanical vs. electronic)

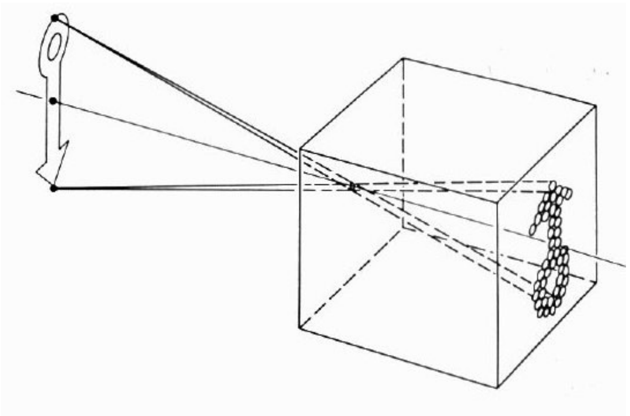
Interlaced vs. progressive scan video
- even/odd rows from different exposures vs entire picture



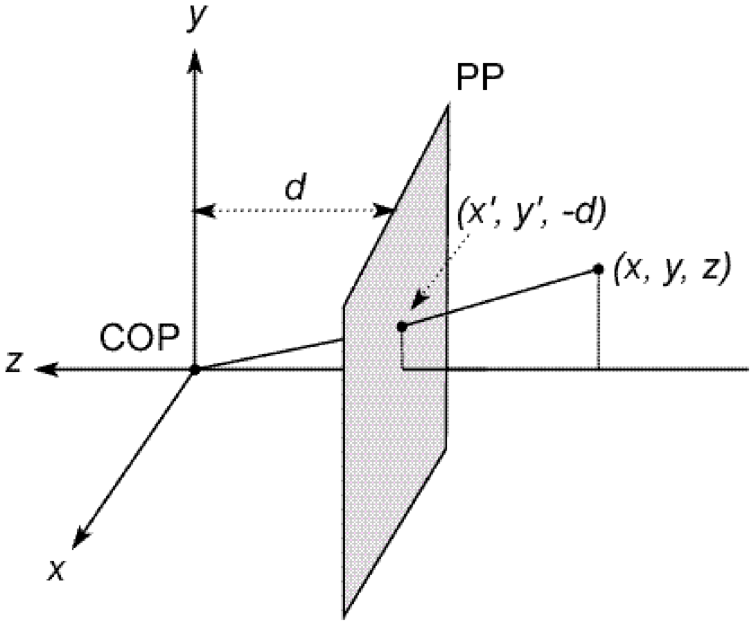
Interlaced Progressive

Modeling projection

Pinhole camera



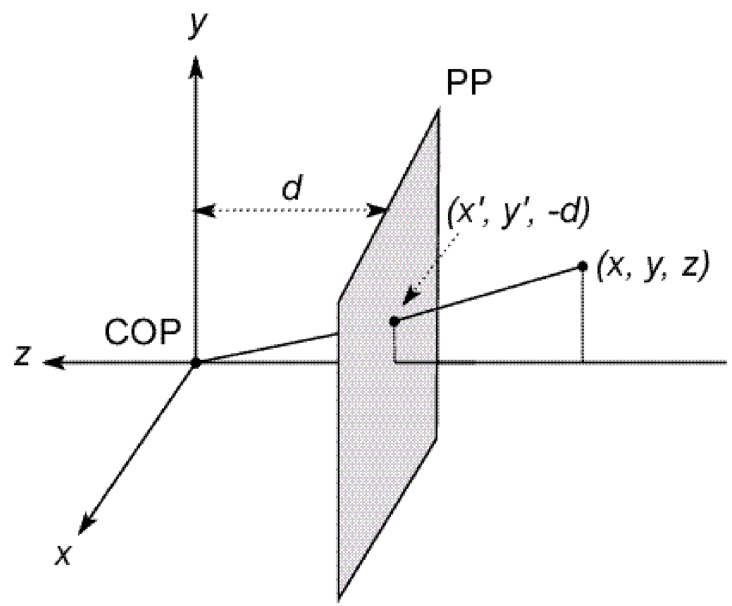
Model



The coordinate system

- We will use the pin-hole camera as an approximation
- Put the optical center (**C**enter **O**f **P**rojection) at the origin
- Put the image plane (**P**rojection **P**lane) *in front* of the COP – Why?
  - avoids inverted image and geometrically equivalent
- The camera looks down the *negative* z axis
  - we need this if we want right-handed-coordinates

Modeling projection



Projection equations

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

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

- Get projection coordinates on image by throwing out last coordinate:

