MIT COMPUTER VISION

MIT CSAIL

6.869: Advances in Computer Vision

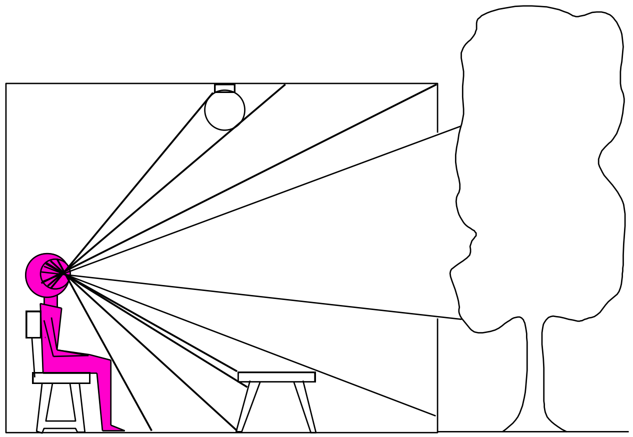
Antonio Torralba, 2012

Lecture 9

Image formation

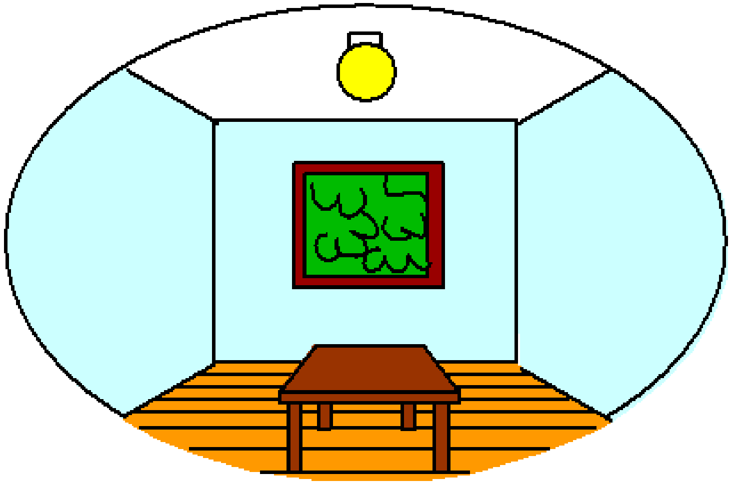
Image formation

3D world



