MIT COMPUTER VISION

MIT CSAIL

6.869: Advances in Computer Vision

Bill Freeman, Antonio Torralba, and Phillip Isola, 2018

Lecture 2

Image formation

bring:

gumby, big flash light, apertures, straw camera, one paper bag camera

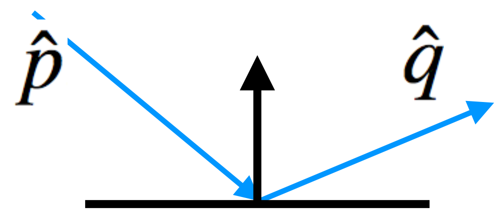
lens, holder, laser pointer, corner camera pieces of paper, tape measure (for lens focal length).

Imaging

- Forming images with pinholes and edges
- Forming images with lenses
- More general imaging devices

Light striking a surface

- Bidirectional reflectance distribution function



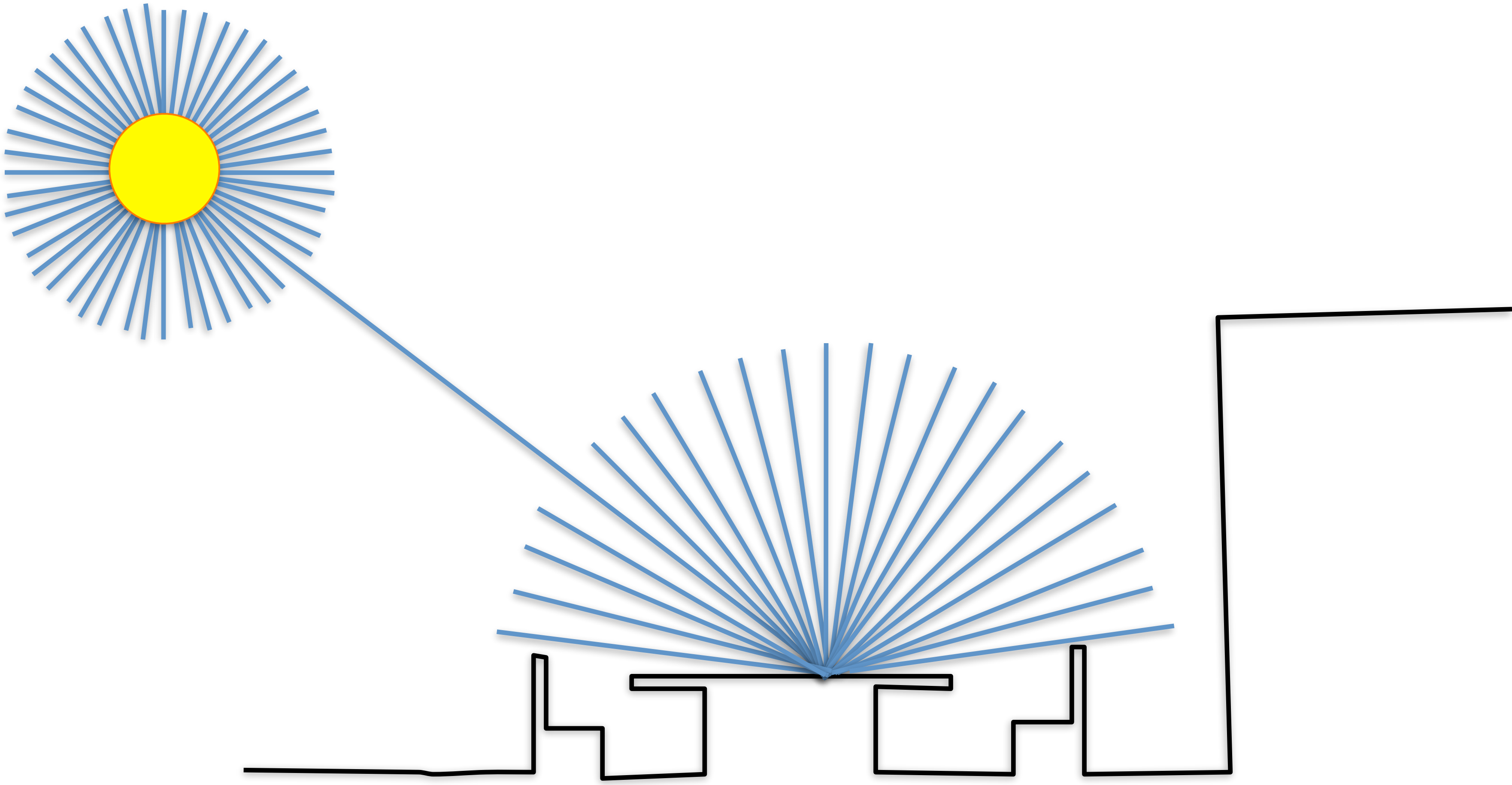
$$I_{\text{out}} = F(I_{\text{in}}, \hat{n}, \lambda, \hat{p}, \hat{q})$$

- For a Lambertian surface:

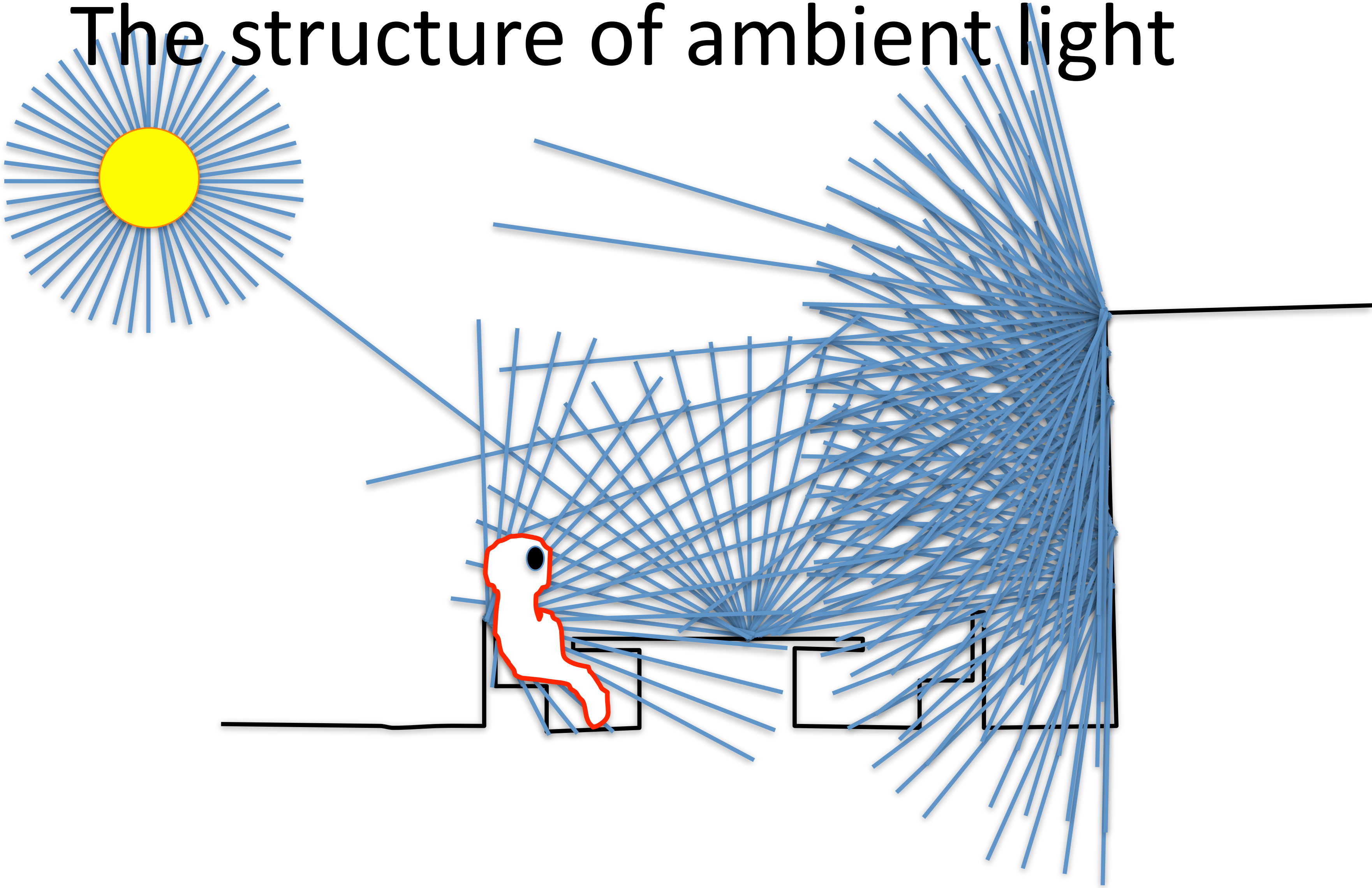
$$I_{\text{out}} = F_L(I_{\text{in}}, \hat{n}, \hat{p}) = A I_{\text{in}}(\lambda) \cos(\hat{n} \cdot \hat{p})$$


light from the sun
light reflecting from
the object surface

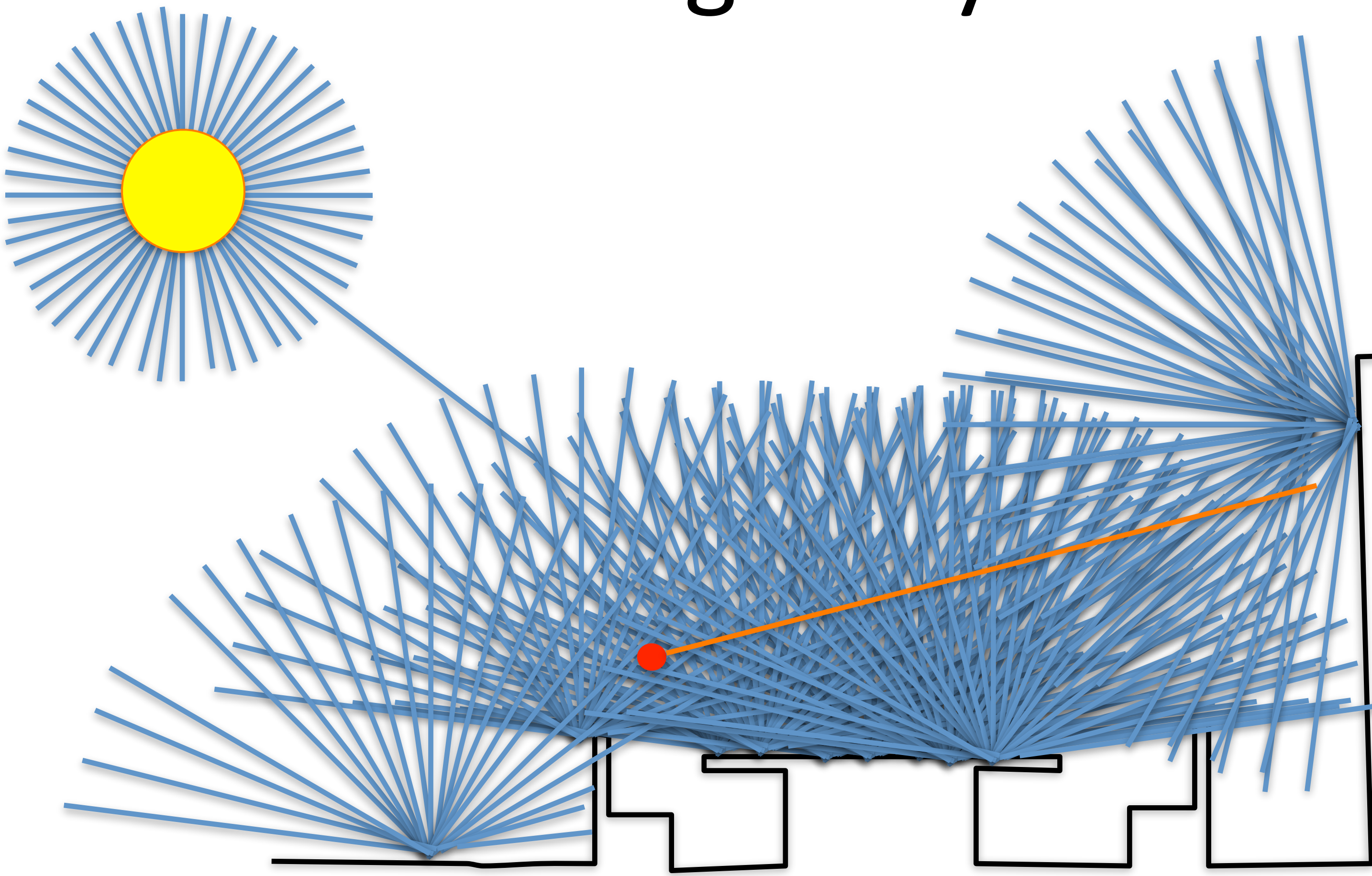
The structure of ambient light



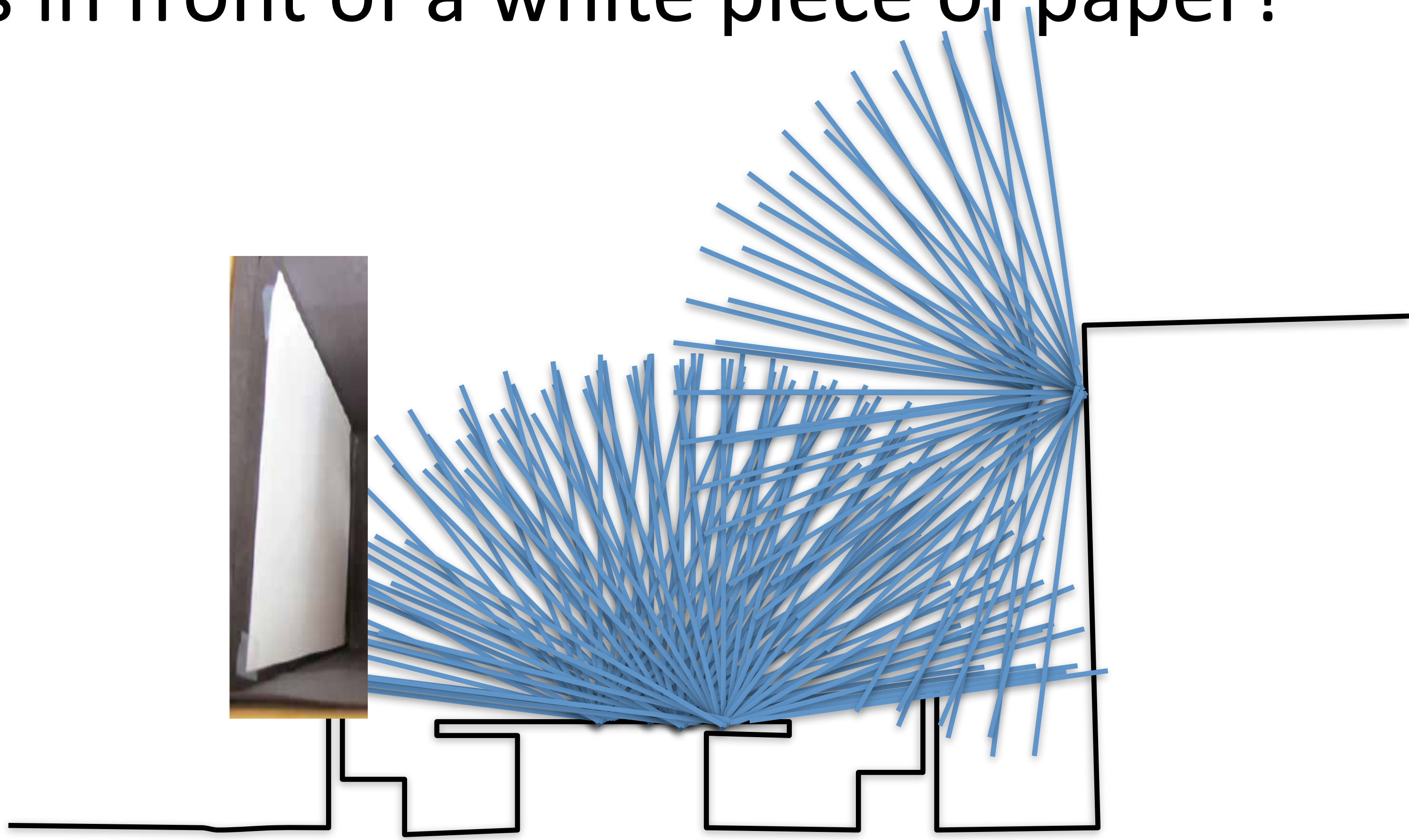
The structure of ambient light



All light rays

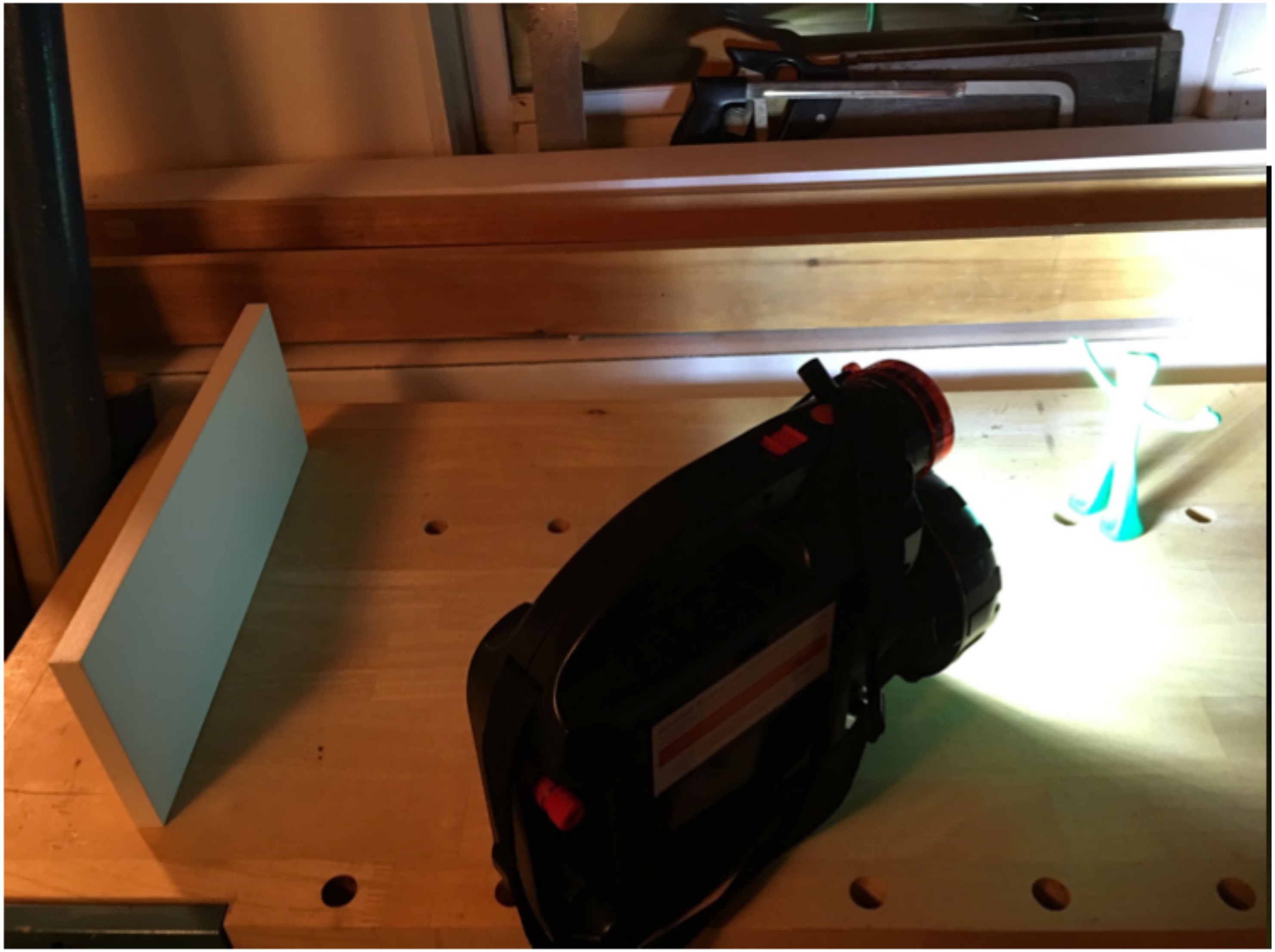


Why don't we generate an image when an object is in front of a white piece of paper?

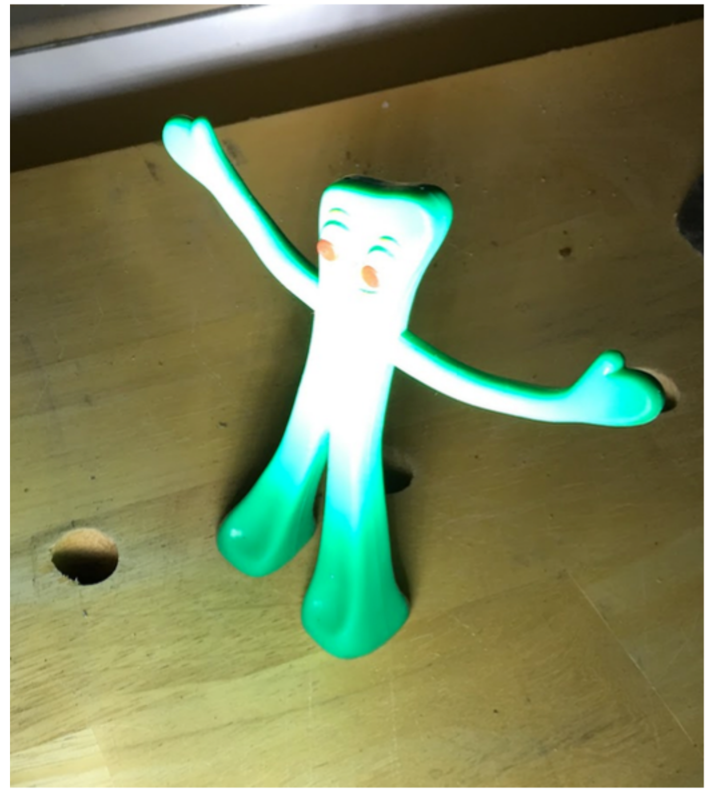


Why is there no picture appearing on the paper?

Let's check, do we get an image?