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

Homogeneous coordinates

Is this a linear transformation? $(x, y, z) \to (-d\frac{x}{z},\ -d\frac{y}{z})$

- 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

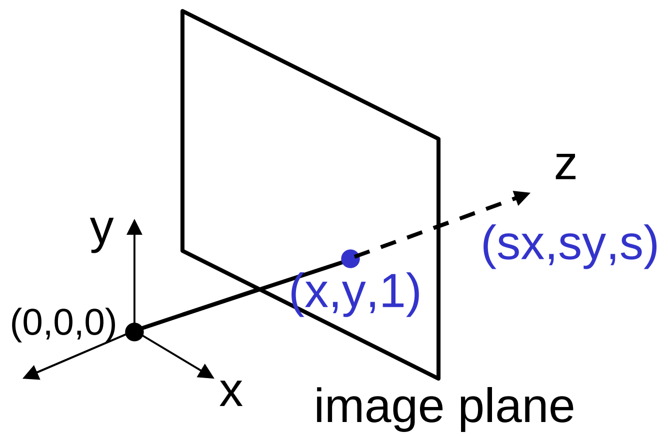
$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

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

Homogenous coordinates: Geometric intuition

Homogenous coords provide a way of extending N-d space to (N+1)-d space

- a point in the 2D image is treated as a *ray* in 3D projective space
- Each *point* (x,y) on the image plane is represented by the *ray* (sx,sy,s)
  - all points on the ray are equivalent: (x, y, 1) ≡ (sx, sy, s)
- Go back to 2D by dividing with last coordinate: (sx,sy,s)/s → (x,y)



Modeling Projection

Projection is a matrix multiply using homogeneous coordinates:

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

divide by third coordinate and throw it out to get image coords

This is known as **perspective projection**

- The matrix is the **projection matrix**
- Can also formulate as a 4x4 (today's handout does this)

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

divide by fourth coordinate and throw last two coordinates out

Perspective Projection

How does scaling the projection matrix change the transformation?

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

Scaling by c:

$$\begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & -c/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} cx \\ cy \\ -cz/d \end{bmatrix} \Rightarrow \left(-d\frac{x}{z}, \quad -d\frac{y}{z}\right)$$

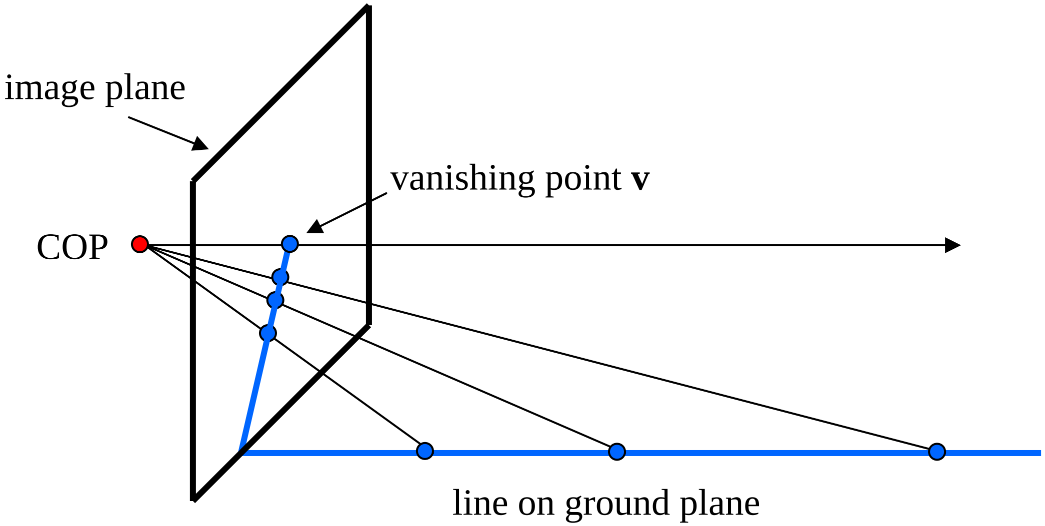
Same result if (x,y,z) scaled by c. This implies that:

In the image, a larger object further away (scaled x,y,z) can have the same size as a smaller object that is closer

Hence…

"Oh. Sorry for yelling. I thought you were much farther away."