Point of observation

2D image

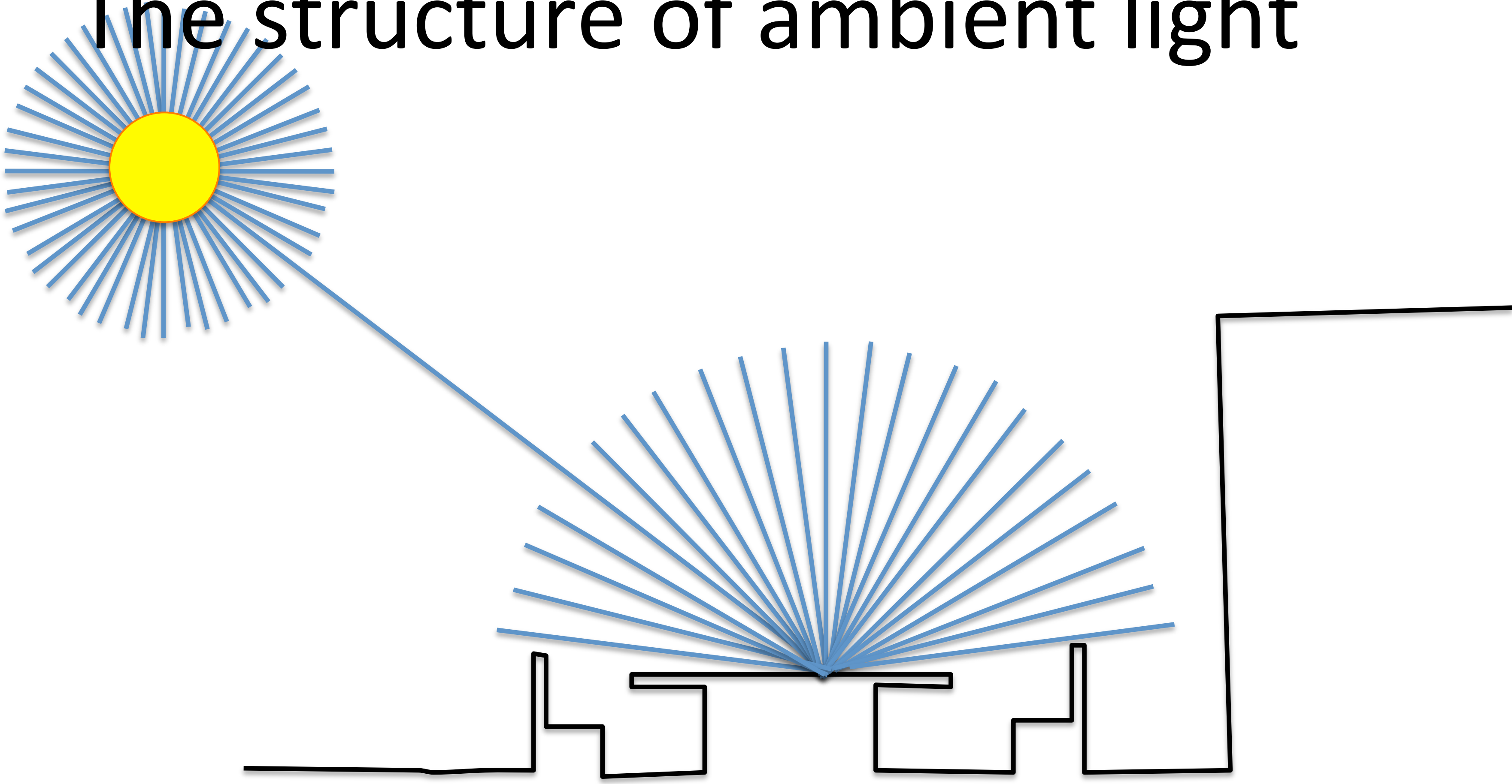


Cameras, lenses, and calibration

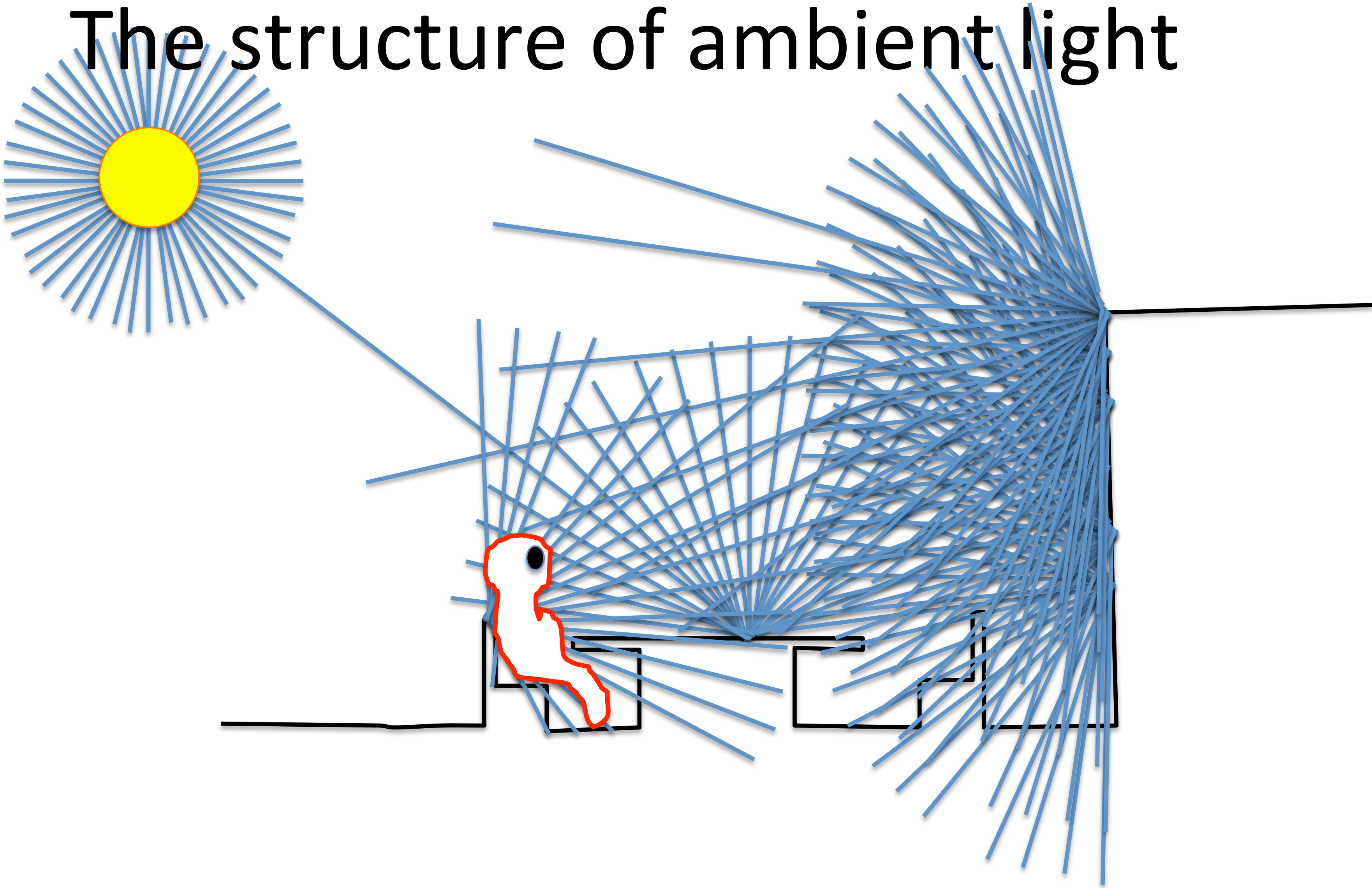
- Camera models
- Projection equations

Images are projections of the 3-D world onto a 2-D plane…