Let’s check, do we get an image? No



To make an image, we need to have only a subset of all the rays strike the sensor or surface

The camera obscura
The pinhole camera

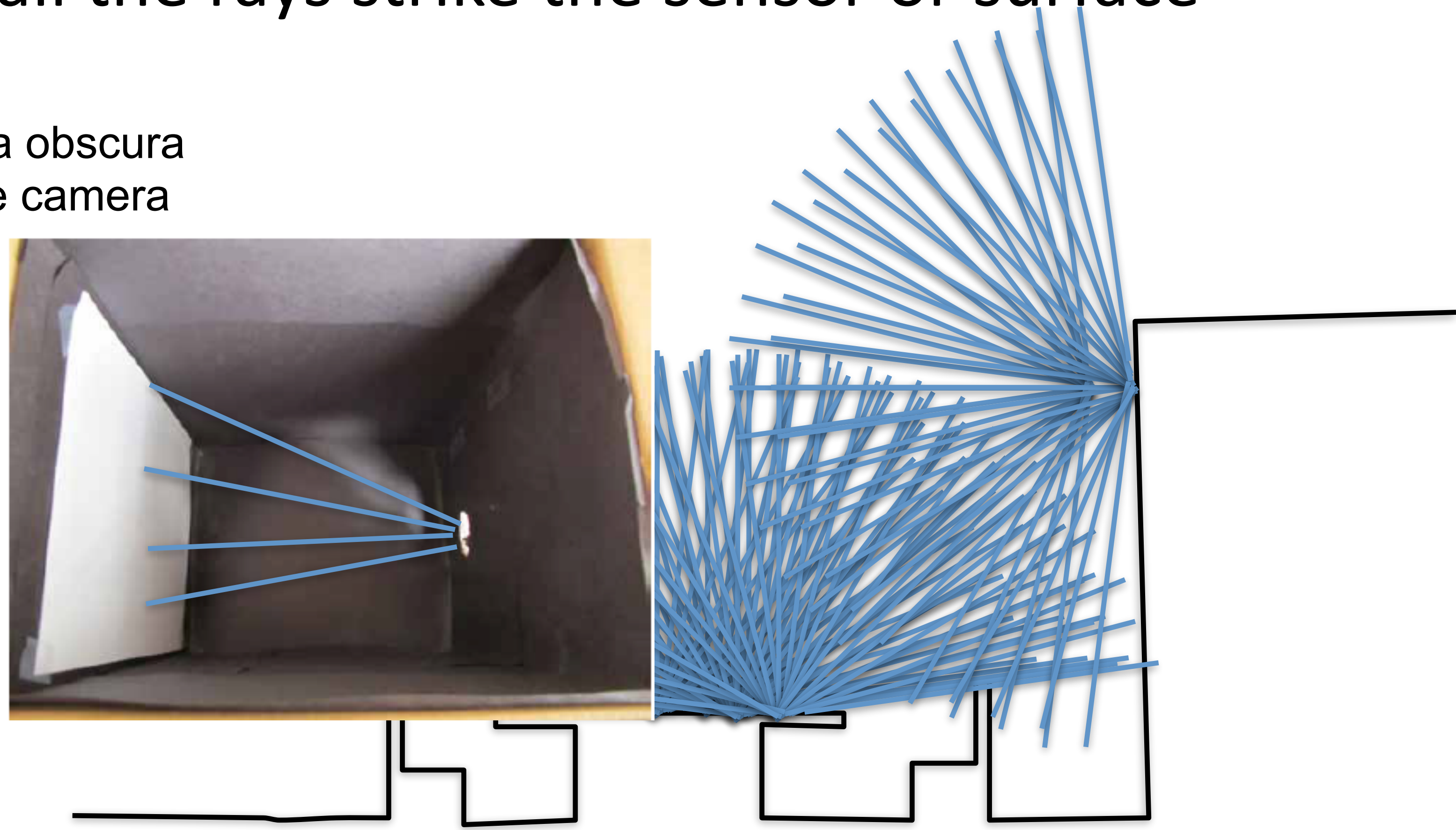
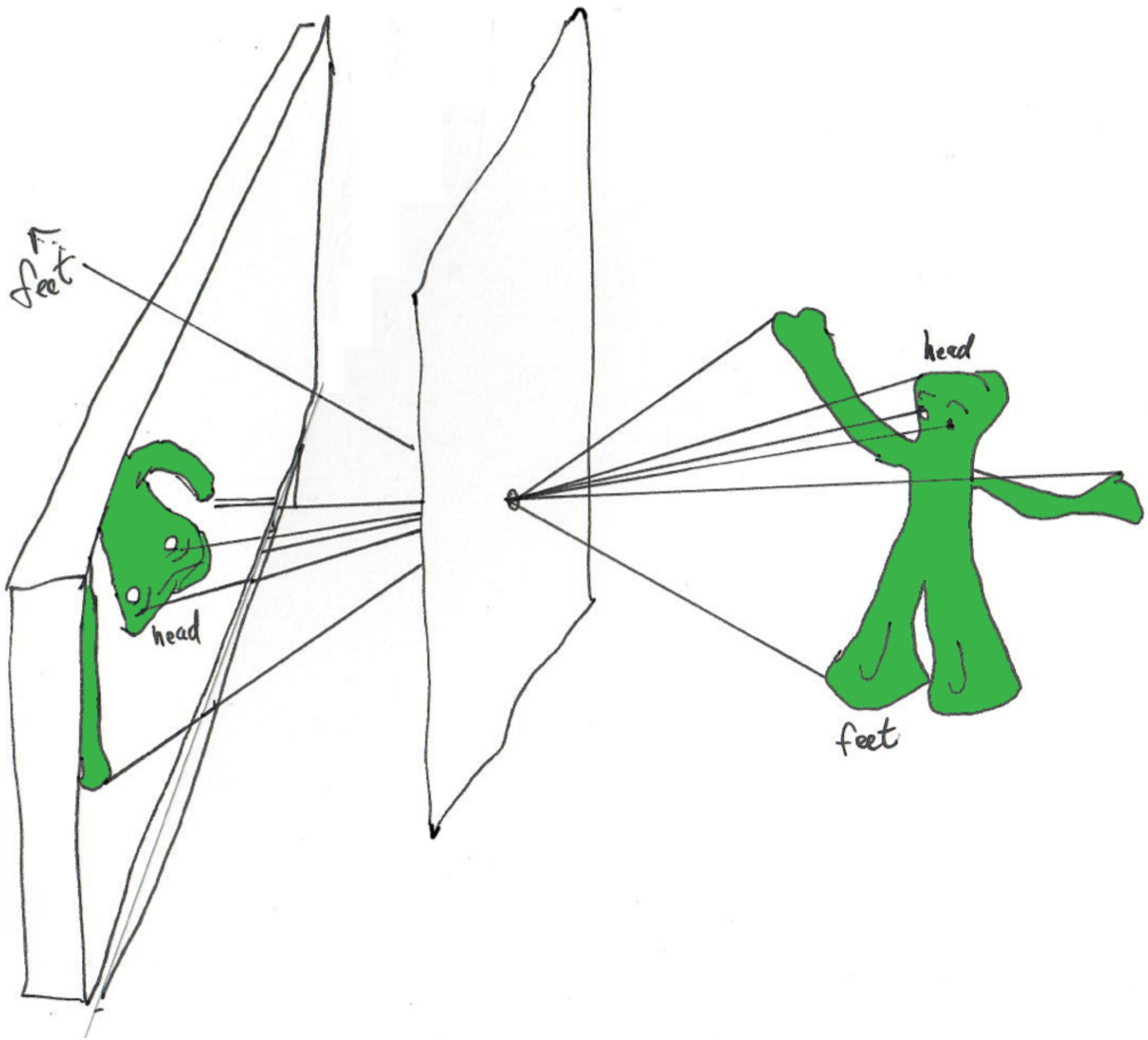
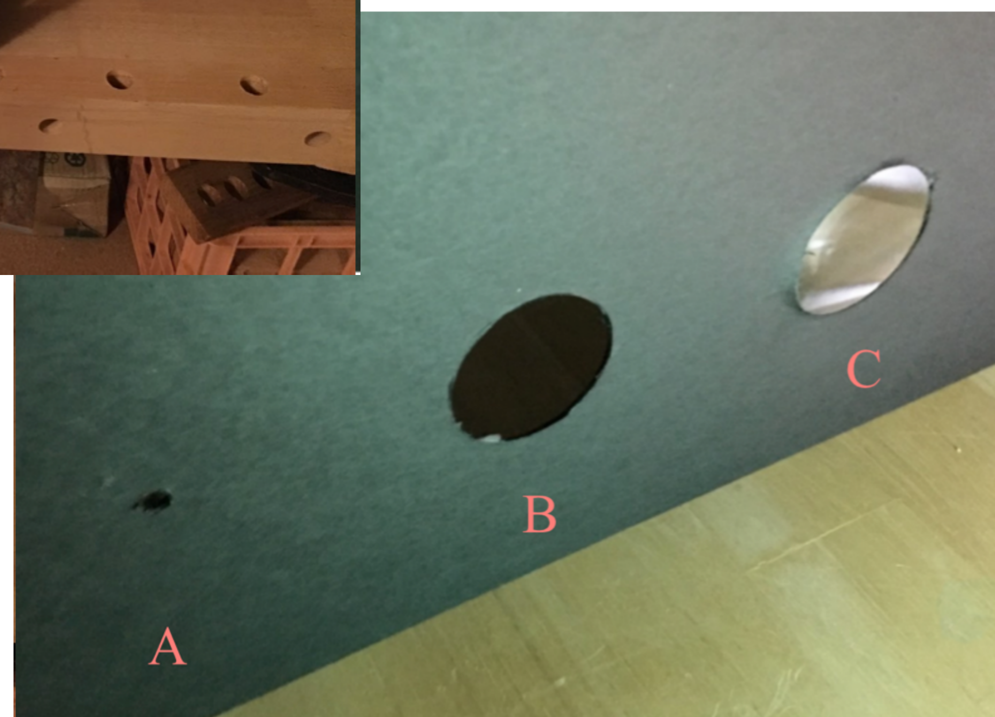


image is inverted



Let's try putting different occluders in between



light on wall past pinhole

grocery bag pinhole camera

grocery bag pinhole camera

grocery bag pinhole camera

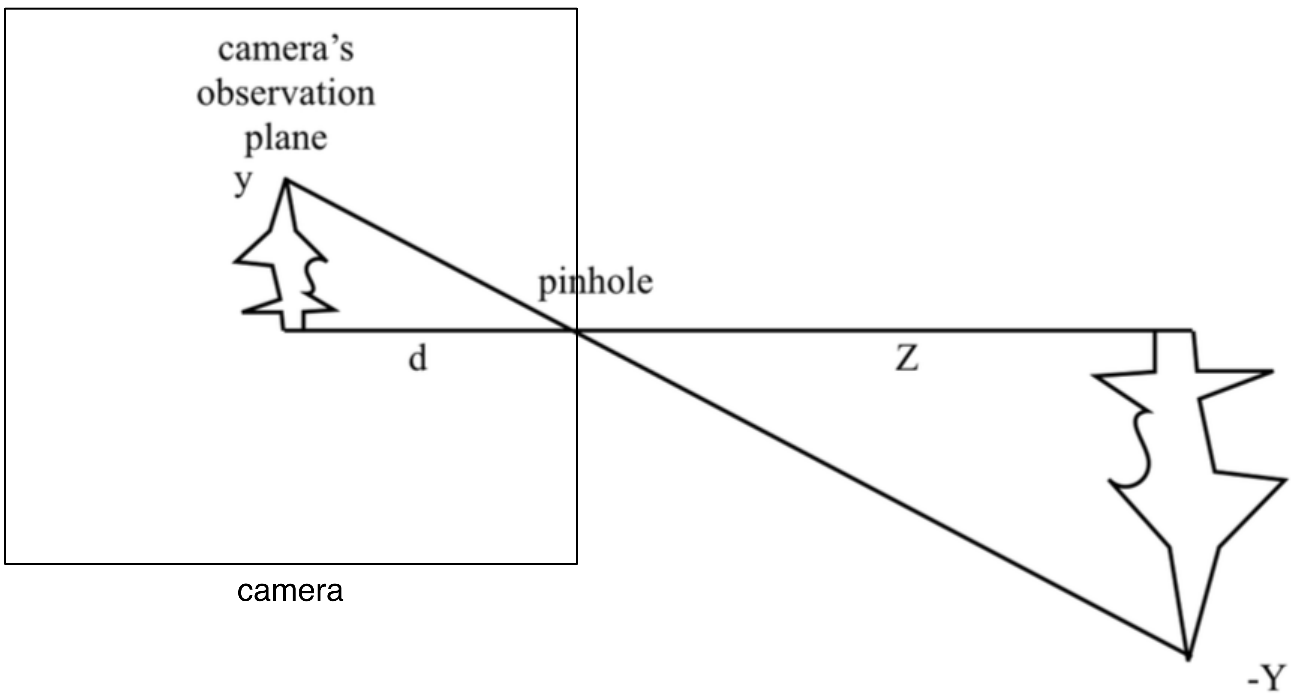
view from outside the bag

http://www.youtube.com/watch?v=FZyCFxsyx8o

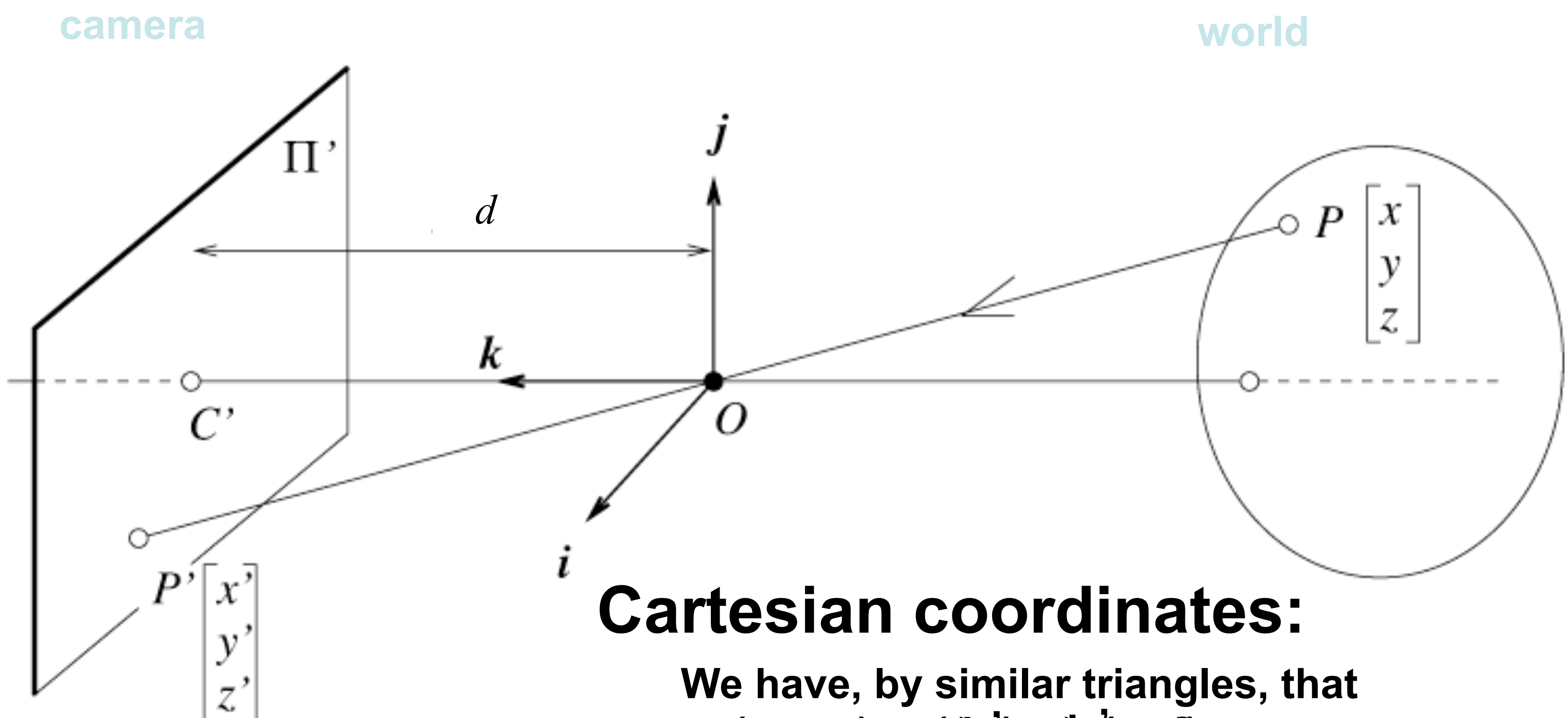
view from inside the bag

http://youtu.be/-rhZaAM3F44

Perspective projection



Perspective projection



Cartesian coordinates:

We have, by similar triangles, that
(x, y, z) -> (dx'/z, dy'/z, d')
Ignore the third coordinate, and get

$$(x,y,z) \rightarrow \left(d\,\frac{x'}{z},\ d\,\frac{y'}{z}\right)$$

Figure from: Forsyth and Ponce.

Geometric properties of projection

- Points go to points
- Lines go to lines
- Planes go to whole image or half-planes.
- Polygons go to polygons
- Degenerate cases
  - line through focal point to point
  - plane through focal point to line

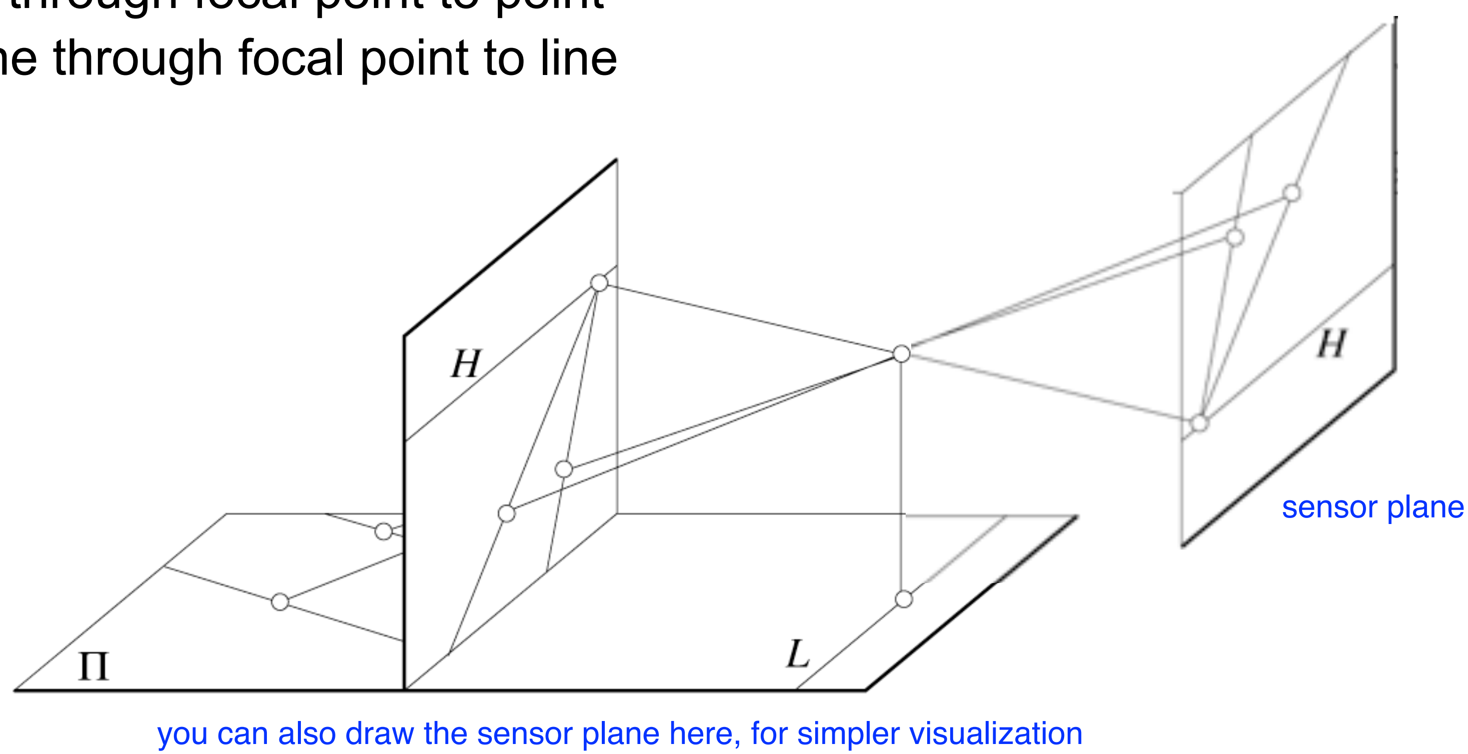
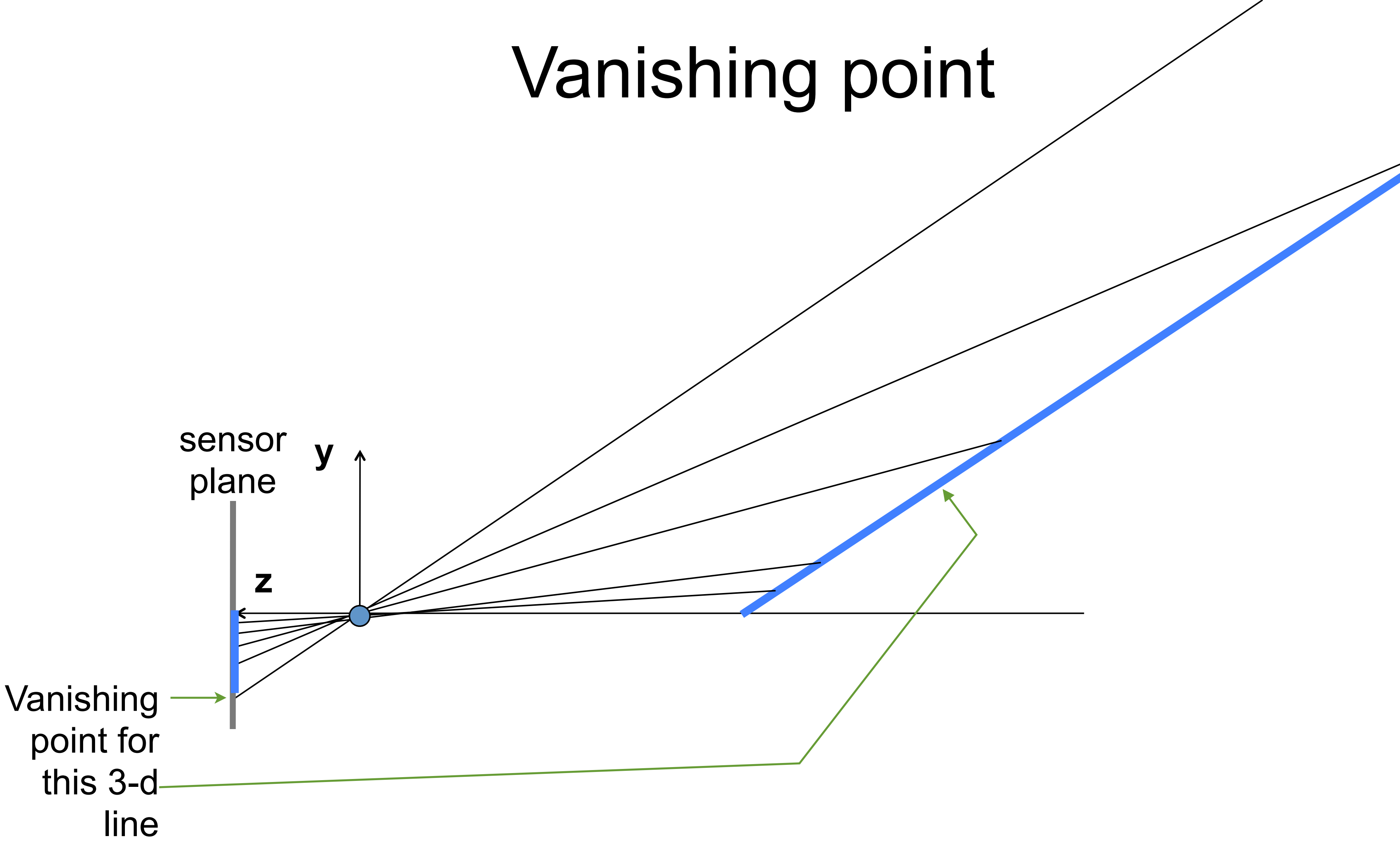


Figure from: Forsyth and Ponce.