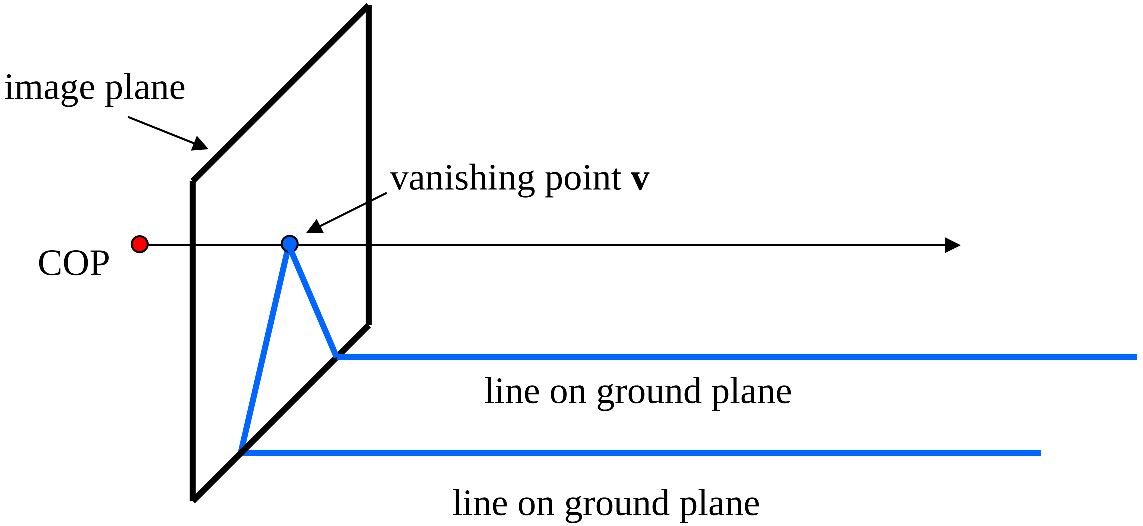
Vanishing points



Vanishing point

- projection of a point at infinity

Vanishing points



Properties

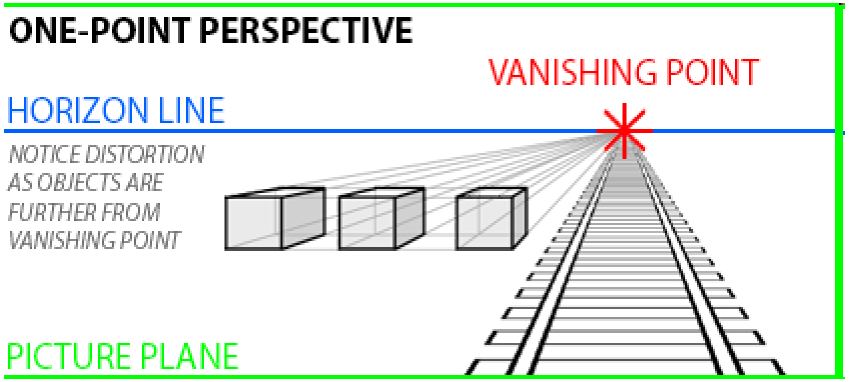
- Any two parallel lines have the same vanishing point **v**
- The ray from **COP** through **v** is parallel to the lines