The structure of ambient light

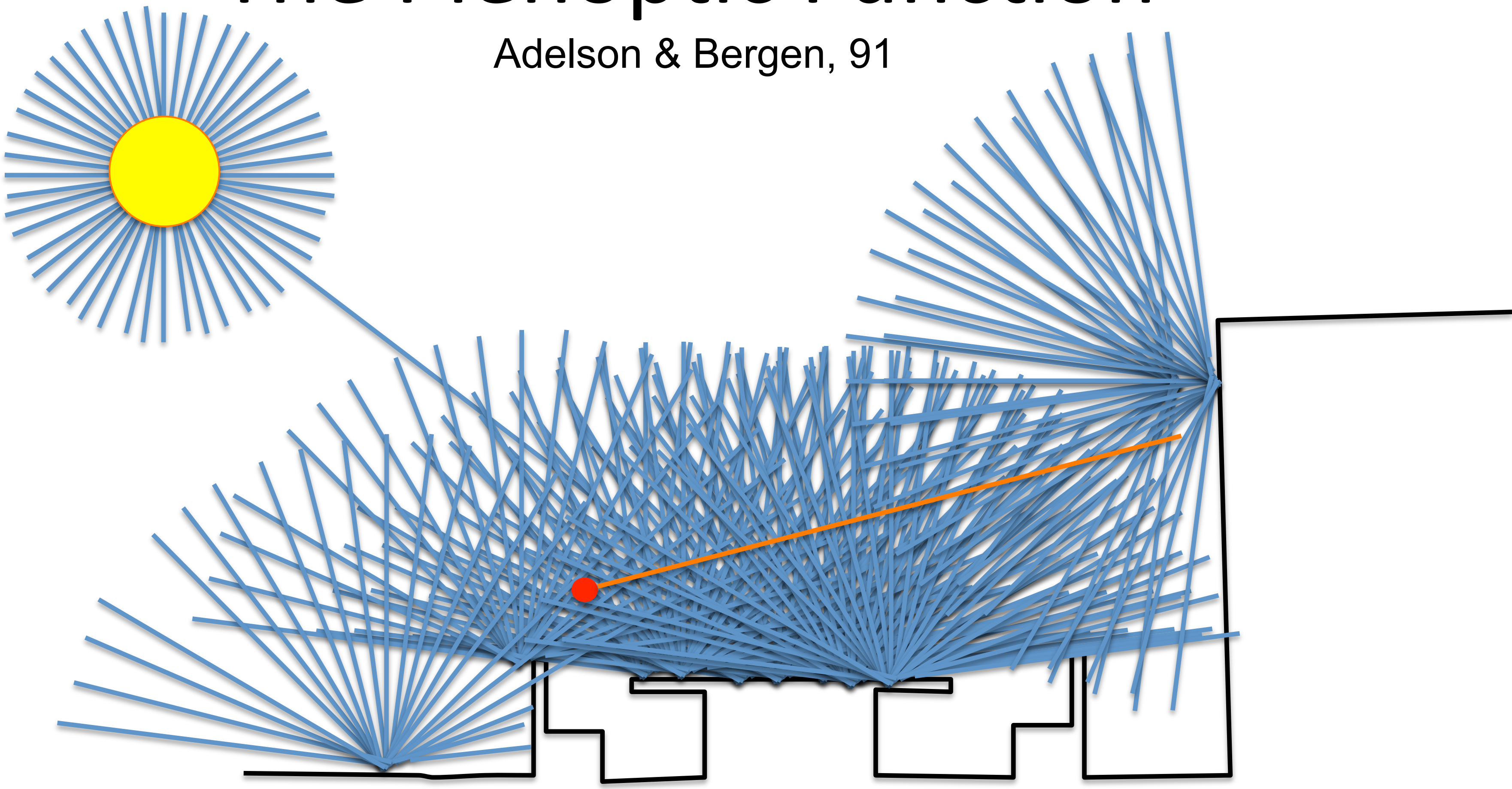


The structure of ambient light



The Plenoptic Function

Adelson & Bergen, 91

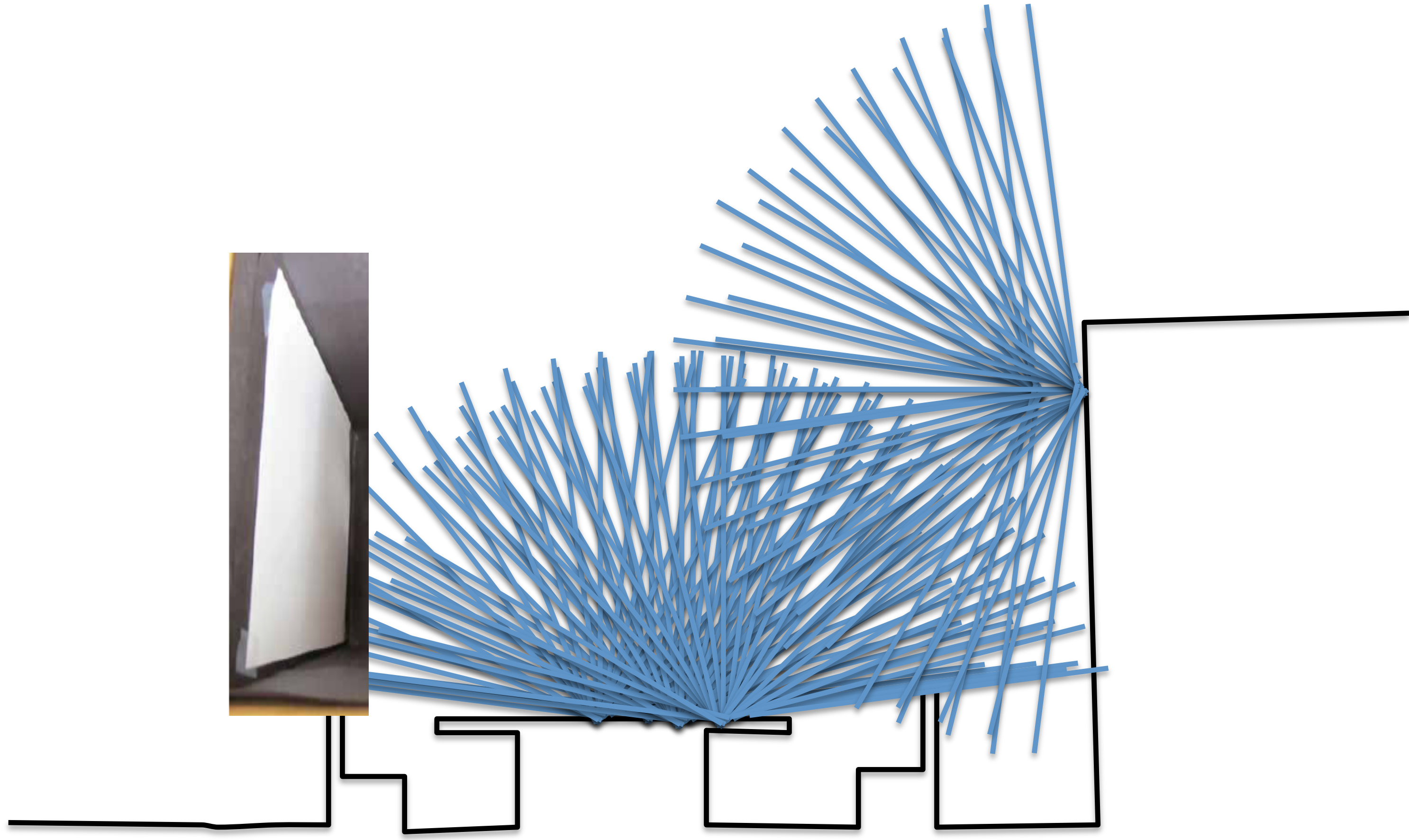


The intensity P can be parameterized as:

$$P\,(\theta, \phi, \lambda, t, X, Y, Z)$$

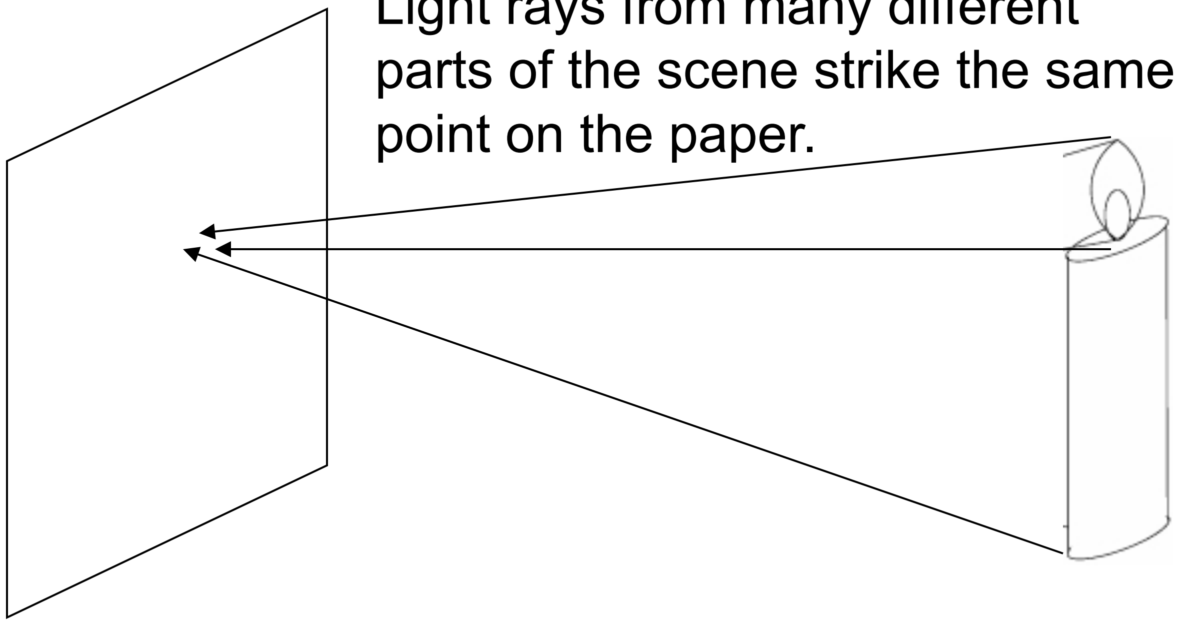
"The complete set of all convergence points constitutes the permanent possibilities of vision." Gibson

Measuring the Plenoptic function



Why is there no picture appearing on the paper?

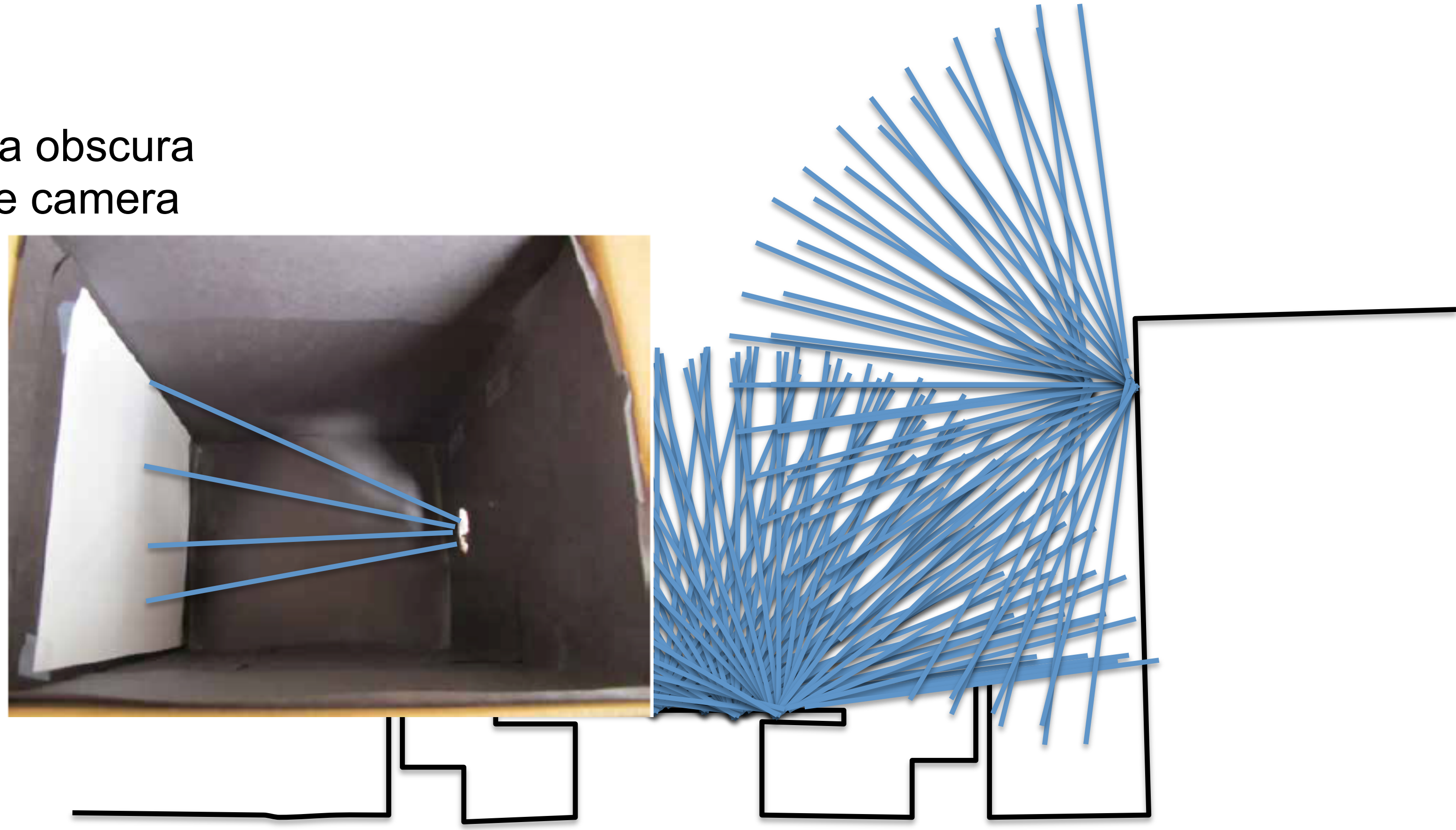
Light rays from many different parts of the scene strike the same point on the paper.