Vanishing point



Line in 3-space

$$x(t) = x_0 + at$$
$$y(t) = y_0 + bt$$
$$z(t) = z_0 + ct$$

Perspective projection of that line

$$x'(t) = \frac{fx}{z} = \frac{f(x_0 + at)}{z_0 + ct}$$
$$y'(t) = \frac{fy}{z} = \frac{f(y_0 + bt)}{z_0 + ct}$$

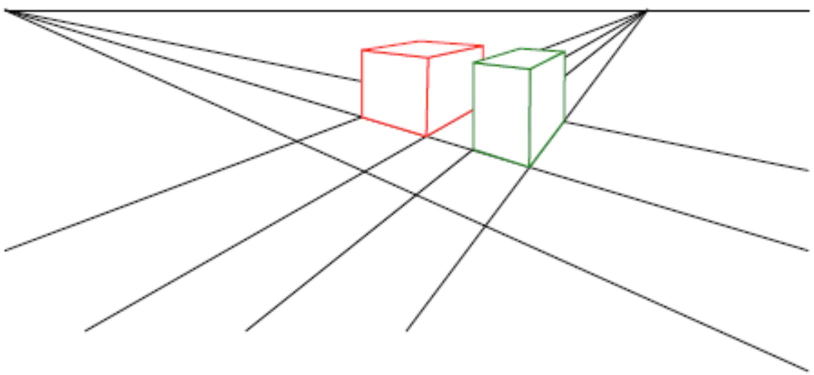
In the limit as $t \longrightarrow \pm\infty$ **we have (for** $c \neq 0$**):**

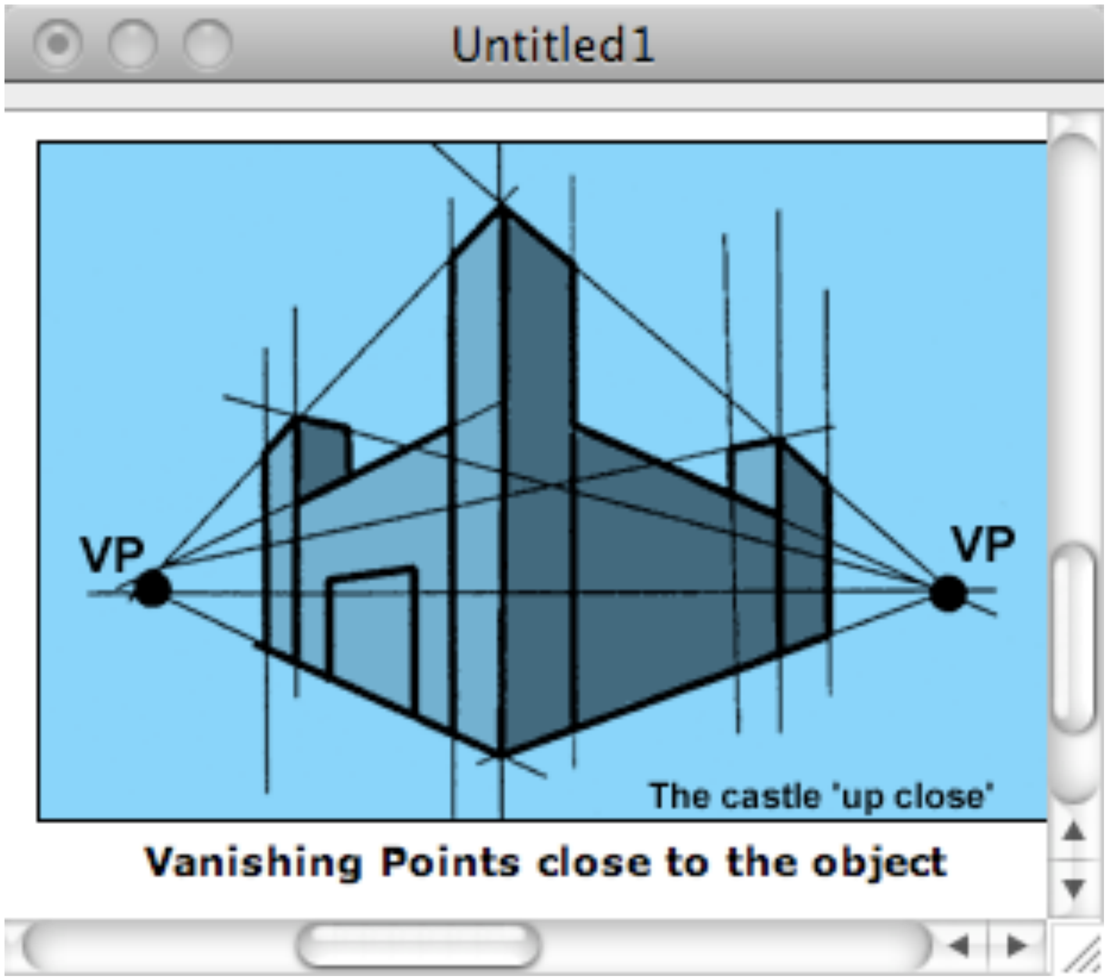
$$x'(t) \longrightarrow \frac{fa}{c}$$
$$y'(t) \longrightarrow \frac{fb}{c}$$

This tells us that any set of parallel lines (same a, b, c parameters) project to the same point (called the vanishing point).

Vanishing points

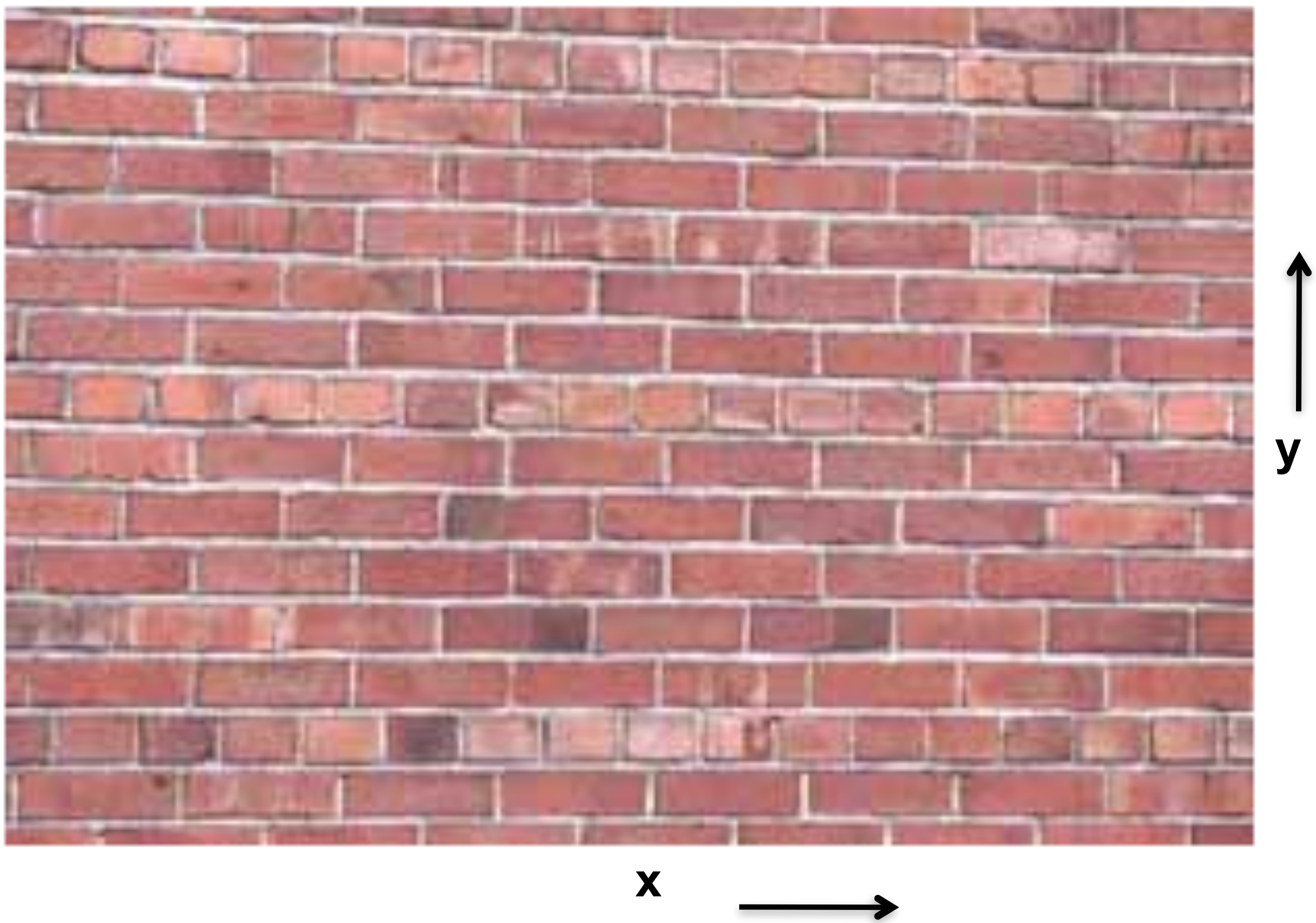
- Each set of parallel lines (=direction) meets at a different point
  - The *vanishing point* for this direction
- Sets of parallel lines on the same plane lead to *collinear* vanishing points.
  - The line is called the *horizon* for that plane





http://www.ider.herts.ac.uk/school/courseware/graphics/two_point_perspective.html

What if you photograph a brick wall head-on?



Brick wall line in 3-space

$$x(t) = x_0 + at$$

$$y(t) = y_0$$

$$z(t) = z_0$$

Perspective projection of that line

$$x'(t) = \frac{f \cdot (x_0 + at)}{z_0}$$

$$y'(t) = \frac{f \cdot y_0}{z_0}$$

All bricks have same z_0. Those in same row have same y_0

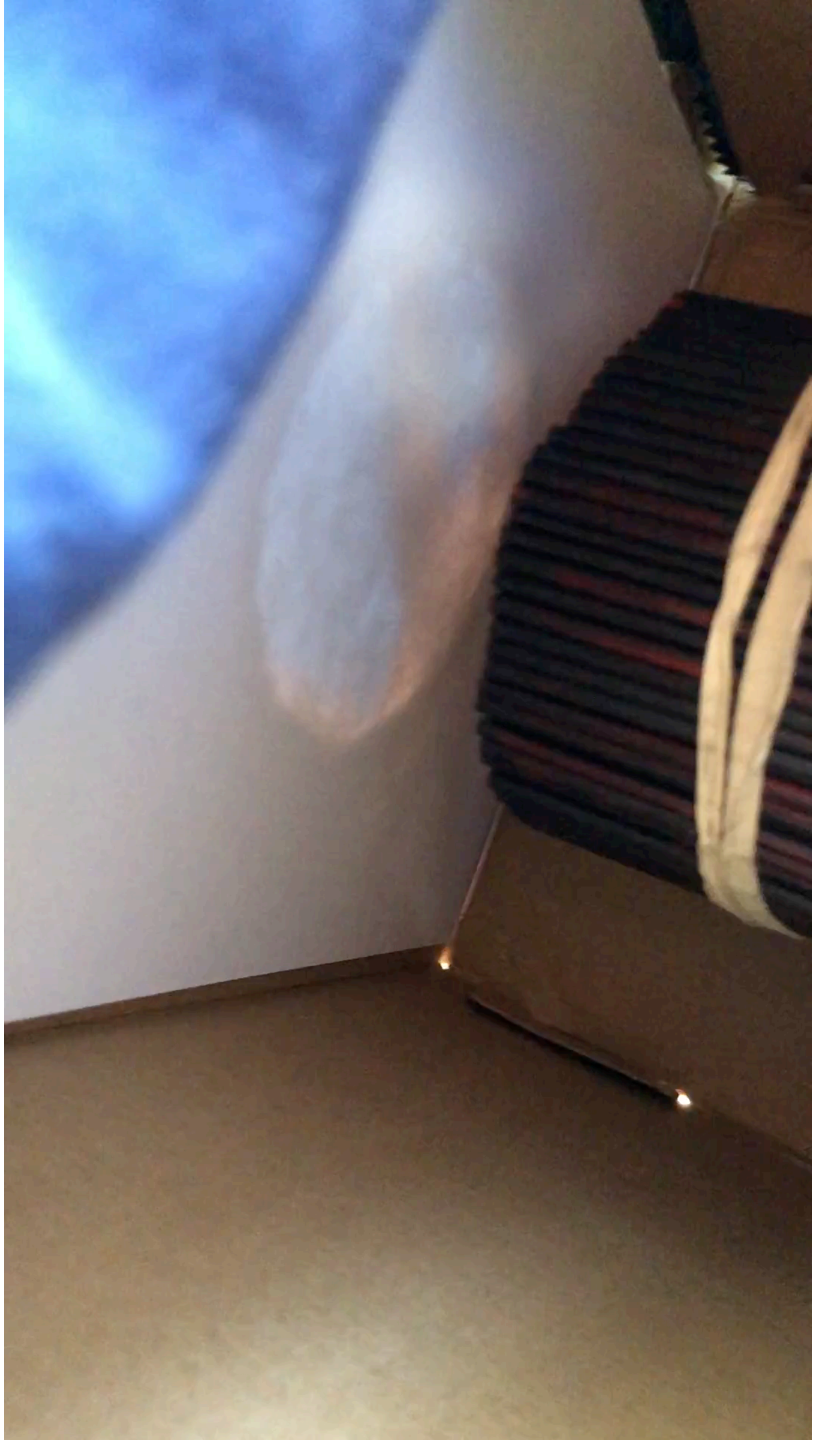
Thus, a brick wall, photographed head-on, gets rendered as set of parallel lines in the image plane.

Straw camera

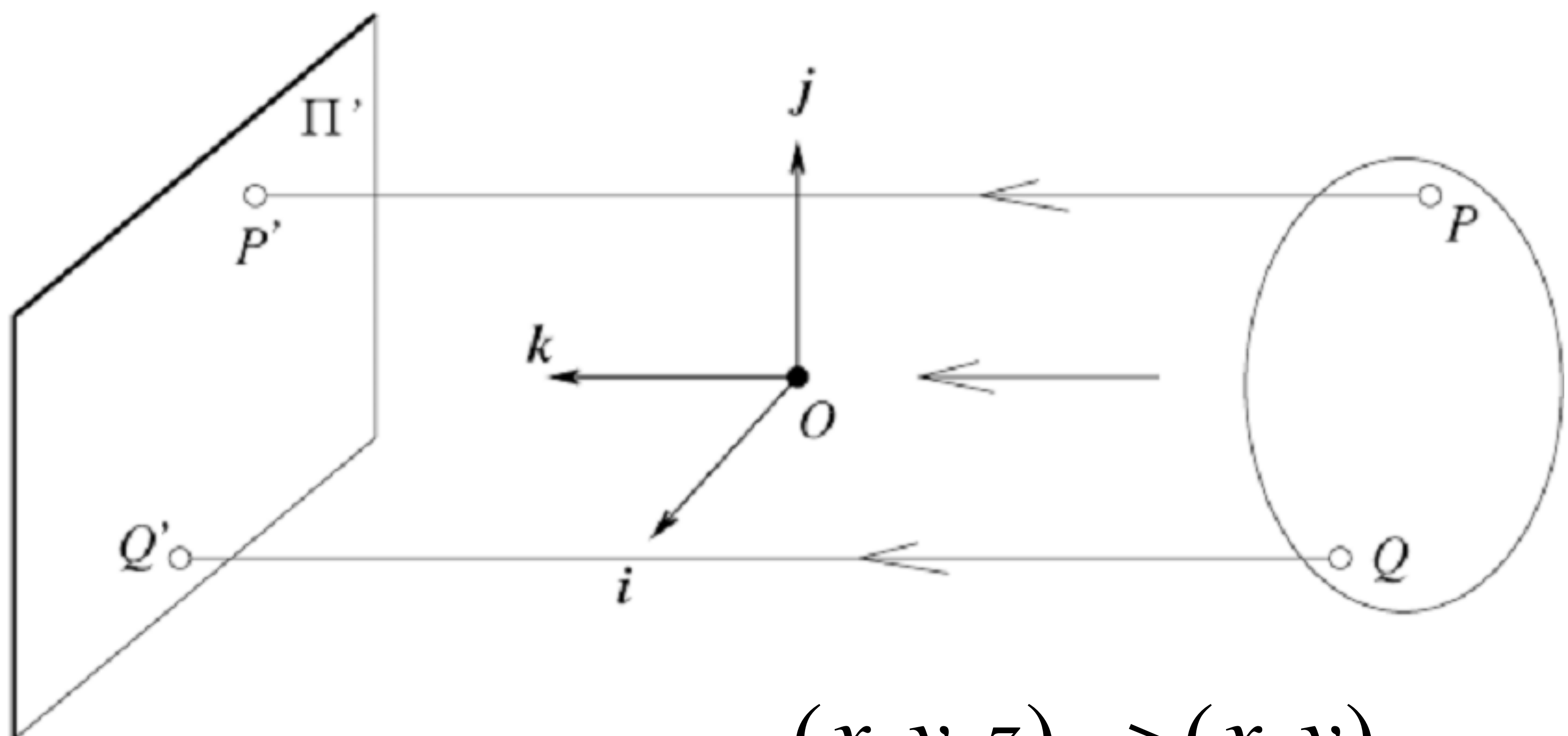
(a)

(b)

Straw camera



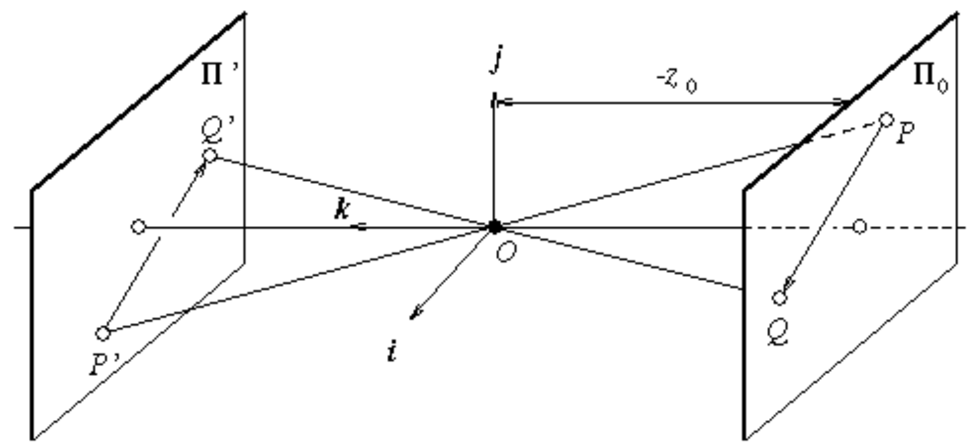
Other projection models: Orthographic projection



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

Other projection models: Weak perspective

- Issue
  - perspective effects, but not over the scale of individual objects
  - collect points into a group at about the same depth, then divide each point by the depth of its group
  - Adv: easy
  - Disadv: only approximate



$$(x, y, z) \rightarrow \left(\frac{fx}{z_0}, \frac{fy}{z_0} \right)$$

Three camera projections

3-d point → **2-d image position**

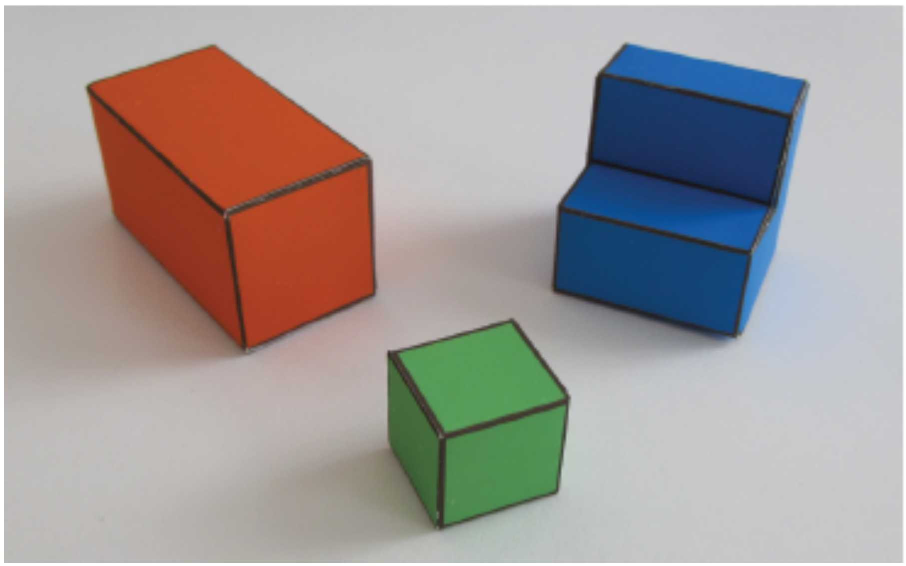
(1) Perspective: $(x, y, z) \rightarrow \left(\frac{fx}{z}, \frac{fy}{z} \right)$

(2) Weak perspective: $(x, y, z) \rightarrow \left(\frac{fx}{z_0}, \frac{fy}{z_0} \right)$

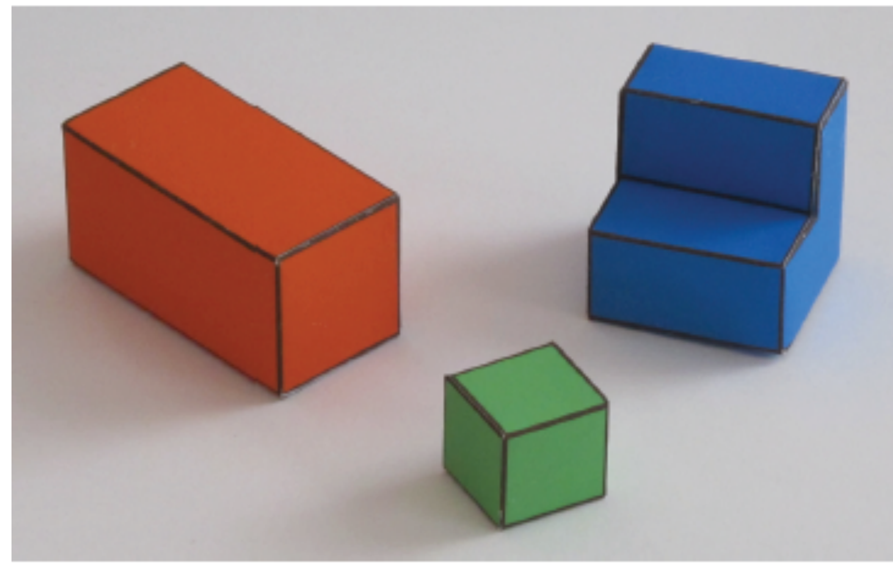
(3) Orthographic: $(x, y, z) \rightarrow (x, y)$

which is perspective, which orthographic?