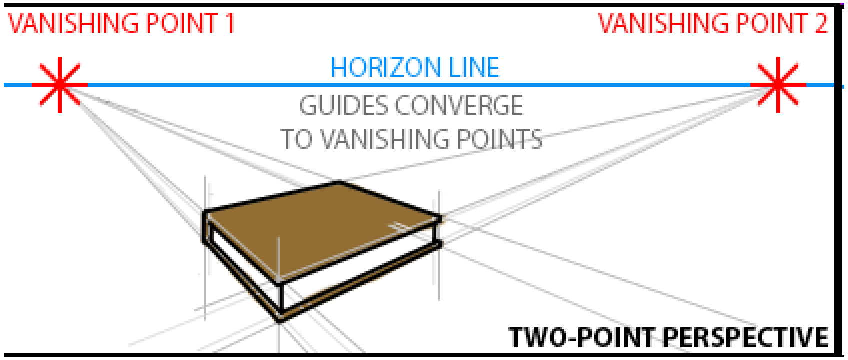
Examples in Real Images

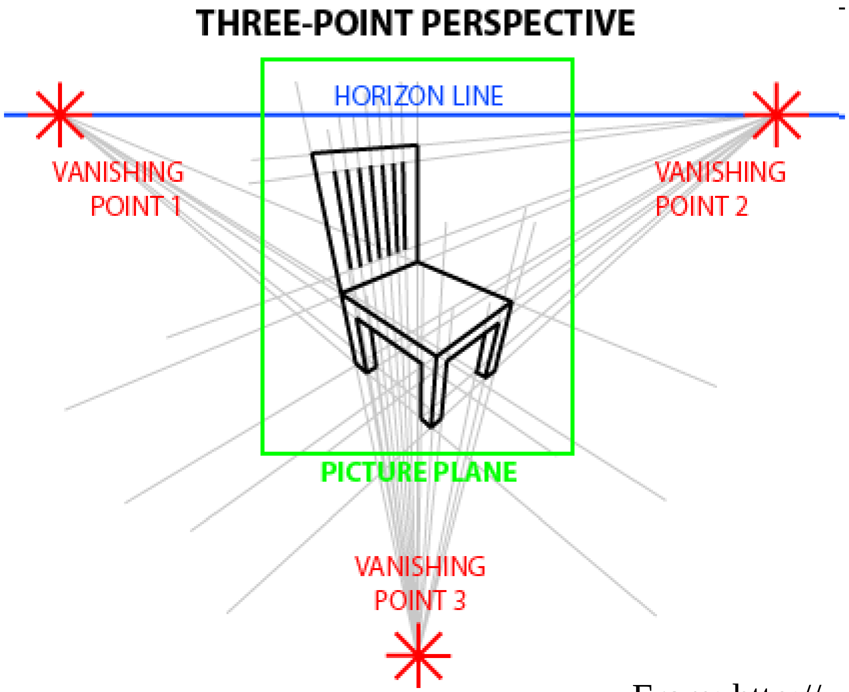
http://stevewebel.com/

http://phil2bin.com/

An image may have more than one vanishing point







From: http://www.atpm.com/9.09/design.shtml

Use in Art

Leonardo Da Vinci's Last Supper

Simplified Projection Models

Weak Perspective and Orthographic

Weak Perspective Projection

Recall Perspective Projection:

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

Suppose relative depths of points on object are much smaller than average distance z_{av} to COP

Then, for each point on the object,

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

where $c = -d / z_{av}$

(Projection reduced to **uniform scaling** of all object point coordinates)

Orthographic Projection

Suppose d $\rightarrow \infty$ in perspective projection model:

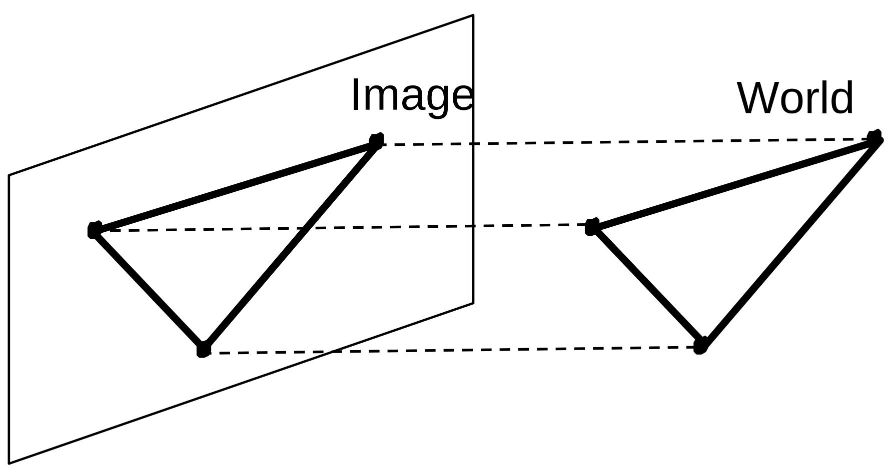
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

Then, we have $z \rightarrow$ -∞ so that $-d/z \rightarrow 1$

Therefore: $\boldsymbol{(x, y, z) \rightarrow (x, y)}$

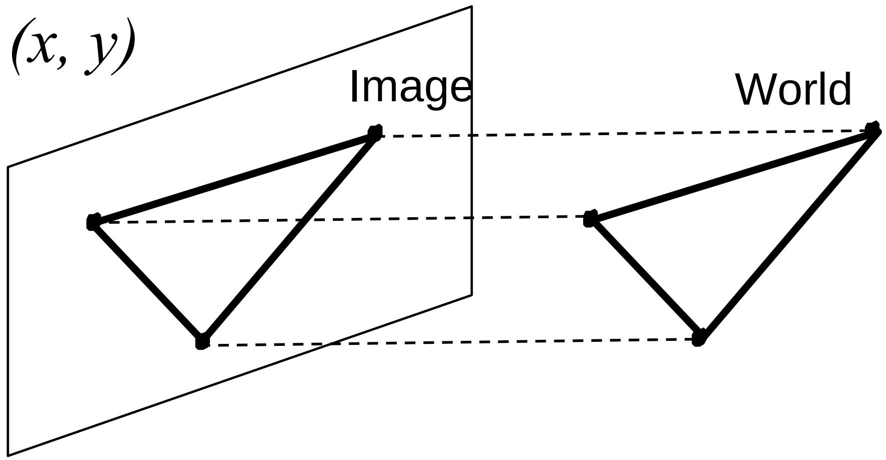
This is called orthographic or “parallel projection

Good approximation for telephoto optics



Orthographic projection

$(x, y, z) \rightarrow (x, y)$



What's the projection matrix in homogenous coordinates?