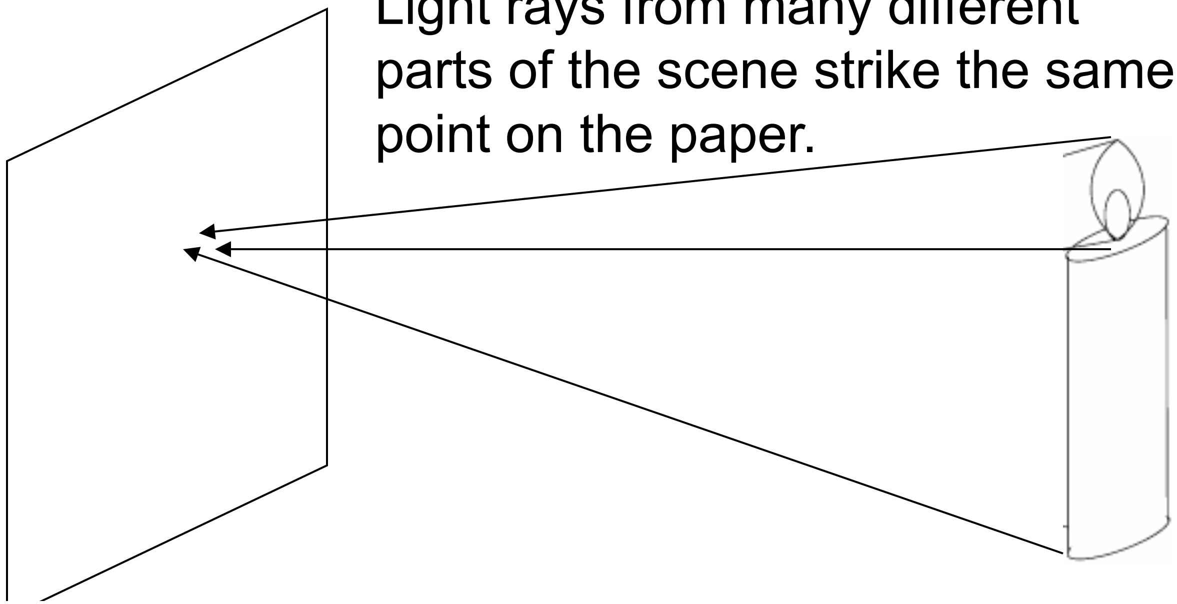
Forsyth & Ponce

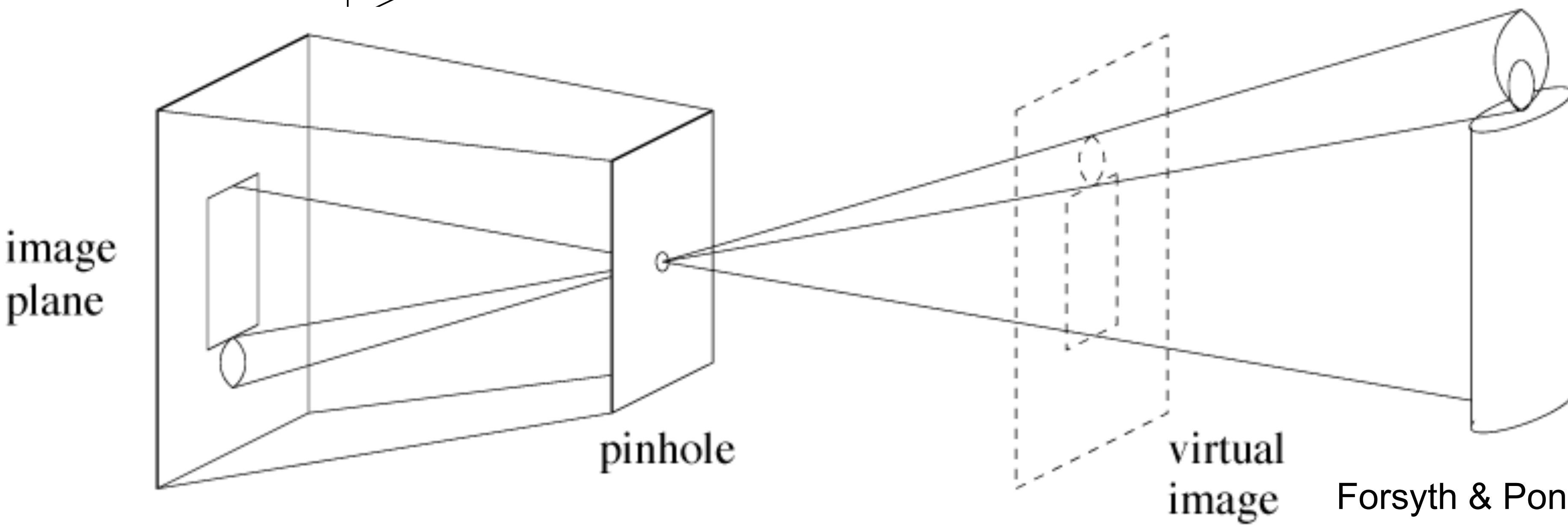
Measuring the Plenoptic function

The camera obscura
The pinhole camera



Light rays from many different parts