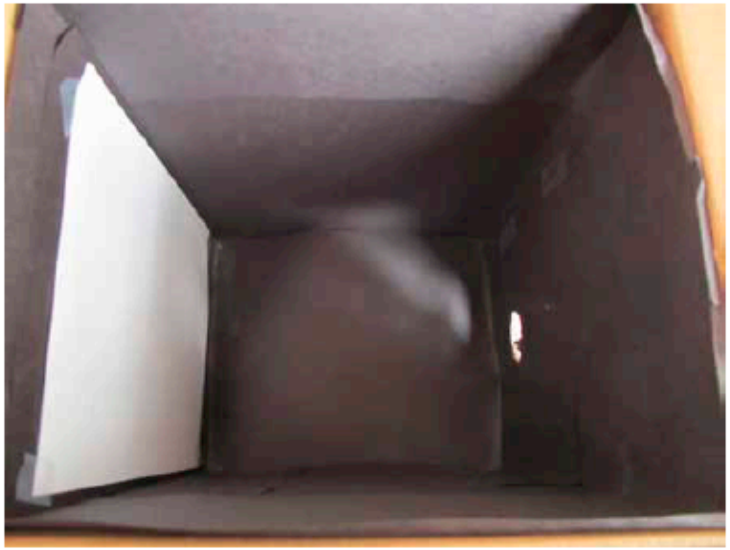
Perspective projection

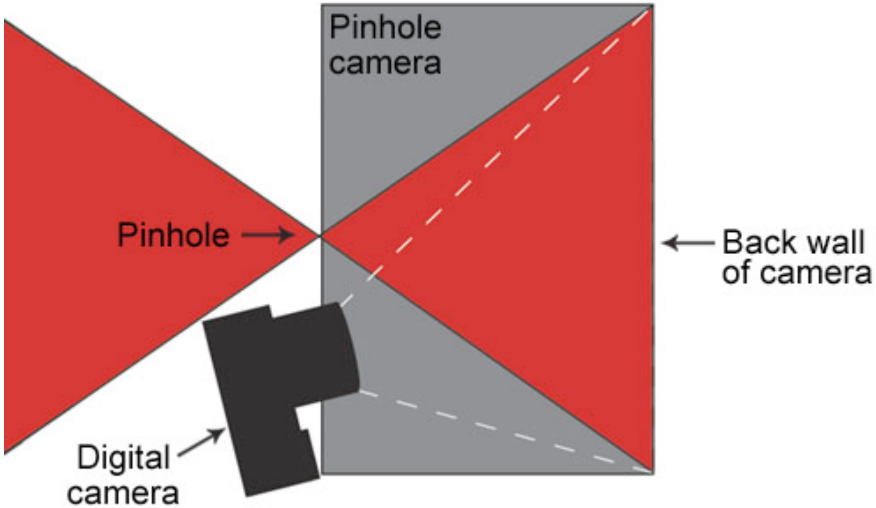


Parallel (orthographic) projection



Problem Set 2

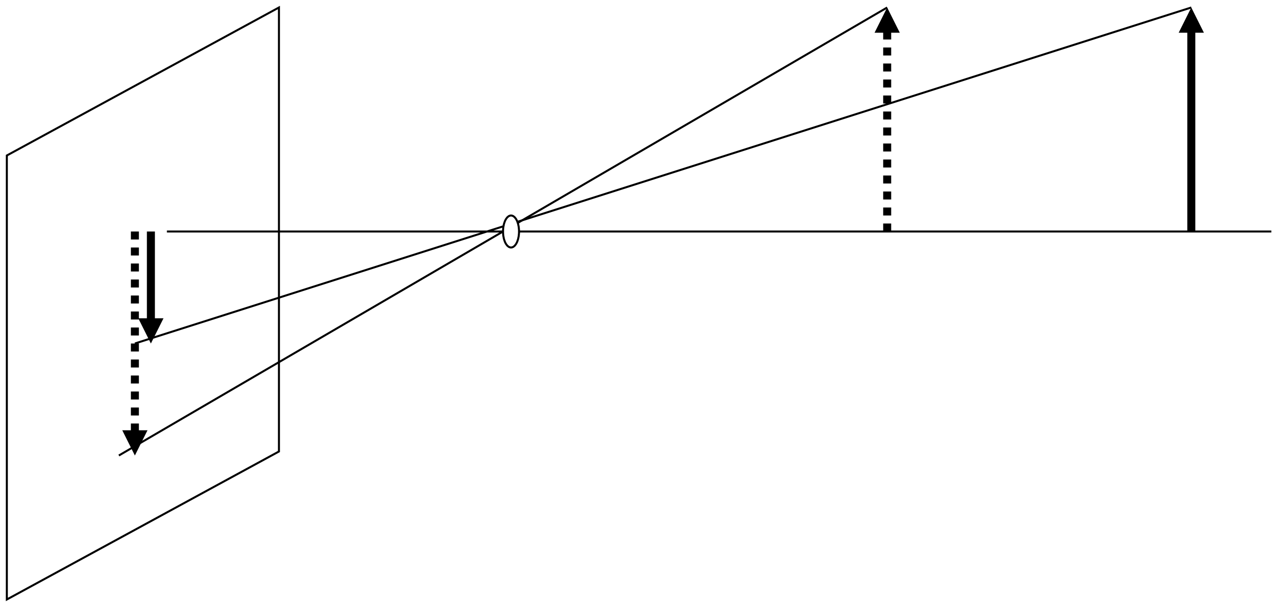




http://www.foundphotography.com/PhotoThoughts/archives/2005/04/pinhole_camera_2.html

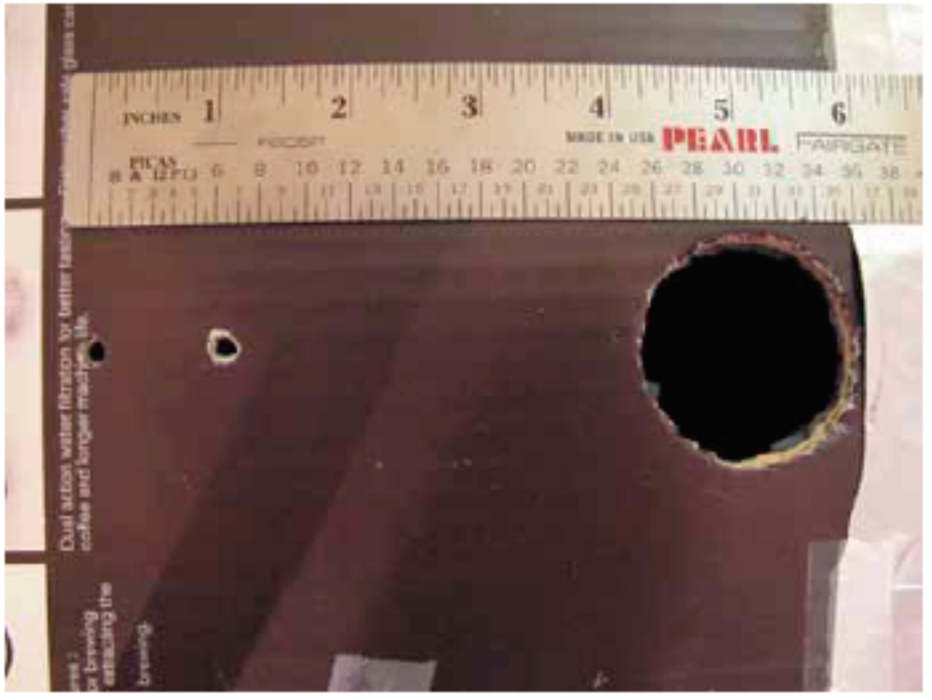
Example images from pinhole camera

Measuring distance

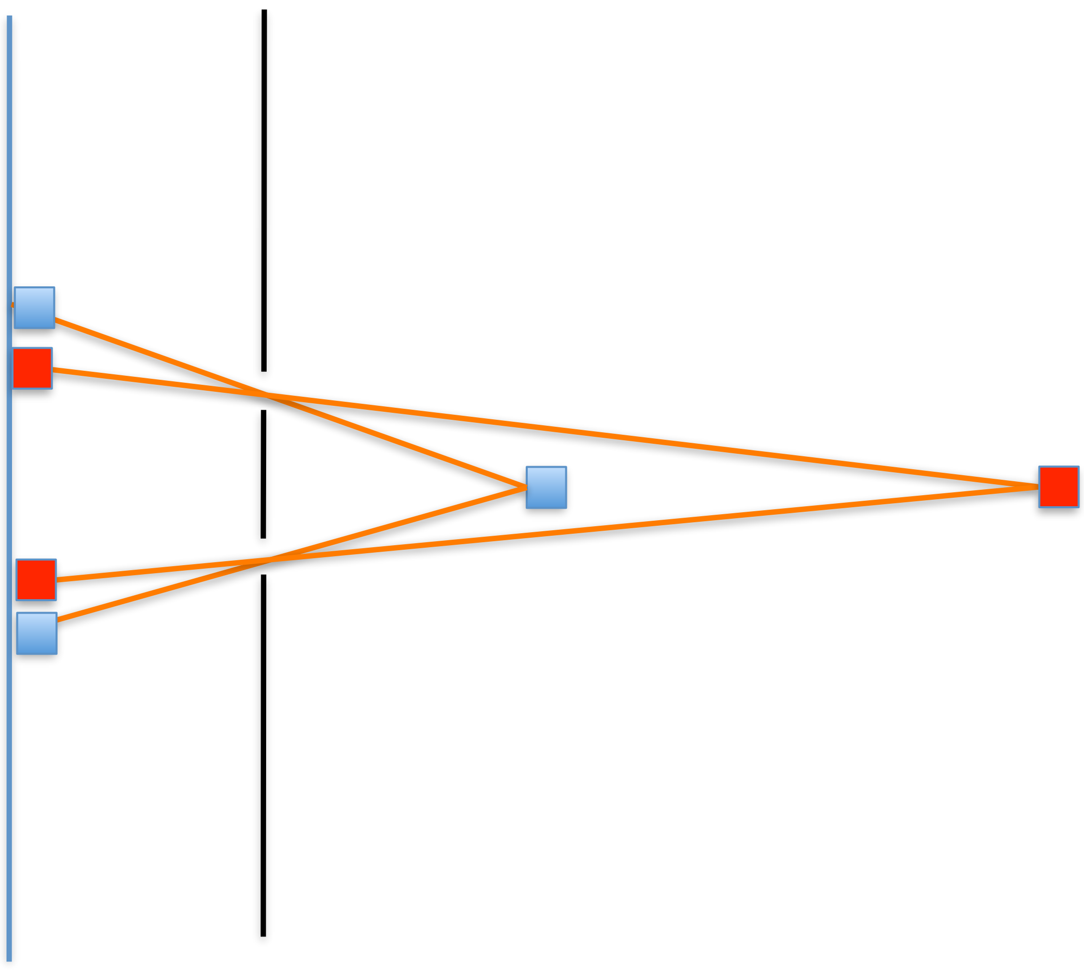


- Object size decreases with distance to the pinhole
- There, given a single projection, if we know the size of the object we can know how far it is.
- But for objects of unknown size, the 3D information seems to be lost.

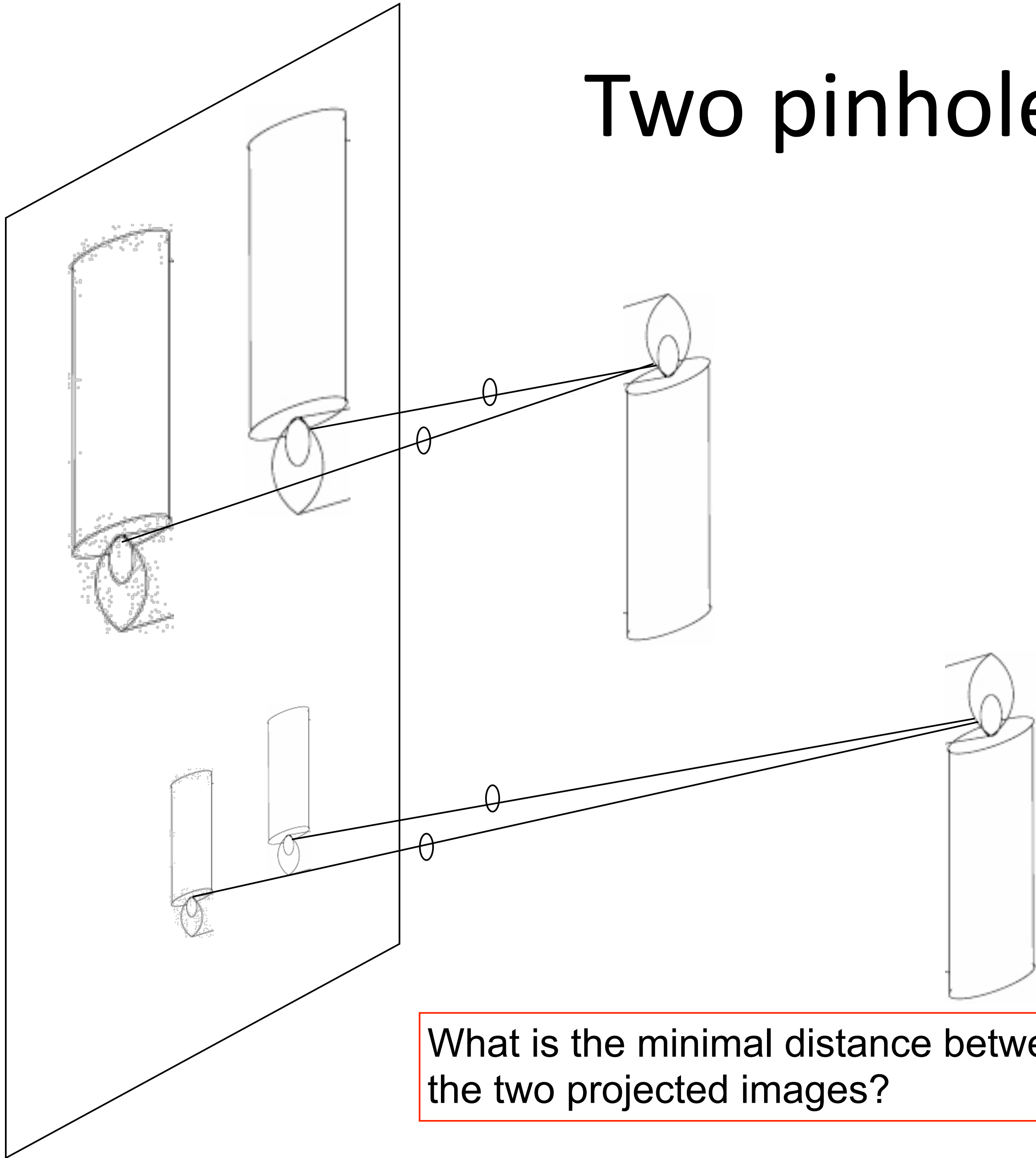
Playing with pinholes



Two pinholes

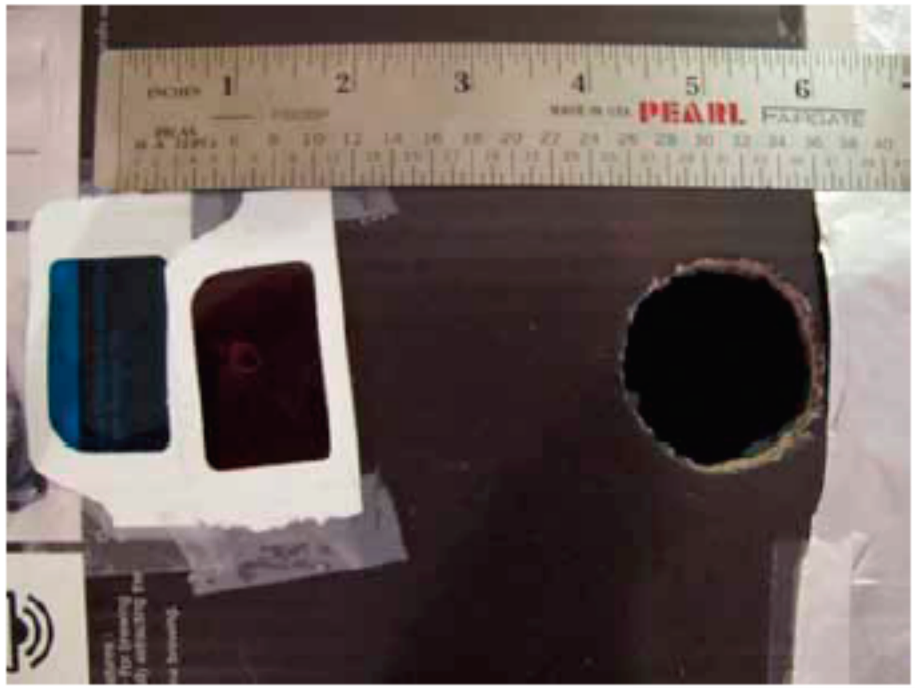


Two pinholes

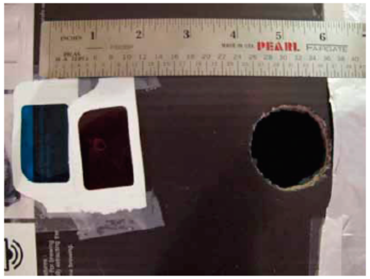


What is the minimal distance between the two projected images?