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

Weak Perspective Revisited

From the previous slides, it follows that:

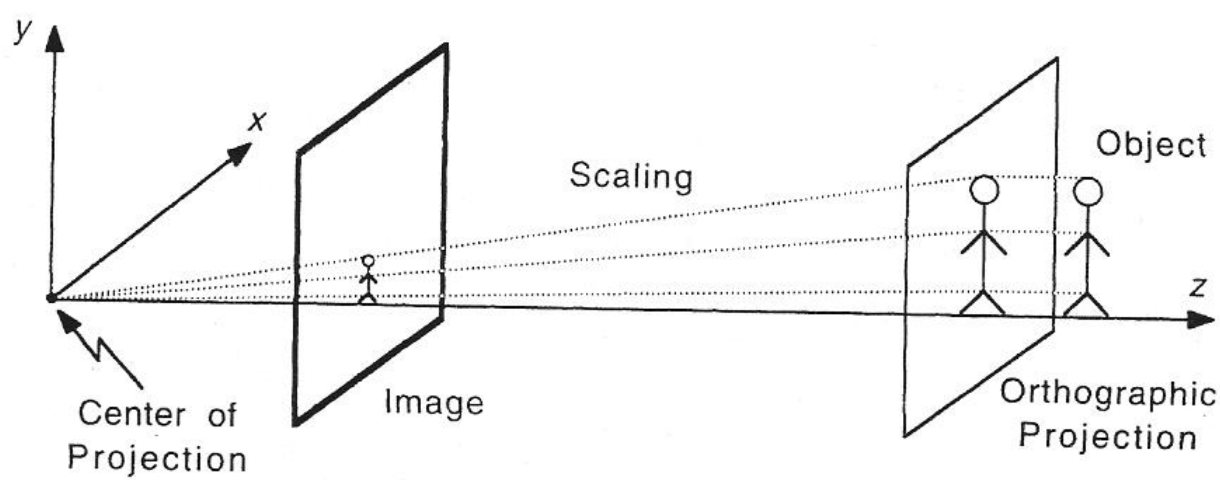
Weak perspective projection $(x, y, z) \rightarrow (cx, cy)$

=

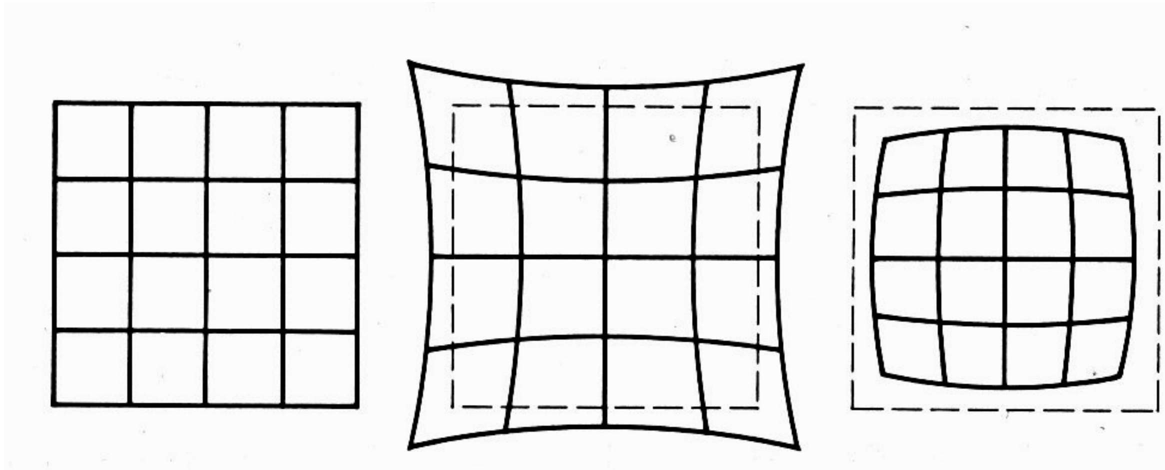
Orthographic projection $(x, y, z) \rightarrow (x, y)$