of the scene strike the same point on the paper.





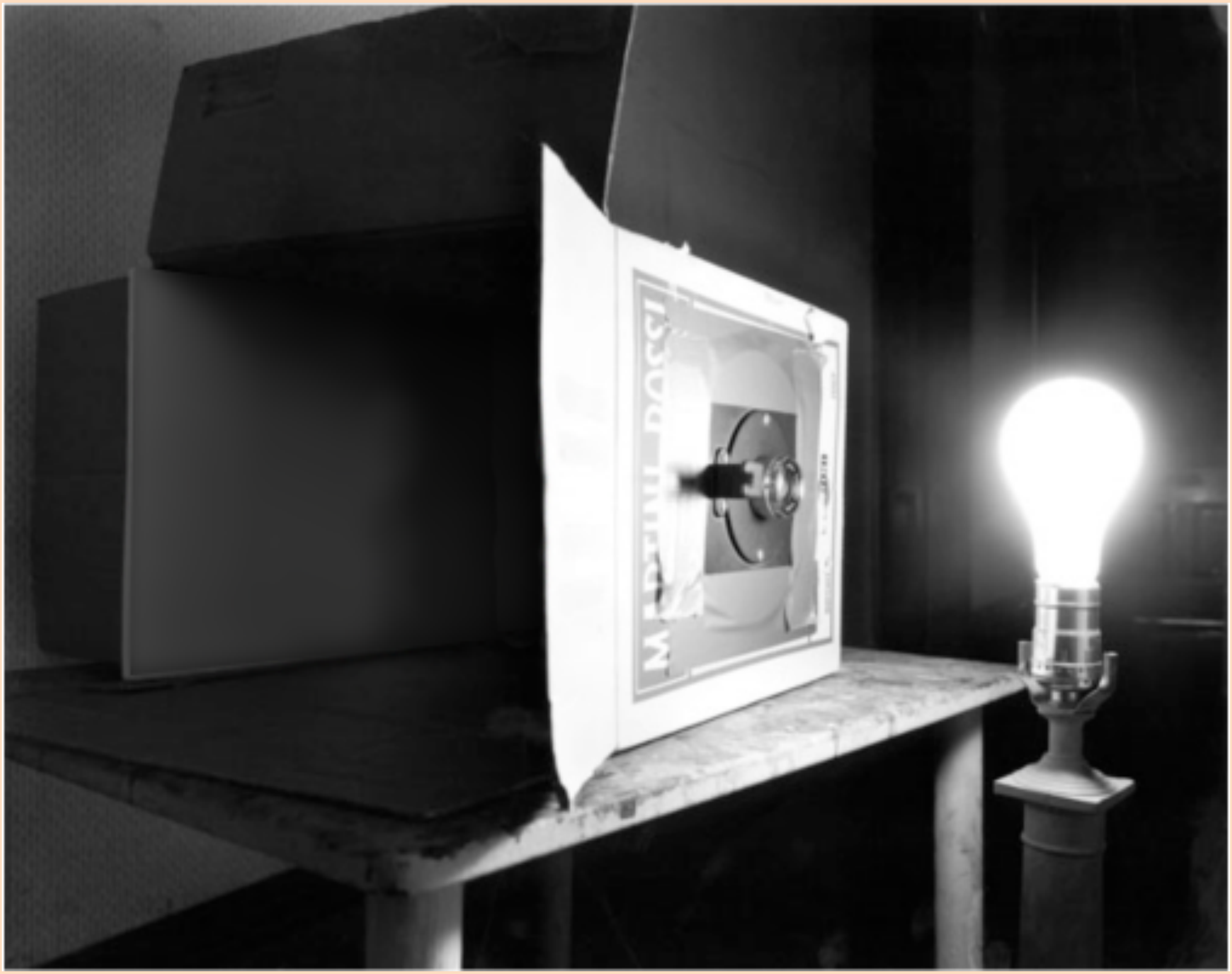
Forsyth & Ponce

The pinhole camera only allows rays from one point in the scene to strike each point of the paper.