Anaglyph pinhole camera



Anaglyph pinhole camera



front of camera

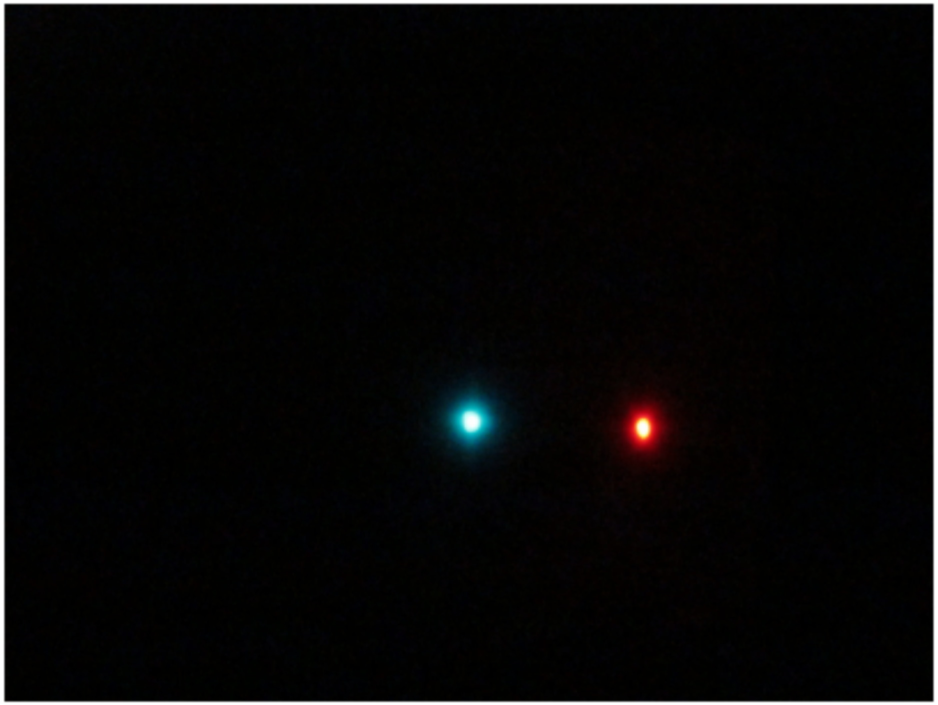
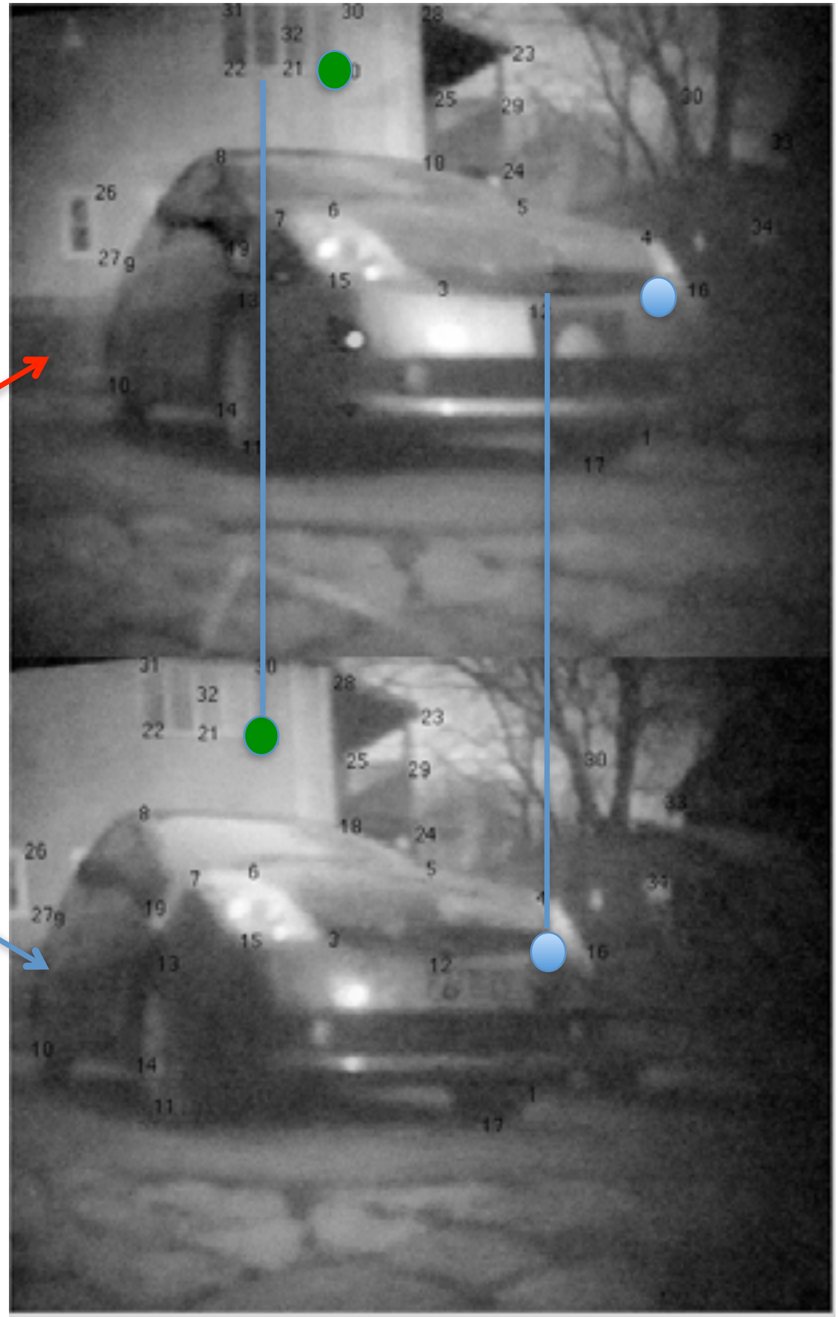


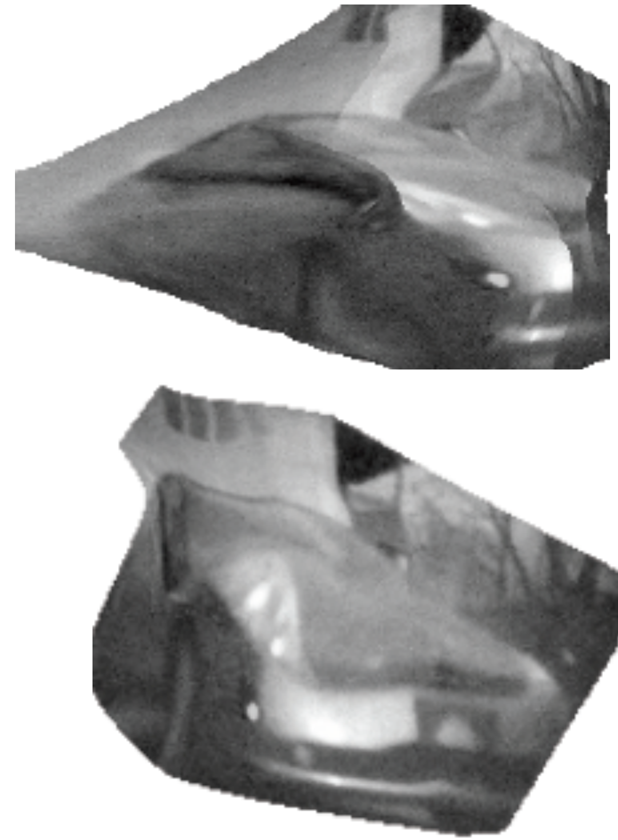
image of a point of light

Anaglyph pinhole camera

Anaglyph



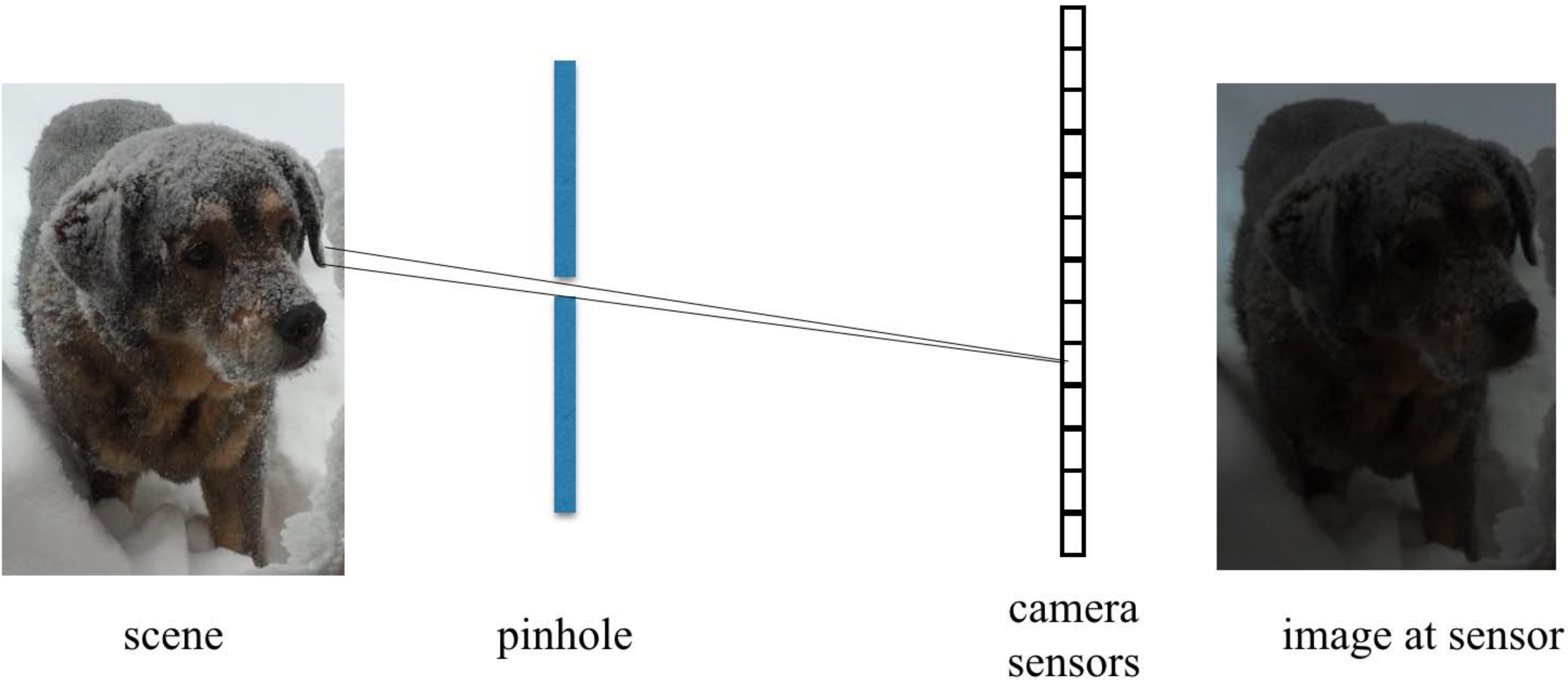
Synthesis of new views



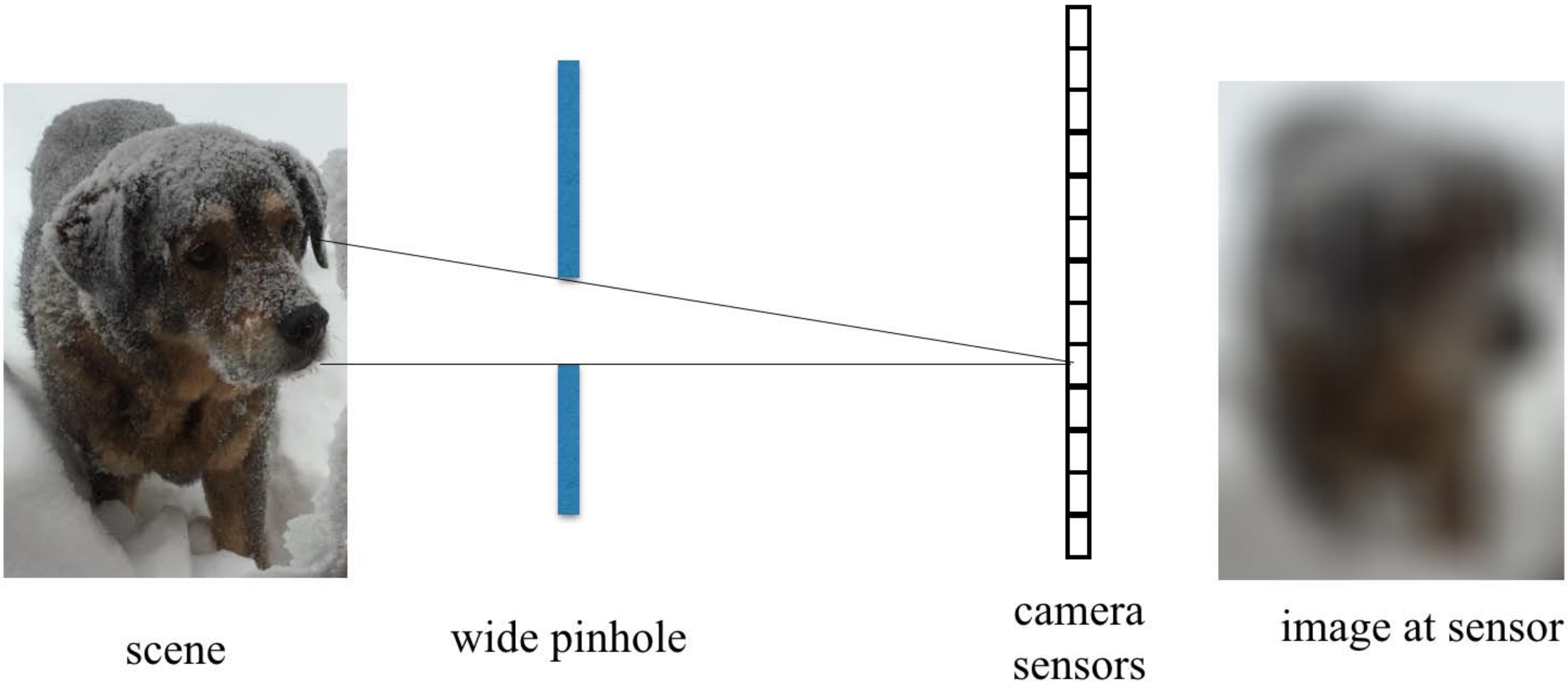
Problem set

- Build the device
- Take some pictures and put them in the report
- Take anaglyph images
- Work out the geometry
- Recover depth for some points in the image

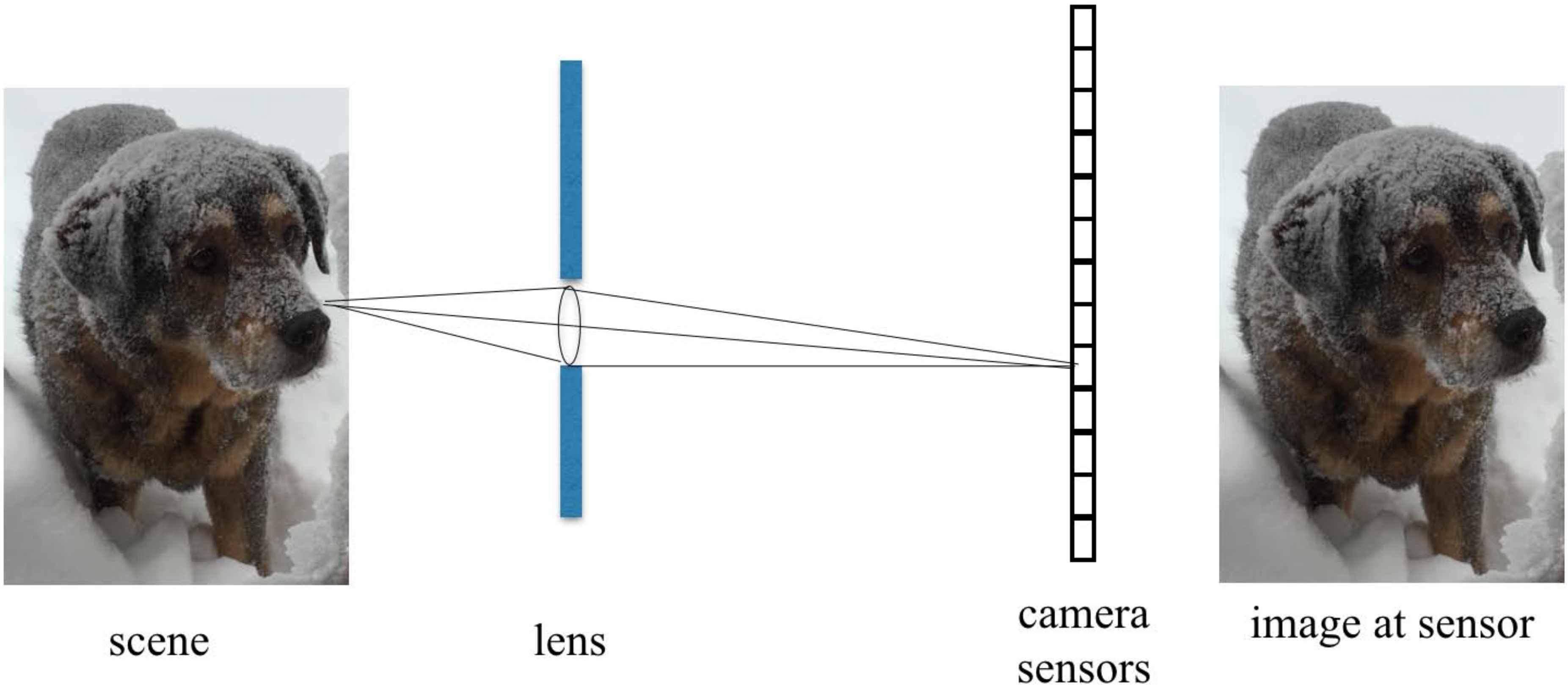
A problem: pinhole camera images are dark, or require long exposures



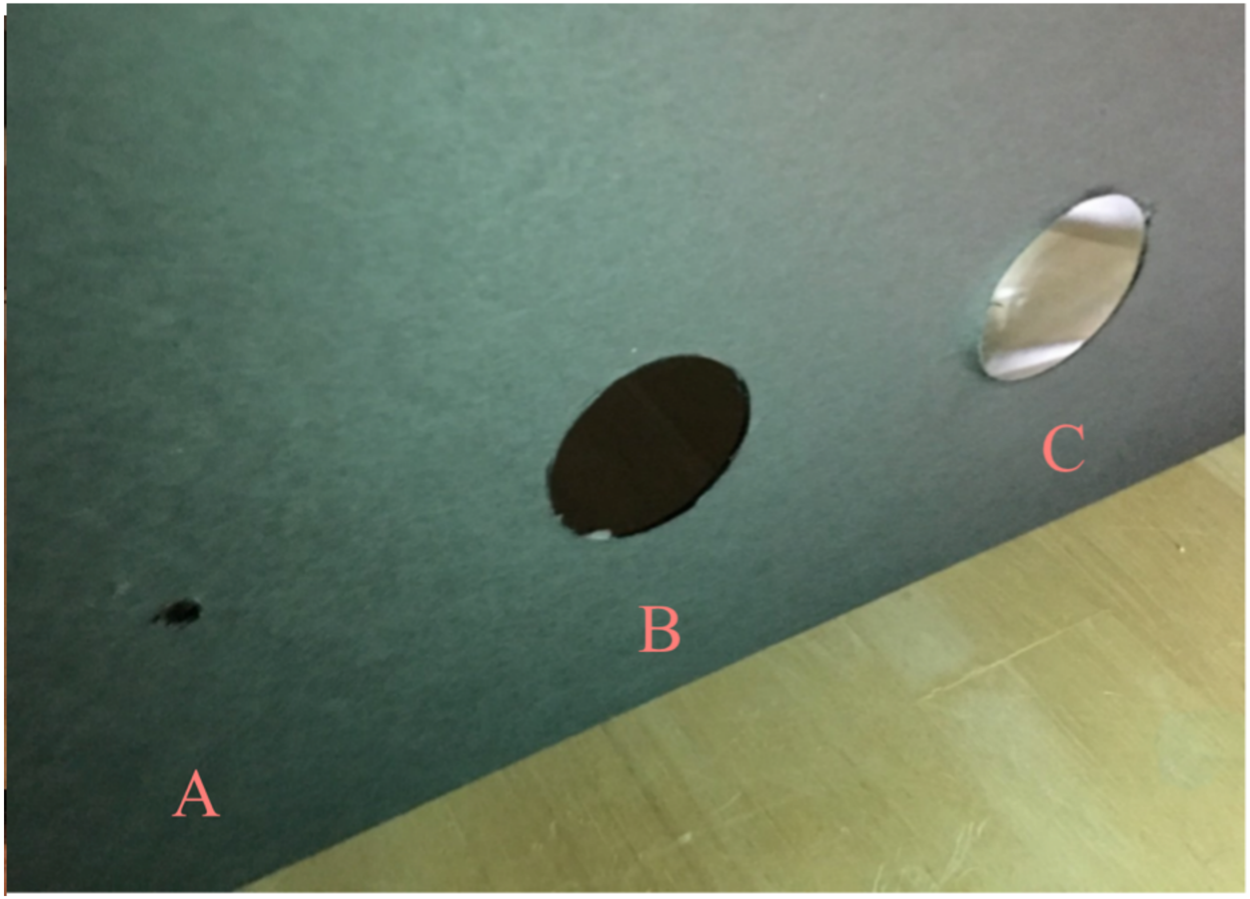
Large aperture gives a brighter image, but at the price of sharpness



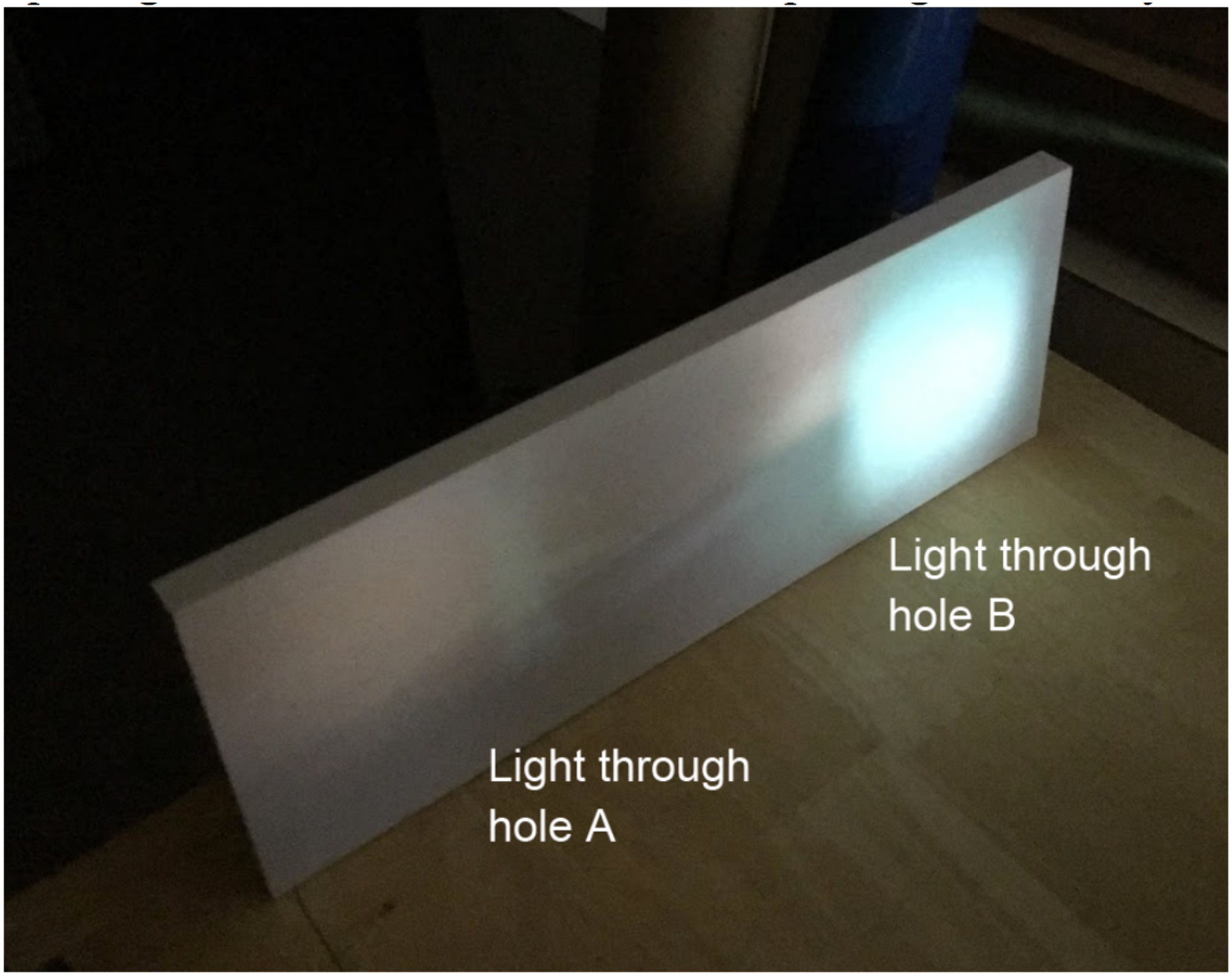
A lens allows a large aperture and a sharp image