followed by

Uniform scaling by a factor $c = -d/z_{av}$



Distortions due to optics

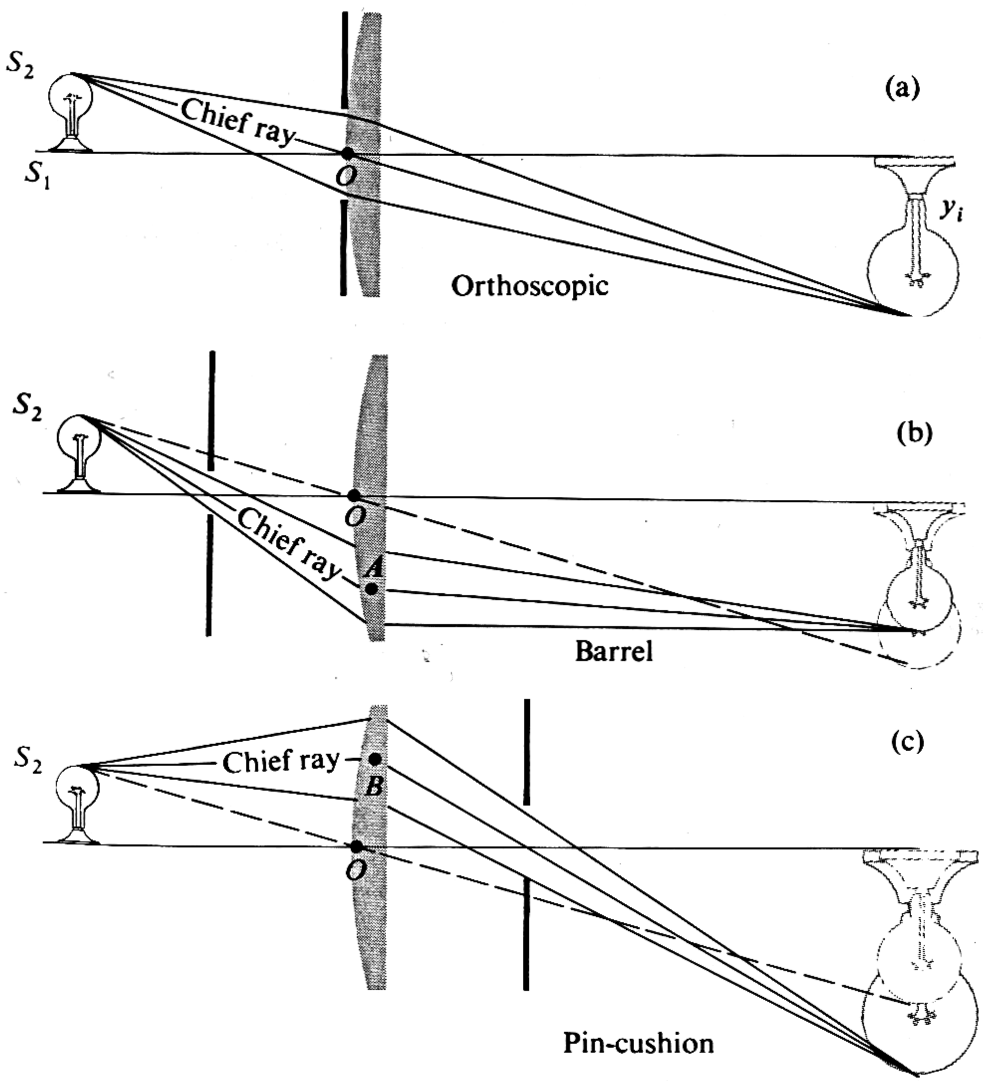


No distortion Pin cushion Barrel

Radial distortion of the image

- Caused by imperfect lenses
- Deviations are most noticeable for rays that pass through the edge of the lens, i.e., at image periphery

Distortion



S_2
Chief ray
S_1
O
(a)
y_i
Orthoscopic
S_2
(b)
O
Chief ray
A
Barrel
S_2
(c)
Chief ray
B
O
Pin-cushion

Modeling Radial Distortion

- Radial distortion is typically modeled as:

$$x = x_d(1 + k_1 r^2 + k_2 r^4)$$

$$y = y_d(1 + k_1 r^2 + k_2 r^4)$$

$$\text{where } r^2 = x_d^2 + y_d^2$$

- (x_d, y_d) are coordinates of distorted points wrt image center
- (x, y) are coordinates of the corrected points
- Distortion is a radial displacement of image points
  - increases with distance from center
- k_1 and k_2 are parameters to be estimated
- k_1 usually accounts for 90% of distortion