Pinhole camera

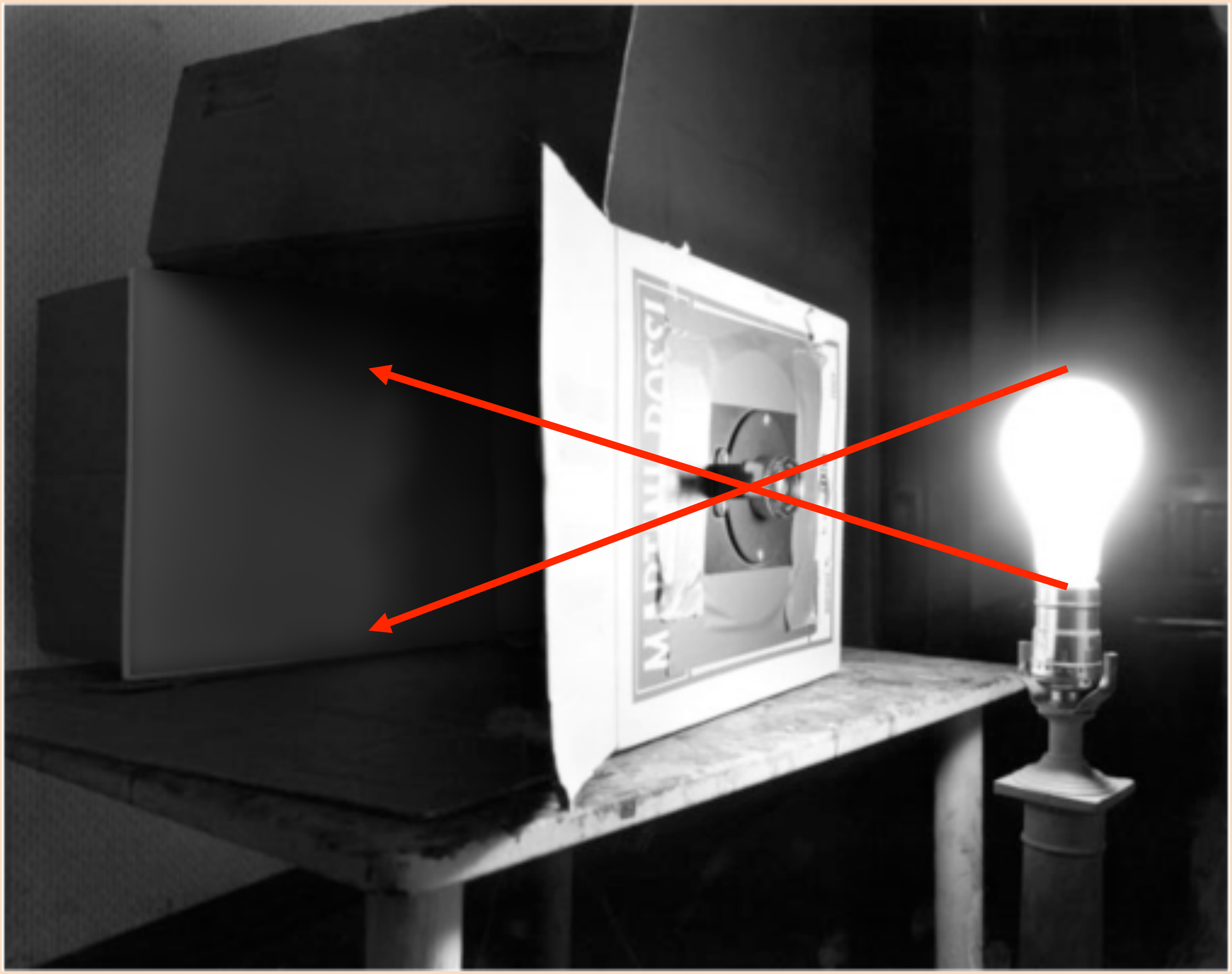
Photograph by Abelardo Morell, 1991

Pinhole camera



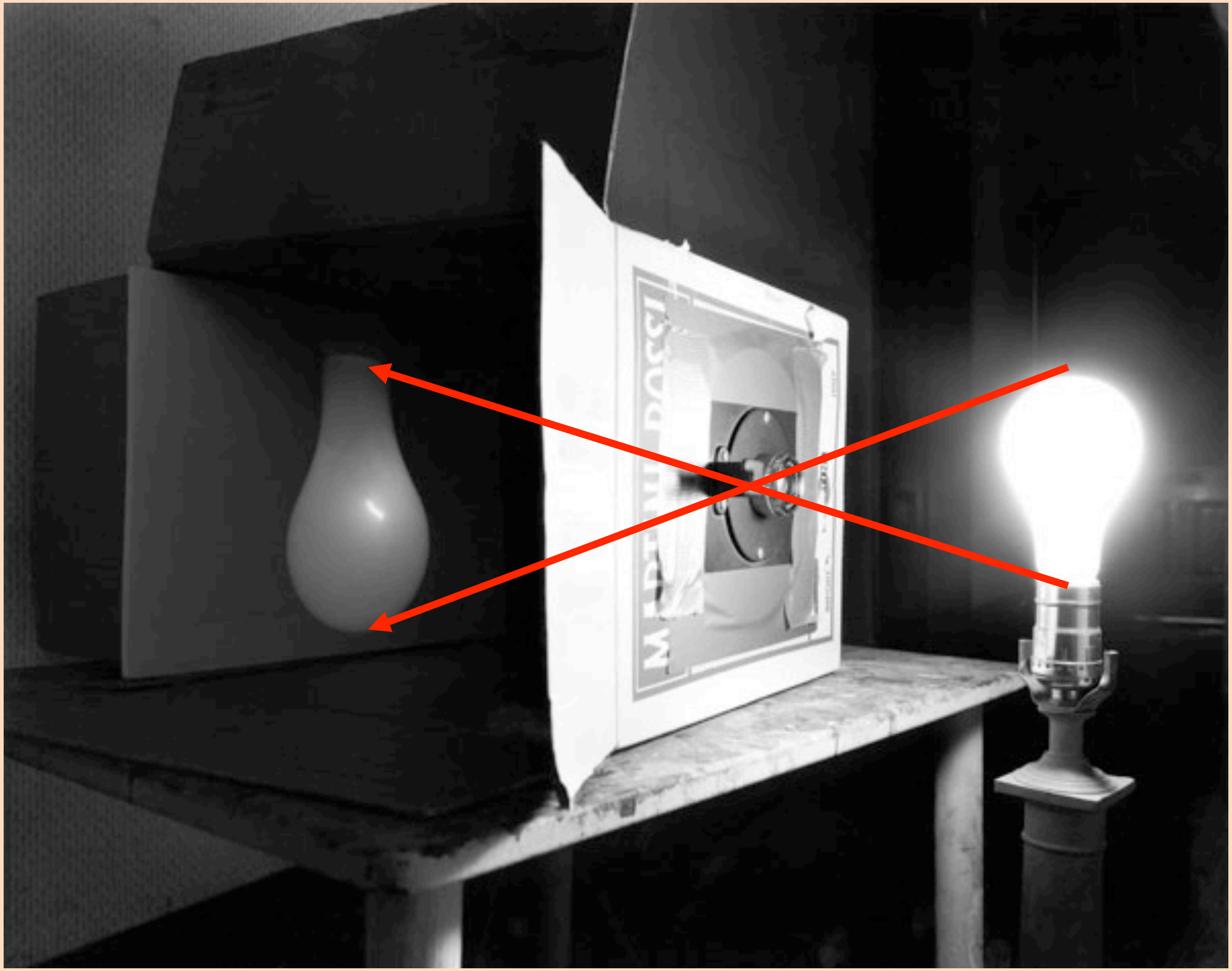
Photograph by Abelardo Morell, 1991

Pinhole camera