Let's try putting different occluders in between the scene and the sensor plane

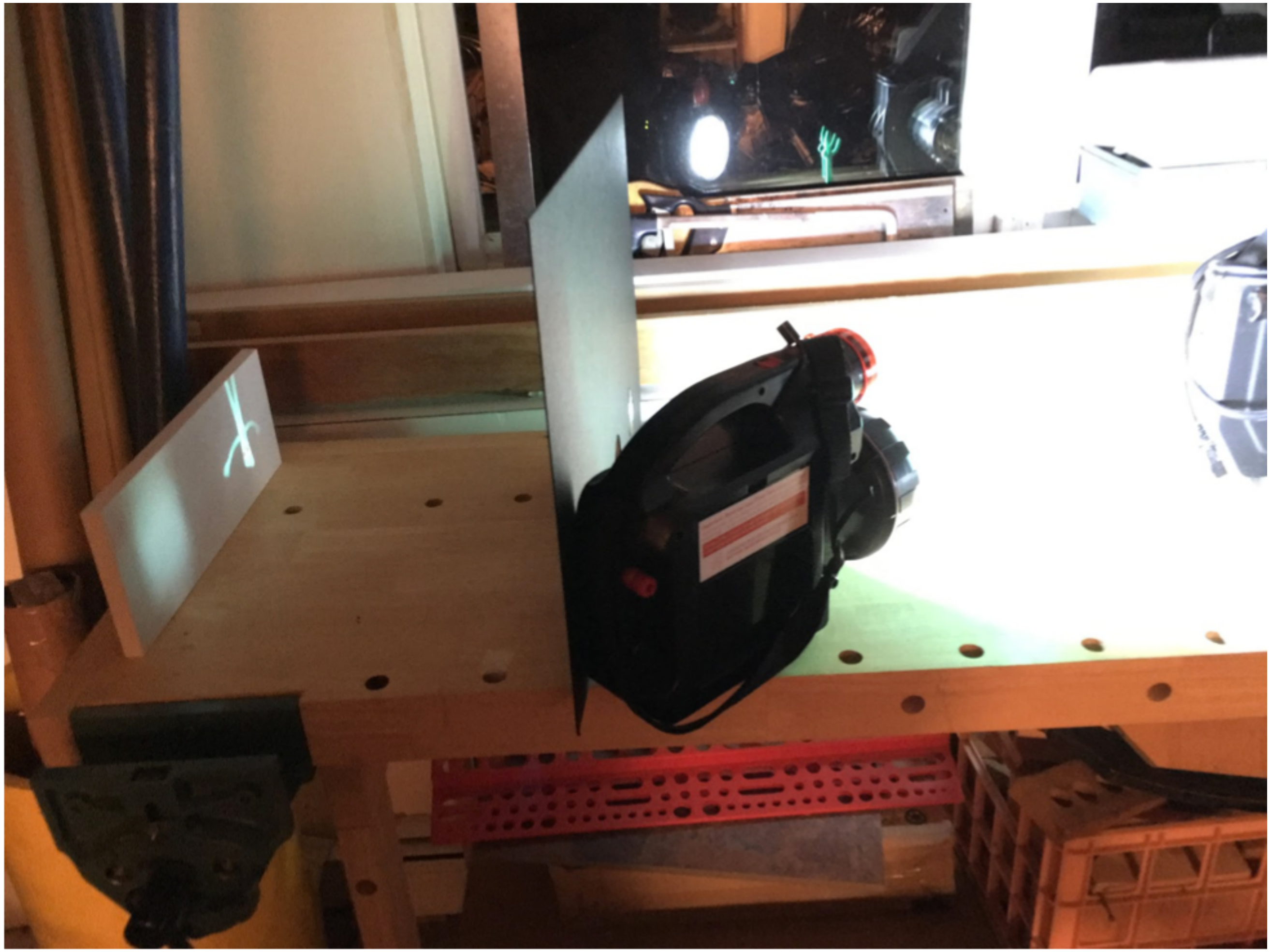


Influence of aperture size: with a small aperture, the image is sharp, but dim. A large aperture gives a bright, but blurry image.

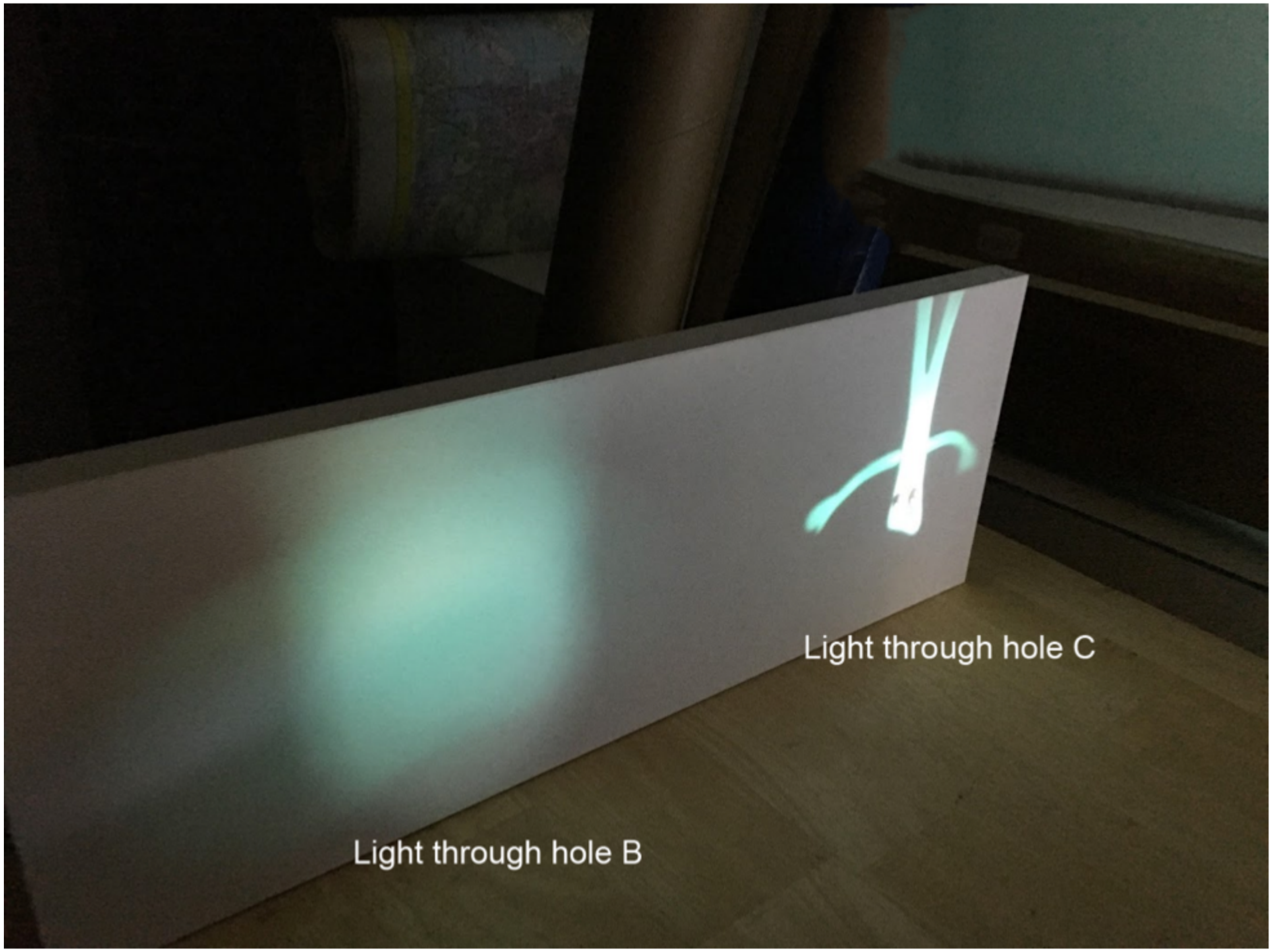


A lens can focus light from one point in the world to one point on the sensor plane.

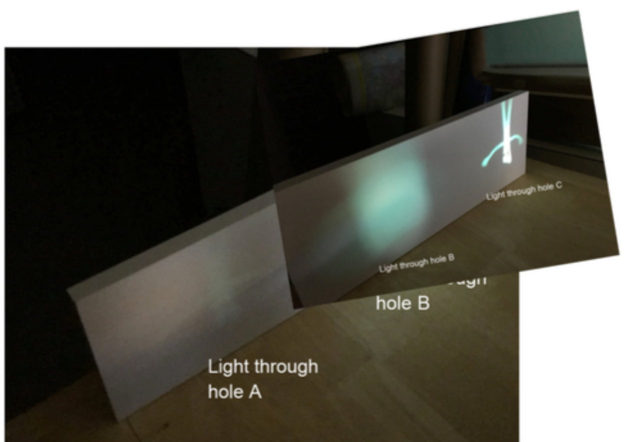
Images through large aperture, with and without lens present



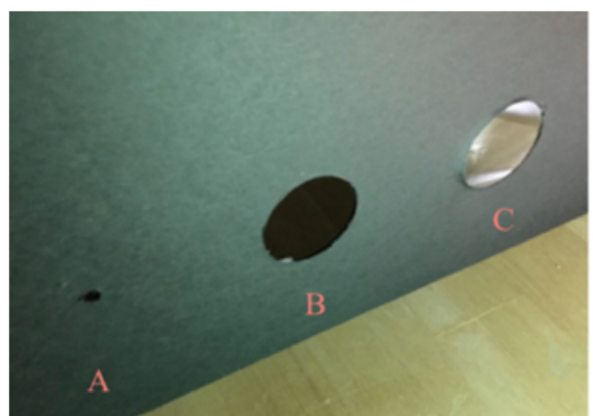
Images through large aperture, with and without lens present



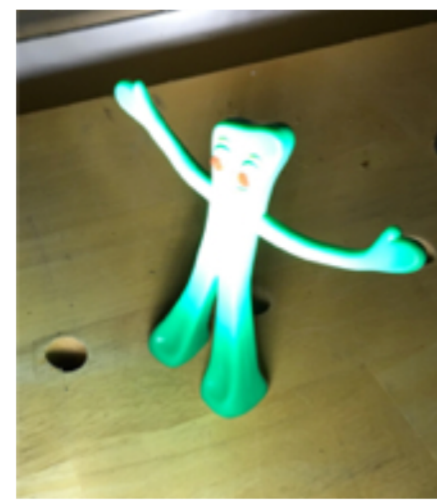
(a)


Light through hole C
Light through hole B
hole B
Light through
hole A

(b)

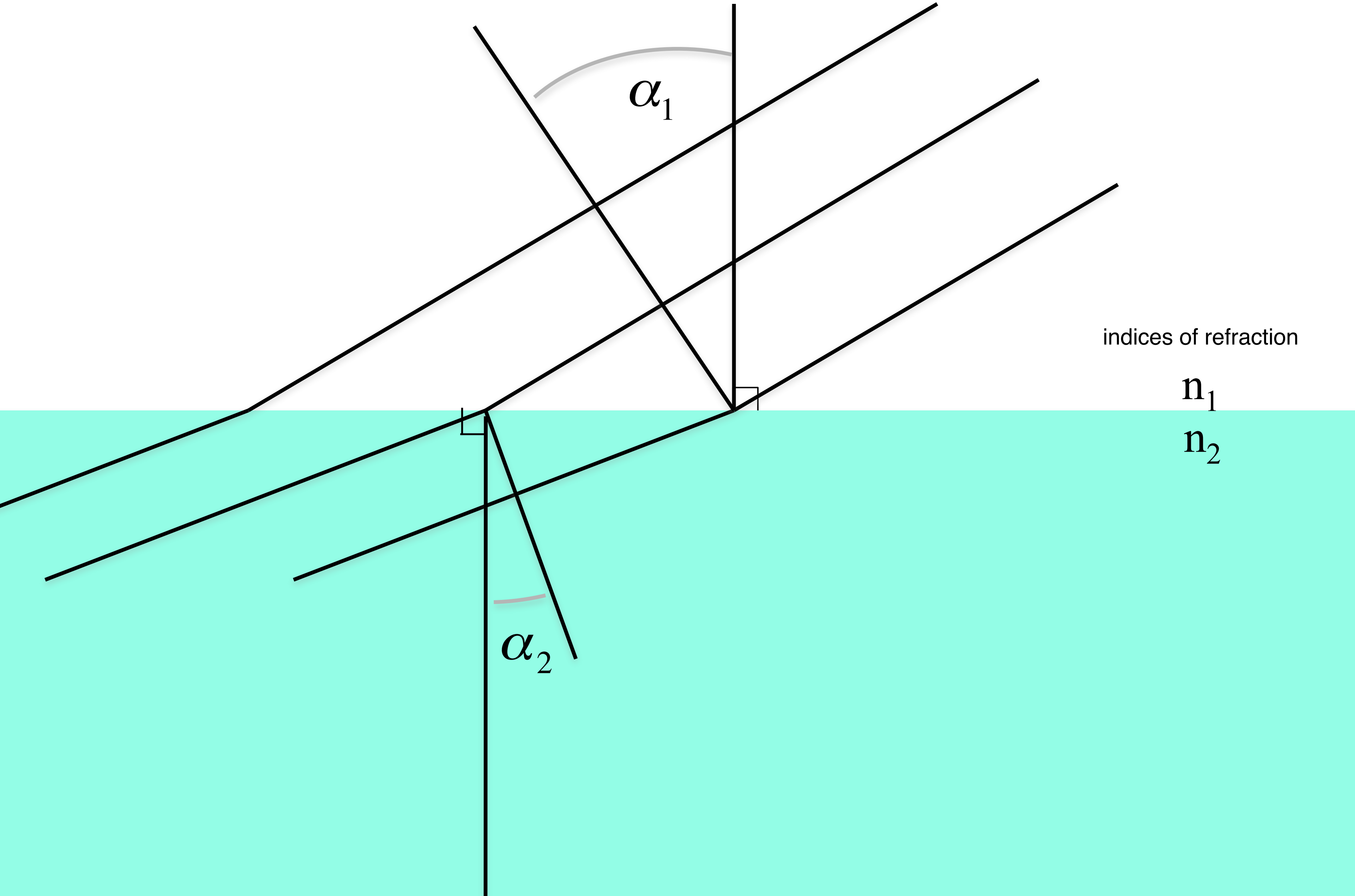

C
B
A

(c)



(d)

Light at a material interface



Light at a material interface

$$\lambda_1 = \frac{c}{\omega\, n_1}$$ wavelength is speed/ freq

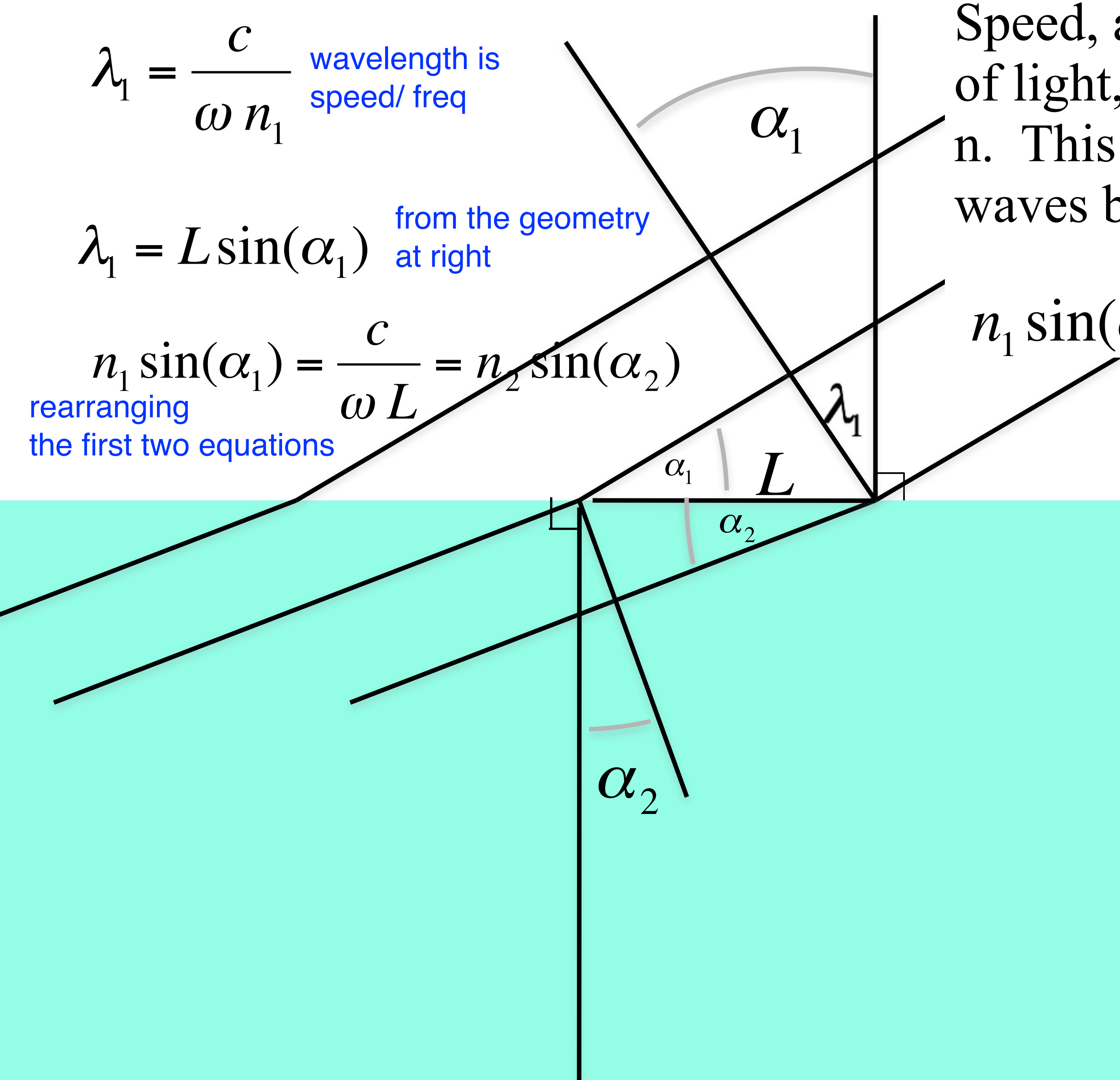
$$\lambda_1 = L \sin(\alpha_1)$$ from the geometry at right

$$n_1 \sin(\alpha_1) = \frac{c}{\omega L} = n_2 \sin(\alpha_2)$$

rearranging the first two equations

Speed, and thus wavelength of light, scales inversely with n. This requires that plane waves bend, according to