Correcting radial distortion

Barrel distortion

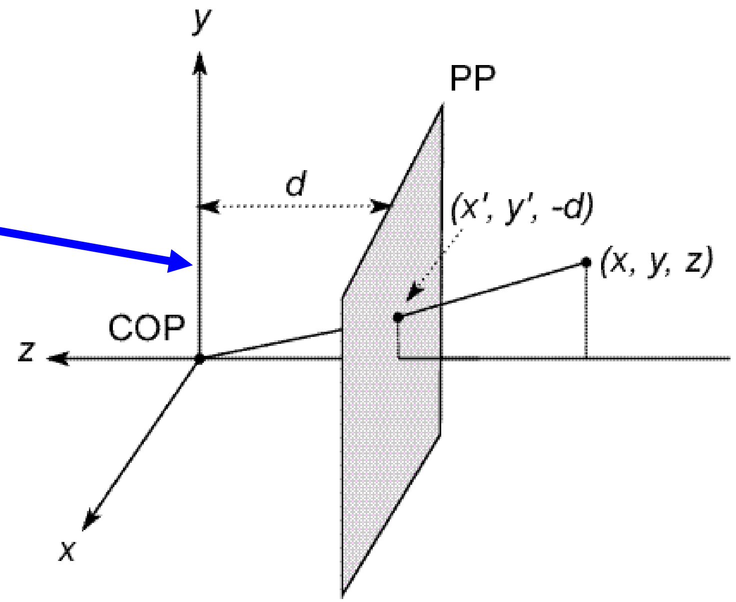
Corrected

from [Helmut Dersch]

Putting it all together: Camera parameters

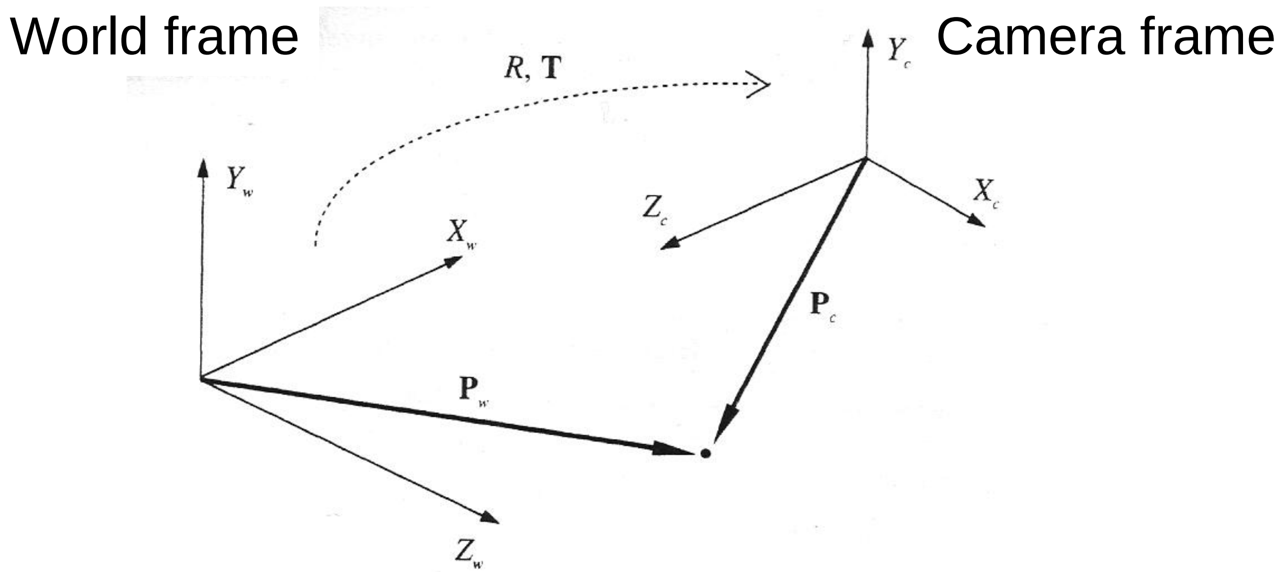
- Want to link coordinates of points in 3D external space with their coordinates in the image
- Perspective projection was defined in terms of *camera reference frame*

Camera reference frame



- Need to find location and orientation of camera reference frame with respect to a known “world” reference frame (these are the *extrinsic parameters*)