$$n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2)$$

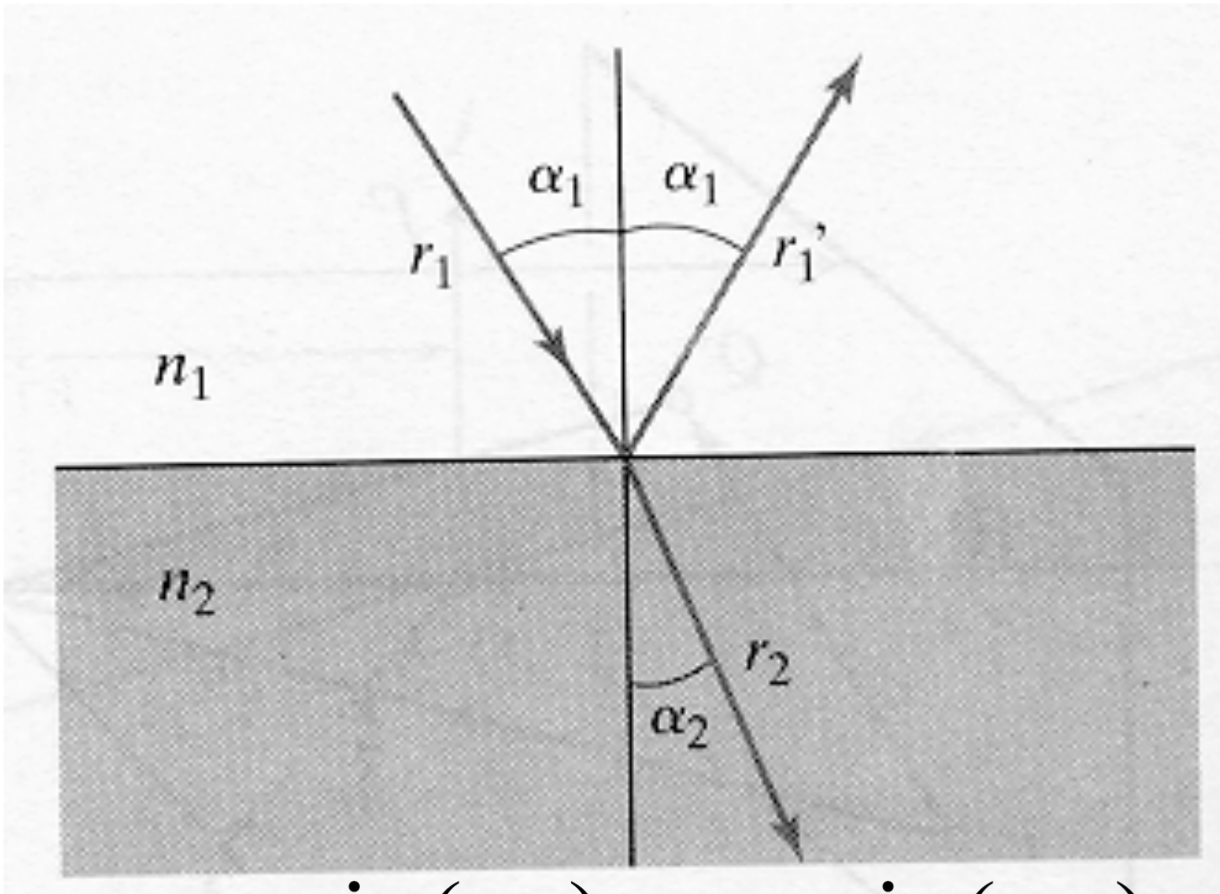


indices of refraction

n_1

n_2

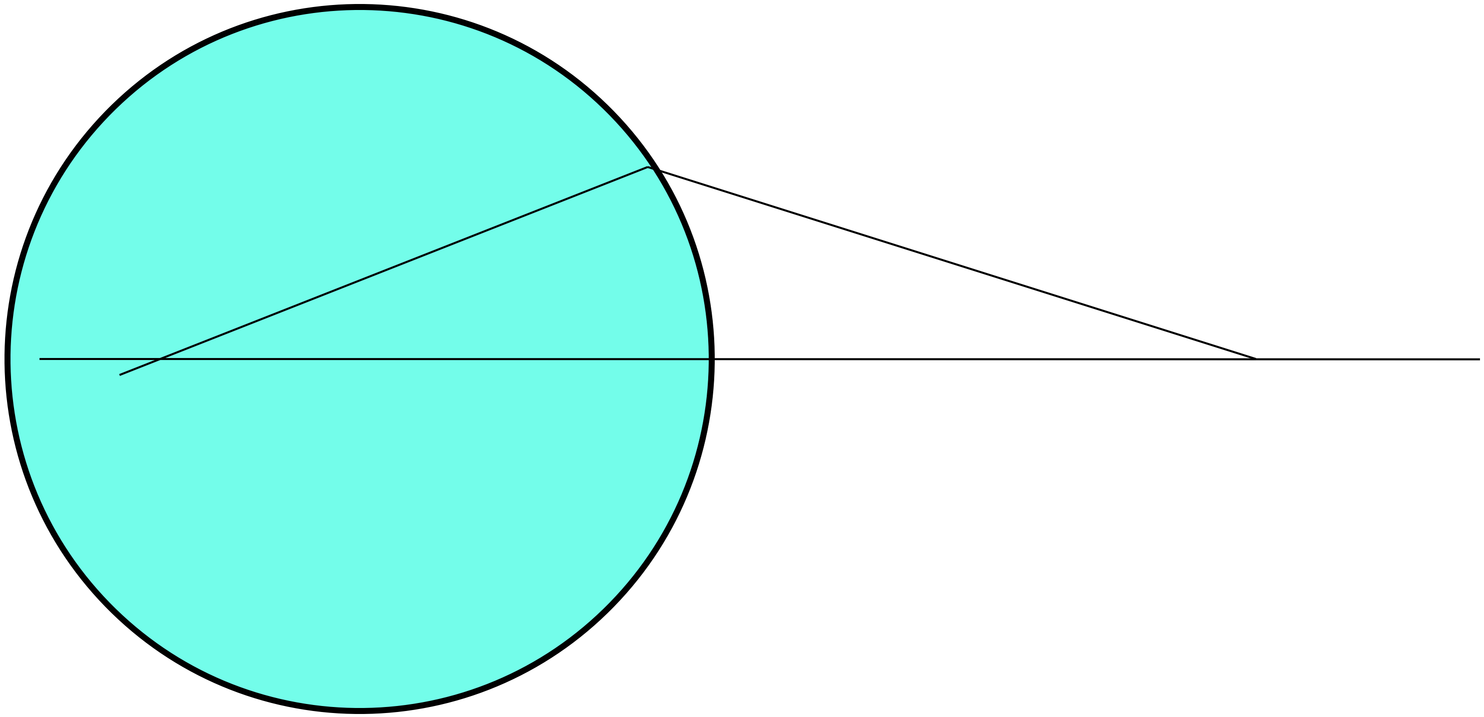
Refraction: Snell's law



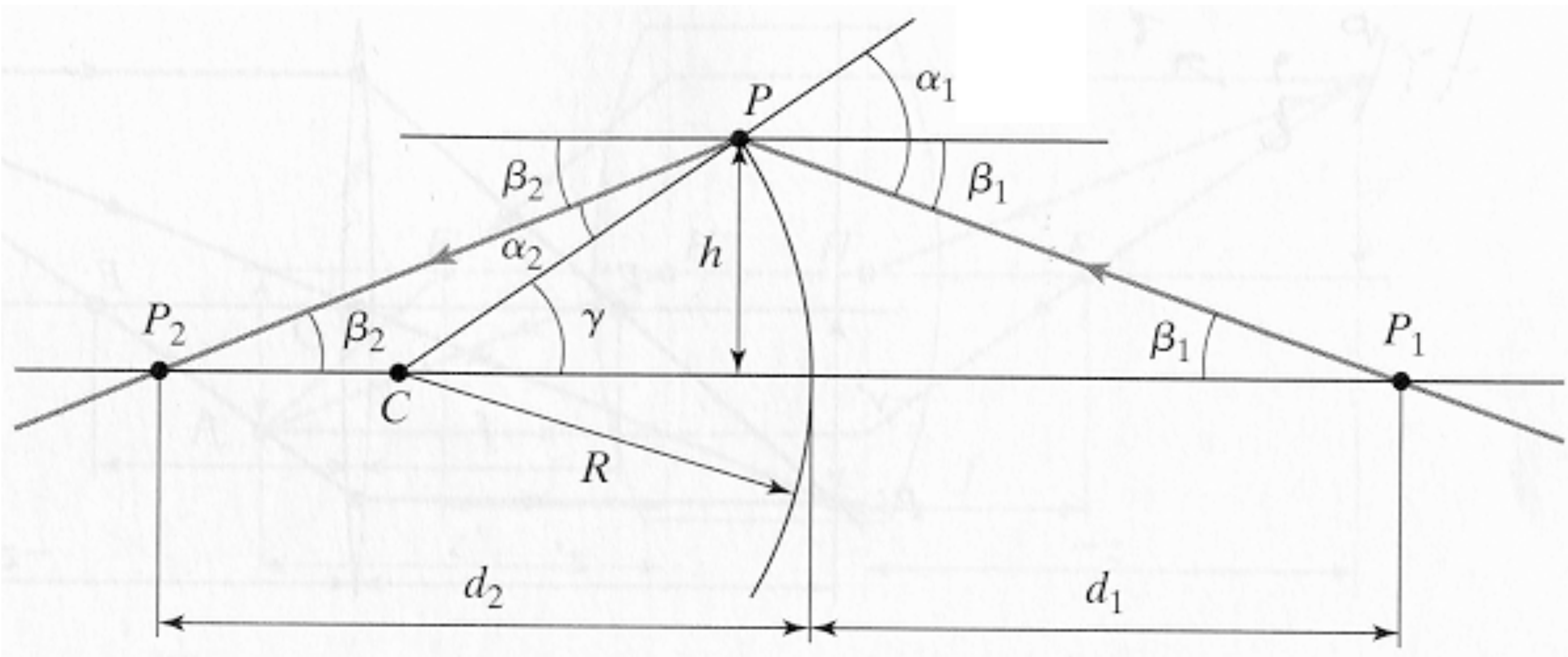
$$n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2)$$

For small angles, $n_1\alpha_1 \approx n_2\alpha_2$

Spherical lens



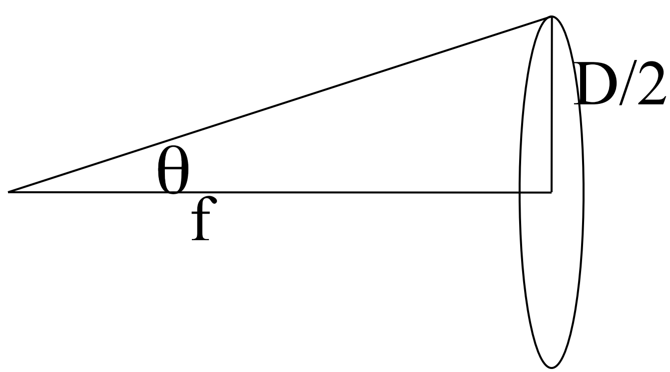
For a spherical lens surface, we can define the relevant angles, apply Snell's law, and find an expression telling how the lens focusses light



Forsyth and Ponce

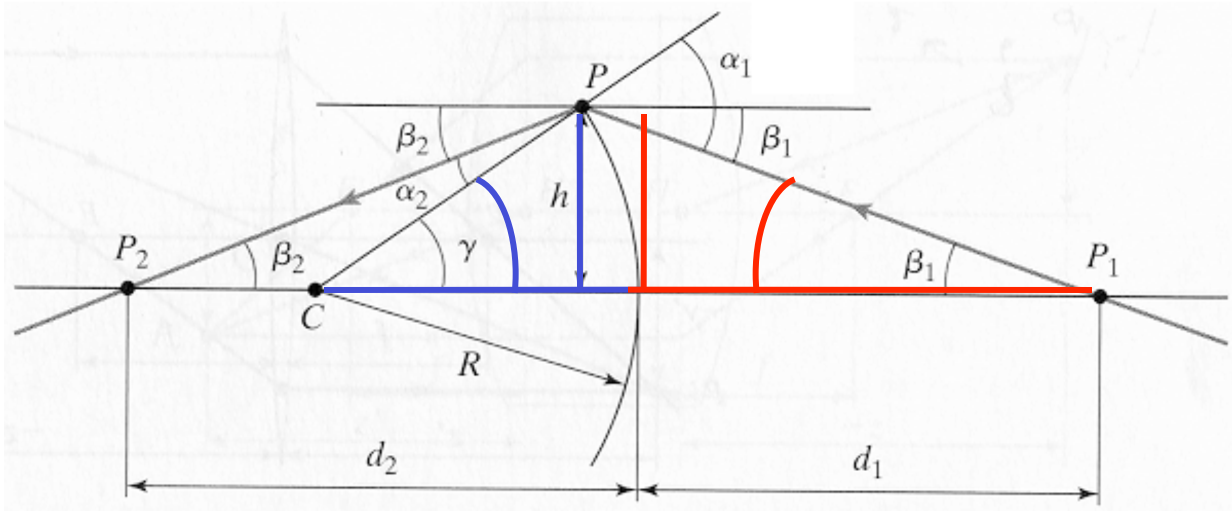
That is easiest to do under the assumptions of "first order optics": small bending angles, and a thin lens

$$\sin(\theta) \approx \theta$$



$$\theta \approx \frac{D/2}{f}$$

Paraxial refraction equation

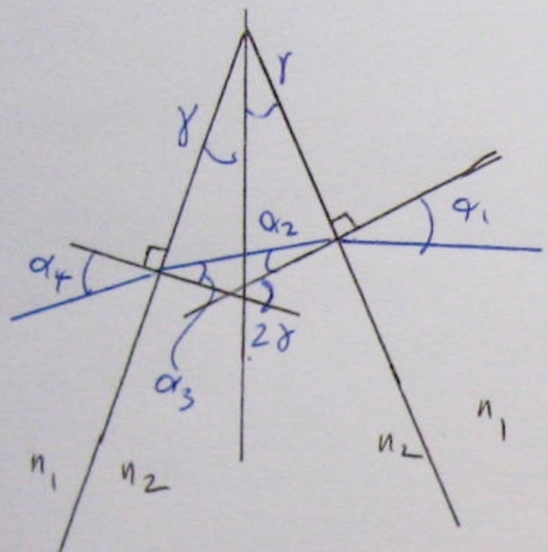


$$\alpha_1 = \boxed{\gamma} + \boxed{\beta_1} \approx h\left(\frac{1}{R} + \frac{1}{d_1}\right)$$

Relates distance from sphere and sphere radius to bending angles of light ray, for lens with one spherical surface.

$$n_1\alpha_1 \approx n_2\alpha_2$$

Deriving the lensmaker's formula



$\alpha_1 = h\left(\frac{1}{R} + \frac{1}{d_1}\right)$ small angle approx

$n_1 \alpha_1 \cong n_2 \alpha_2$ snell's law

$\alpha_2 = 2\gamma - \alpha_3$ geometry

$n_2 \alpha_3 \cong n_1 \alpha_4$ snell's law

$\alpha_4 = h_1\left(\frac{1}{R} + \frac{1}{d_2}\right)$ small angle

$\gamma = \frac{h}{R}$ small angle

$$n_1 \alpha_1 = n_2\left(\frac{2h}{R} - \frac{n_1}{n_2}\alpha_4\right) = h\left(\frac{1}{R} + \frac{1}{d_1}\right)$$

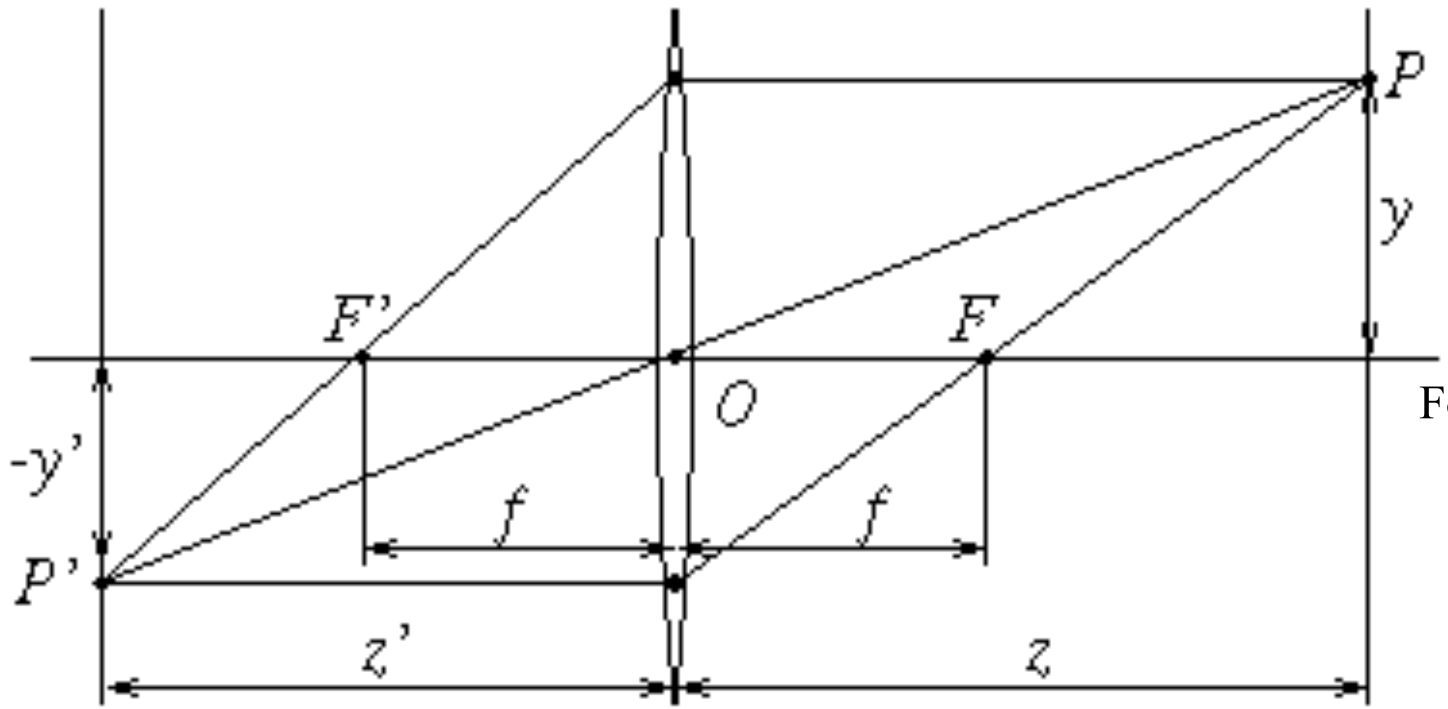
let $n_1 = 1$, $n_2 = n$

cancel h's $\quad n\left(\frac{2}{R} - \frac{1}{n}\left(\frac{1}{R} + \frac{1}{d_2}\right)\right) = \frac{1}{R} + \frac{1}{d_1}$

$$\frac{2n}{R} - \frac{1}{R} - \frac{1}{d_2} = \frac{1}{R} + \frac{1}{d_1}$$

$$\frac{2(n-1)}{R} = \frac{1}{d_1} + \frac{1}{d_2}$$ "Lens maker's formula"

The thin lens, first order optics



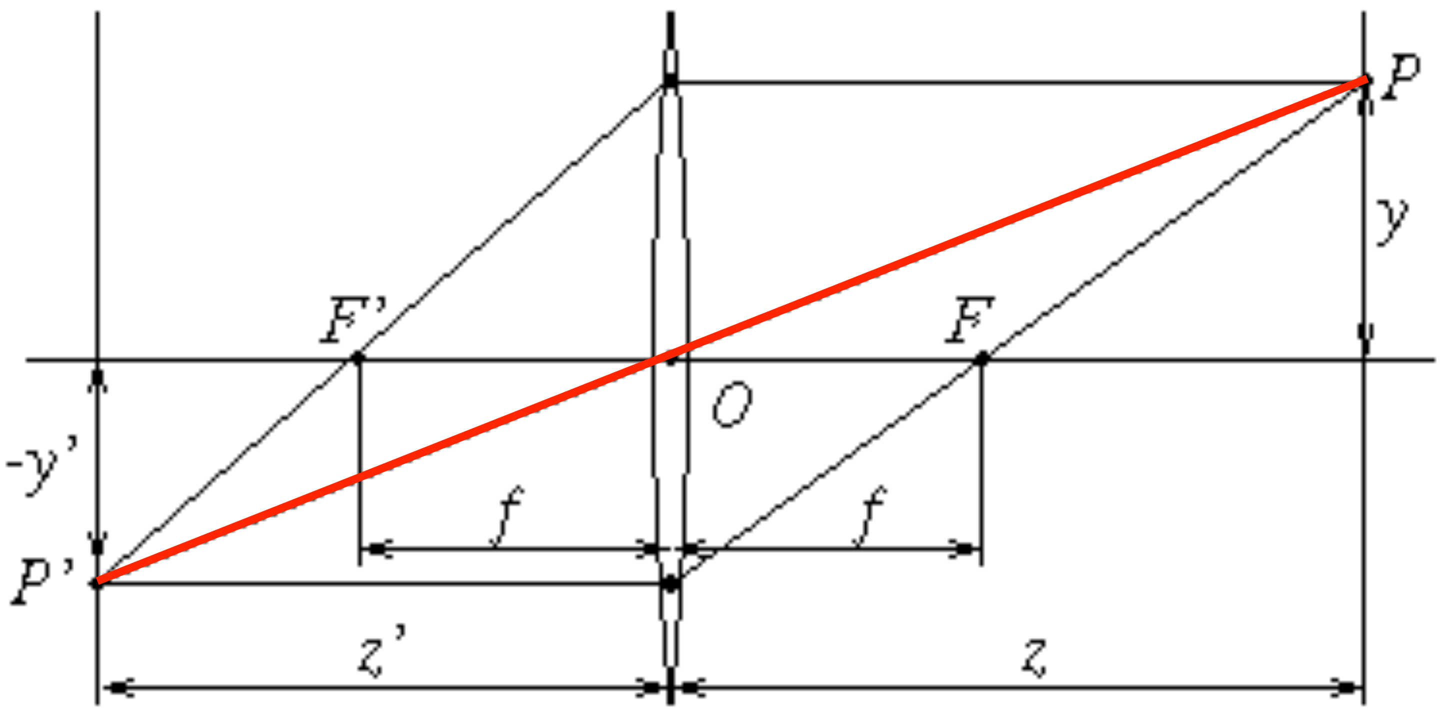
Forsyth&Ponce

The lensmaker's equation:

$$\frac{1}{z'} + \frac{1}{z} = \frac{1}{f} \qquad f = \frac{R}{2(n-1)}$$

How does the mapping from the 3-d world to the image plane compare for a lens and for a pinhole camera?

The perspective projection of a pinhole camera. But note that many more of the rays leaving from P arrive at P’



Lens demonstration

- Verify:
  - Focusing property
  - Lens maker's equation ($f = 25$ inches)
  - The relationship between distances in the world and distances in the sensor plane

more general cameras

general imagers
(from imaging chapter of course notes)

$$\vec{y} = \boldsymbol{A}\vec{x} \tag{1.9}$$

For the case of conventional cameras, where the observed intensities, $\vec{y}$ are an image of the reflected intensities in the scene, $\vec{x}$, then $\boldsymbol{A}$ is approximately an identity matrix.

For more general cameras, $\boldsymbol{A}$ may be very different from an identity matrix, and we will need to estimate $\vec{x}$ from $\vec{y}$. In the presence of noise, there may not be a solution $\vec{x}$ that exactly satisfies Eq. (1.9), so we often seek to satisfy it in a least squares sense. In most cases, $\boldsymbol{A}$ is either not invertable, or is poorly conditioned. It is often useful to introduce a regularizer, an additional term in the objective function to be minimized. If the regularization term favors small $\vec{x}$, then the objective term to minimize, E, could be

$$E = |\vec{y} - \boldsymbol{A}\vec{x}|^2 + \lambda|\vec{x}|^2 \tag{1.10}$$

general imagers
(from imaging chapter of course notes)

Setting the derivative of Eq. (1.10) with respect to the elements of the vector $\vec{x}$ equal to zero, we have

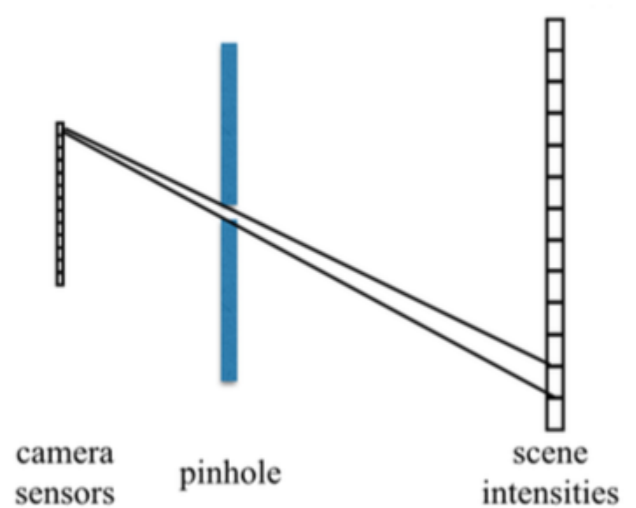
$$\begin{aligned} 0 &= \nabla_x|\vec{y} - \boldsymbol{A}\vec{x}|^2 + \nabla_x\lambda|\vec{x}|^2 && (1.11) \\ &= \boldsymbol{A}^T\boldsymbol{A}\vec{x} - \boldsymbol{A}^T\vec{y} + \lambda\vec{x} && (1.12) \\ & && (1.13) \end{aligned}$$

or

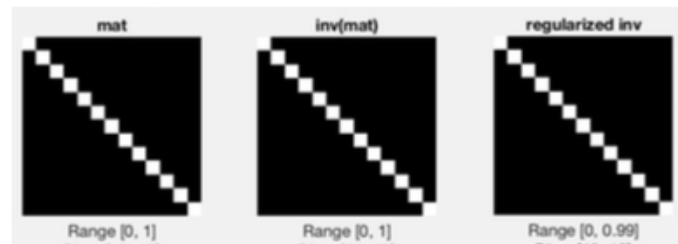
$$\vec{x} = (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\vec{y} \qquad (1.14)$$

See, e.g.: https://en.wikipedia.org/wiki/Matrix_calculus

system matrix, A, for pinhole imager



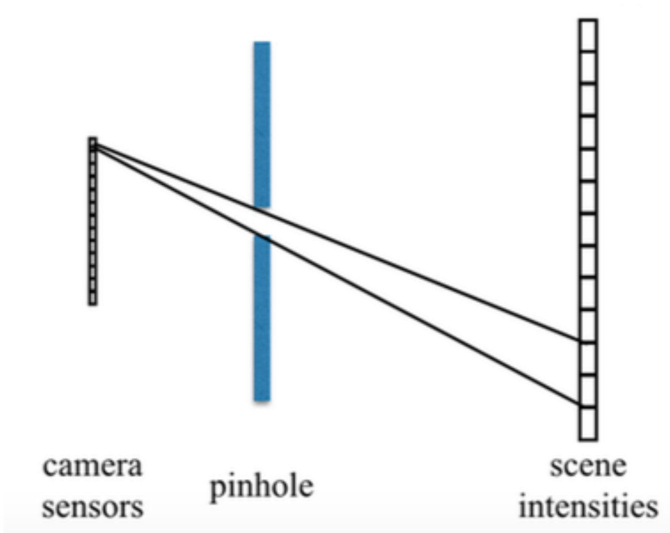
(a)



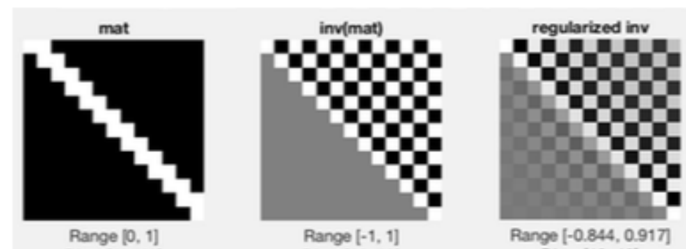
(c)

Figure 1.8
(a) Schematic drawing of a small-hole 1-d pinhole camera.(b) Visualization of imaging matrices: The imaging matrix relating scene intensities to sensor readings; the inverse of that matrix; the regularized inverse. For the small-pinhole imager, all three matrices are approximately identity matrices.

system matrix, A, for large aperture pinhole imager



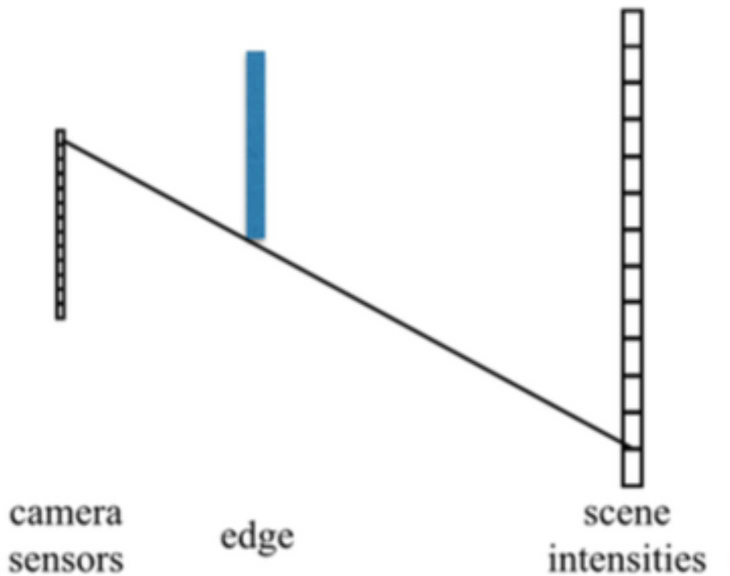
(a)



(b)

Figure 1.9
(a) Large-hole 1-d pinhole camera. (b) Visualization of imaging matrices: The imaging matrix relating scene intensities to sensor readings; the inverse of that matrix; the regularized inverse. For the small-pinhole imager, all three matrices are approximately identity matrices.

system matrix, A, for large aperture pinhole imager

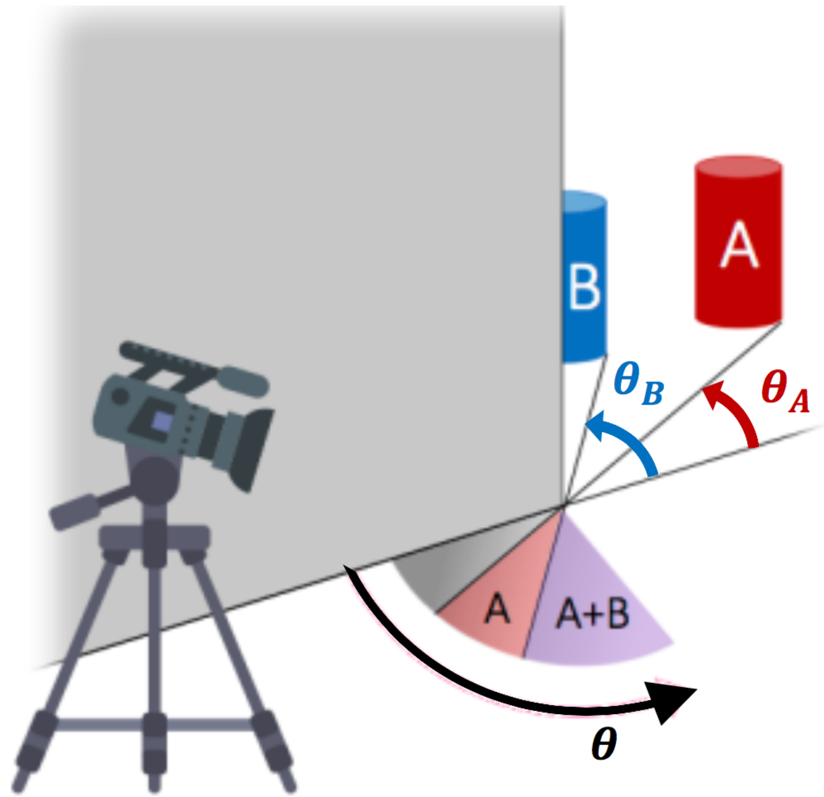


(a)

(c)

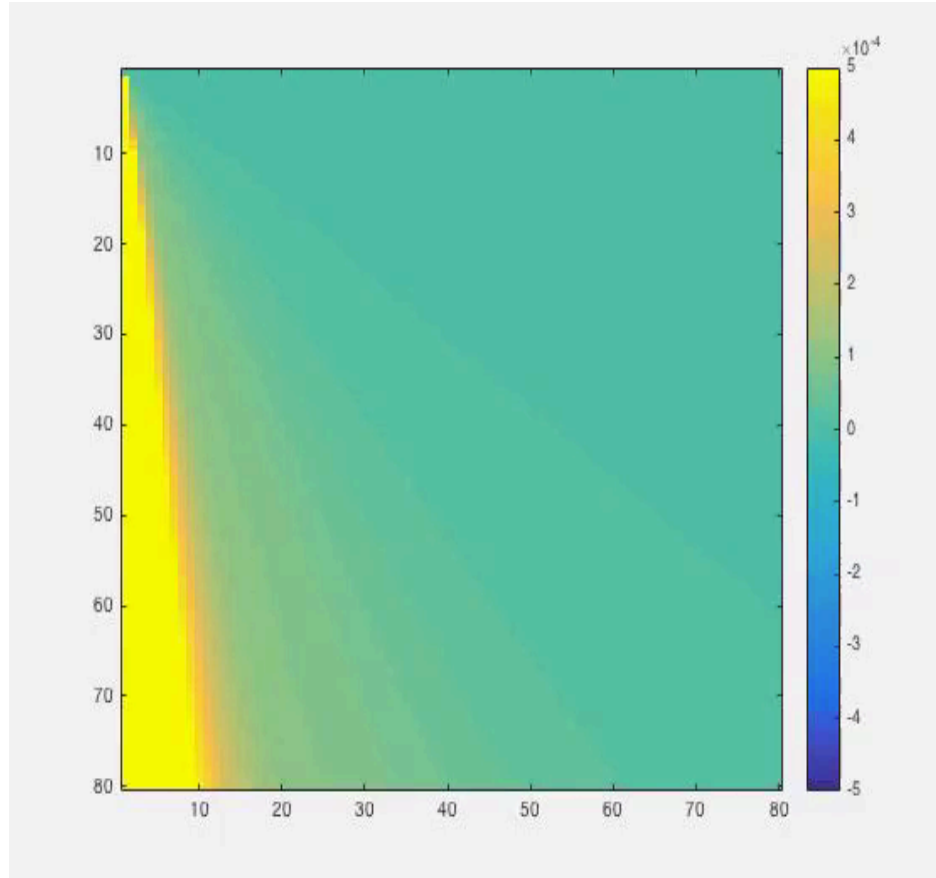
Another occlusion-based camera: edge camera

show intensity demo with cards

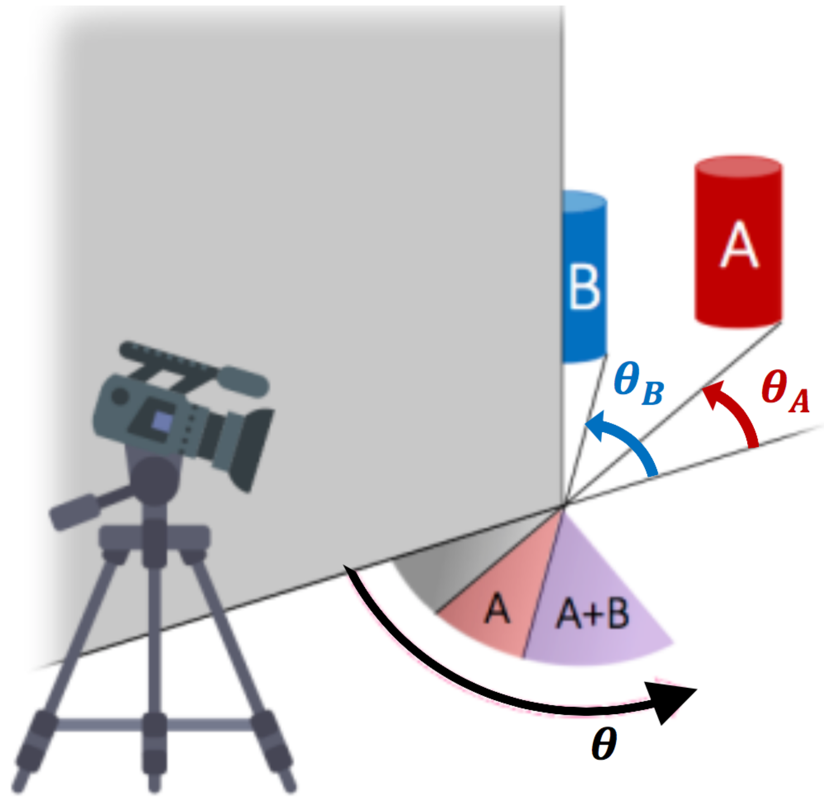

B
A
θ_B
θ_A
A
A+B
θ

Corner Camera 1-D Image Computation

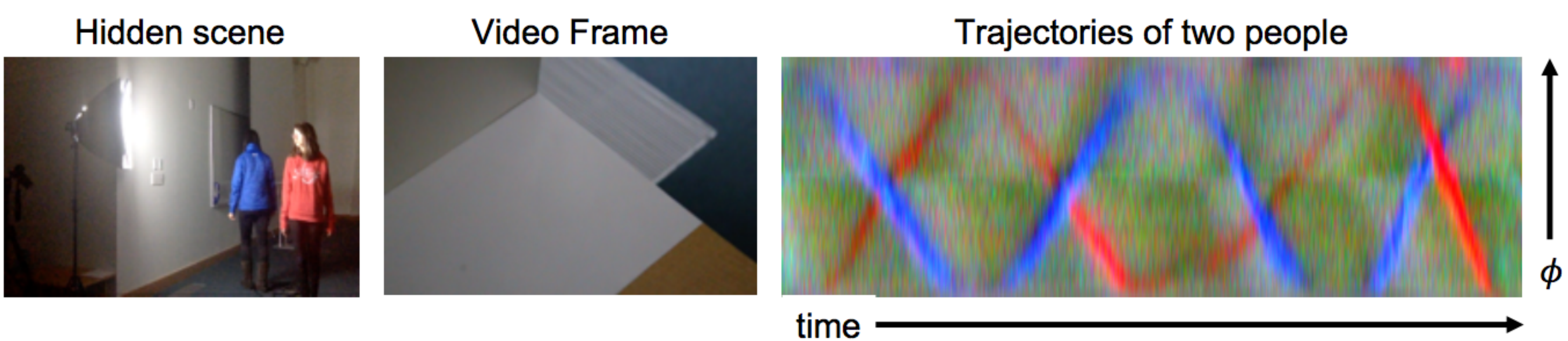
Rectified Image



Images you multiply the rectified ground plane images by to recover the input image around the corner (projected to 1d), for each different angle.



B
A
θ_B
θ_A
A
A+B
θ



Hidden scene
Video Frame
Trajectories of two people
time
ϕ

Experiment Proof of Concept

Experimental Proof of Concept

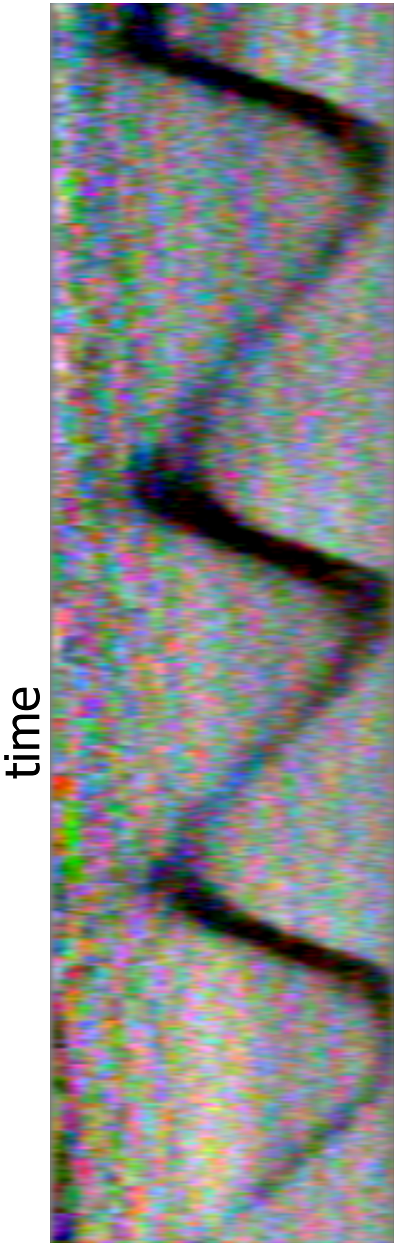
Experimental Proof of Concept

Experimental Proof of Concept

Video Corresponding to 1-D Camera

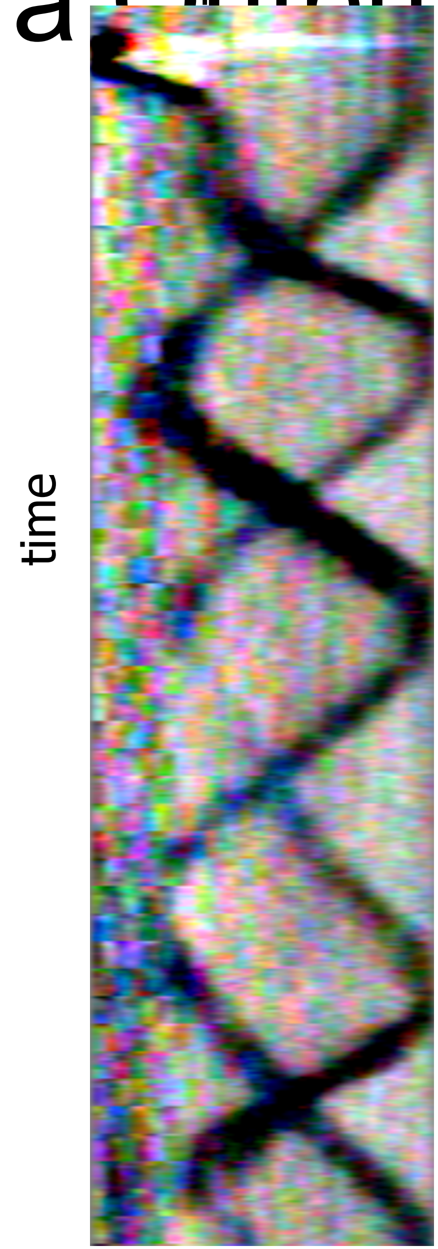
1-D Corner Camera Output

- How many people?
- Where slowed down, where moved quickly?

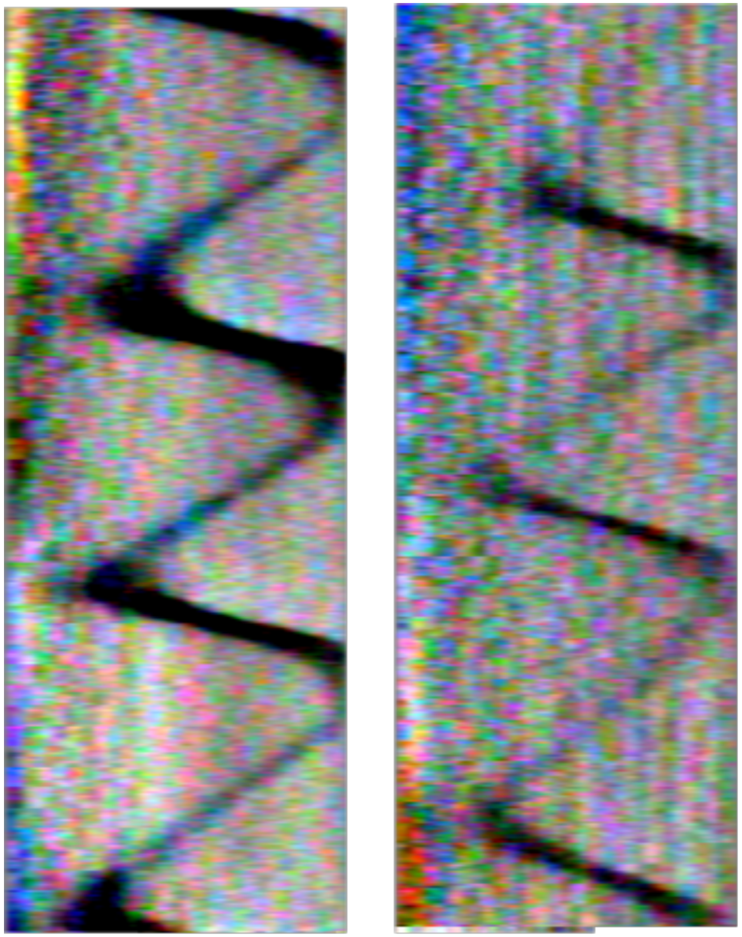


1-D Corner Camera Output

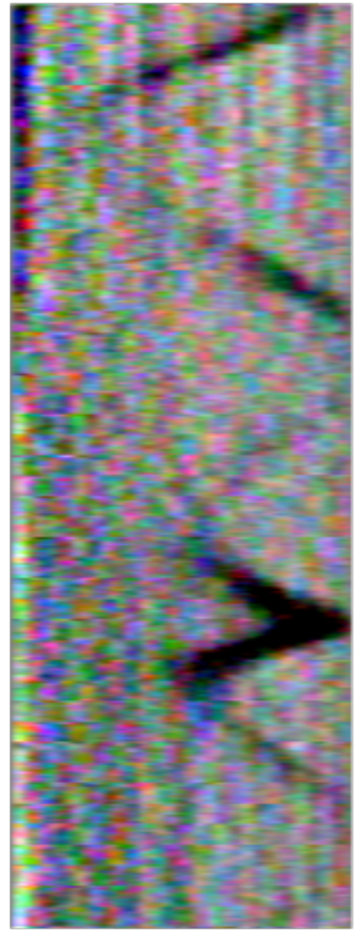
- How many people?
- How fast is each person moving?



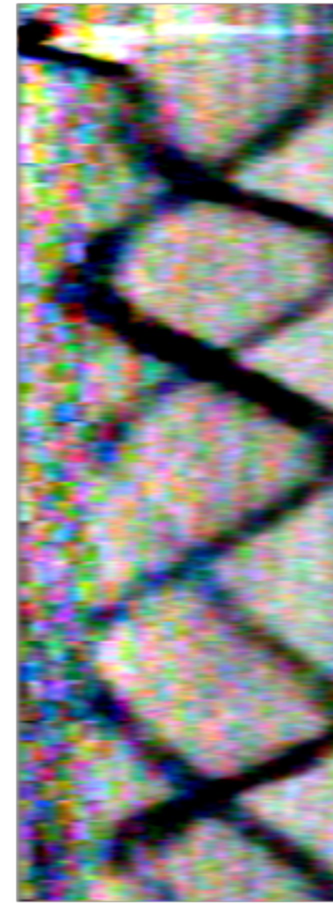
More Corner Camera Videos



1 Person Walking in Circles



1 Person
Randomly Walking



2 People Walking
in Circles

Additional Results

Paper ID: 1983

Summary

- Want to make images
- Pinhole camera models the geometry of perspective projection
- Lenses make it work in practice
- Models for lenses
  - Thin lens, spherical surfaces, first order optics
- Cameras as linear systems to invert