Extrinsic camera parameters



- Parameters that describe the transformation between the camera and world frames:
  - 3D translation vector **T** describing relative displacement of the origins of the two reference frames
  - 3 x 3 rotation matrix *R* that aligns the axes of the two frames onto each other
- Transformation of point $\mathbf{P}_w$ in world frame to point $\mathbf{P}_c$ in camera frame is given by: $\mathbf{P}_c = R(\mathbf{P}_w - \mathbf{T})$

Intrinsic camera parameters

- Parameters that characterize the optical, geometric and digital properties of camera
  - Perspective projection parameter: focal length d in previous slides
  - Distortion due to optics: radial distortion parameters k_1, k_2
  - Transformation from camera frame to pixel coordinates:
    - Coordinates (x_{im},y_{im}) of image point in pixel units related to coordinates (x,y) of same point in camera ref frame by:

      $x = -(x_{im} - o_x)s_x$

      $y = -(y_{im} - o_y)s_y$

      where (o_x,o_y) is the image center and s_x, s_y denote size of pixel

    (Note: - sign in equations above are due to opposite orientations of x/y axes in camera and image reference frames)

- Estimation of extrinsic and intrinsic parameters is called camera calibration
  - typically uses a 3D object of known geometry with image features that can be located accurately

From world coordinates to pixel coordinates

Plugging $\mathbf{P}_c = R(\mathbf{P}_w - \mathbf{T})$, $x = -(x_{im} - o_x)s_x$, and $y = -(y_{im} - o_y)s_y$ into perspective projection equation, we get

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = M_{\text{int}} M_{ext} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

where $(x_{im}, y_{im}) = (x/z, y/z)$

Camera to image ref frame

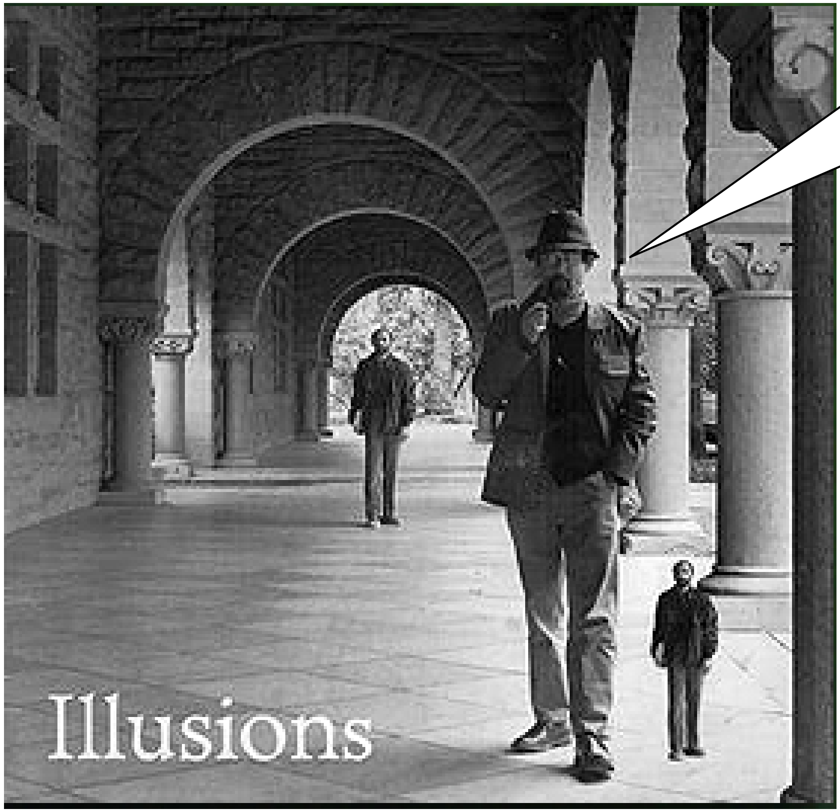
World to camera ref frame

$$M_{\text{int}} = \begin{bmatrix} d/s_x & 0 & o_x \\ 0 & d/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \quad M_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -\mathbf{R}_1\mathbf{T} \\ r_{21} & r_{22} & r_{23} & -\mathbf{R}_2\mathbf{T} \\ r_{31} & r_{32} & r_{33} & -\mathbf{R}_3\mathbf{T} \end{bmatrix}$$

(r_{ij} are the elements of rotation matrix R; $\mathbf{R}_i$ is its *i-th* row)
All parameters estimated by camera calibration procedure

Next time: Image Features & Interest Operators

- Things to do:
  - Work on Project 1: Use Sieg 327 if skeleton software not working on your own computer
  - Readings online



How's this for perspective?