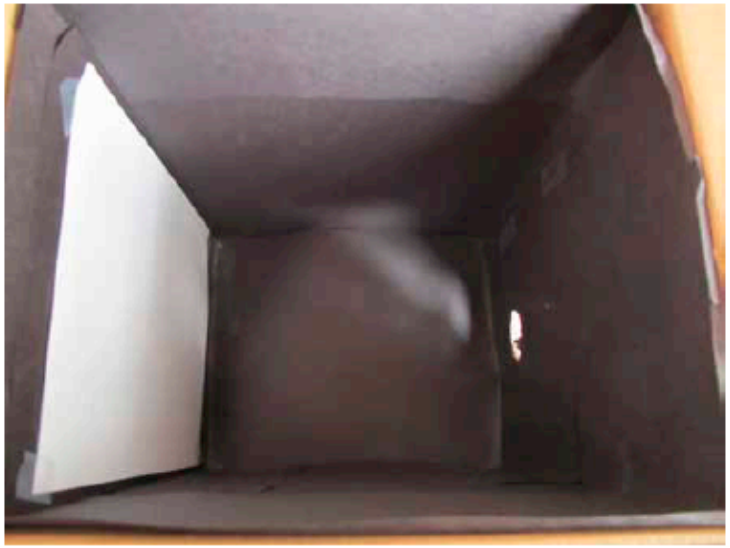
Photograph by Abelardo Morell, 1991

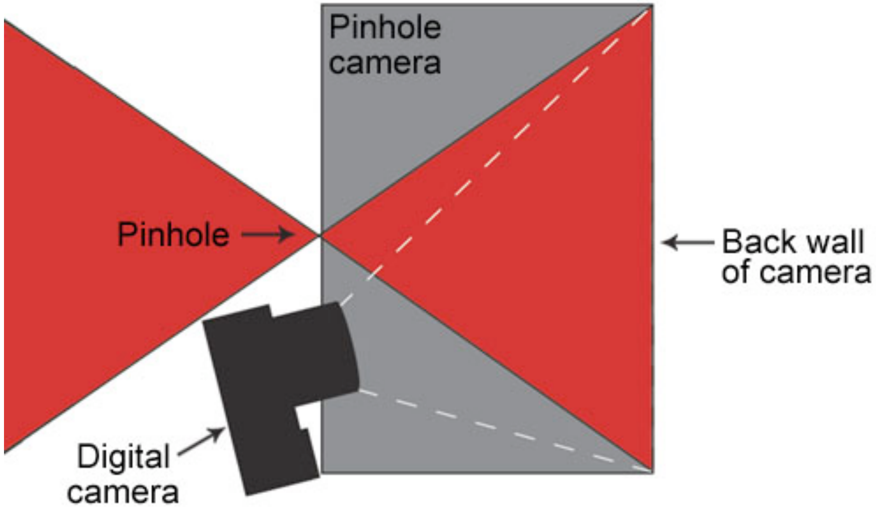
Pinhole camera



Photograph by Abelardo Morell, 1991

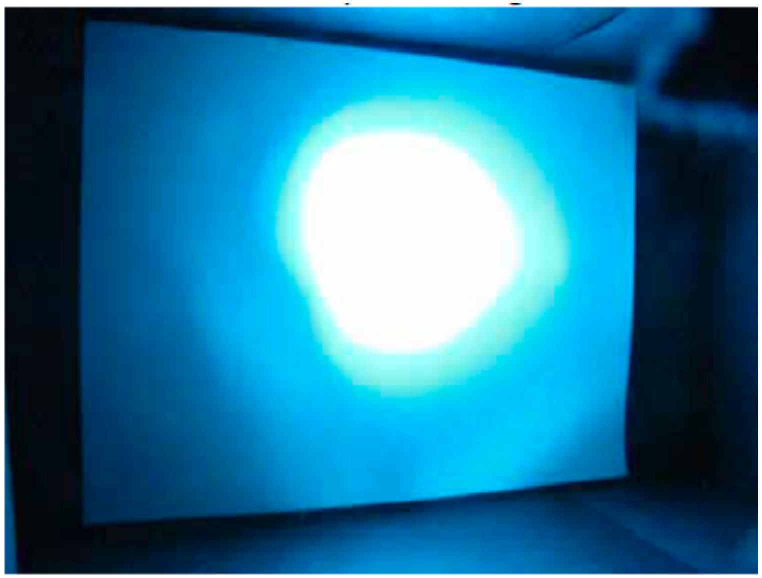
Problem Set 1



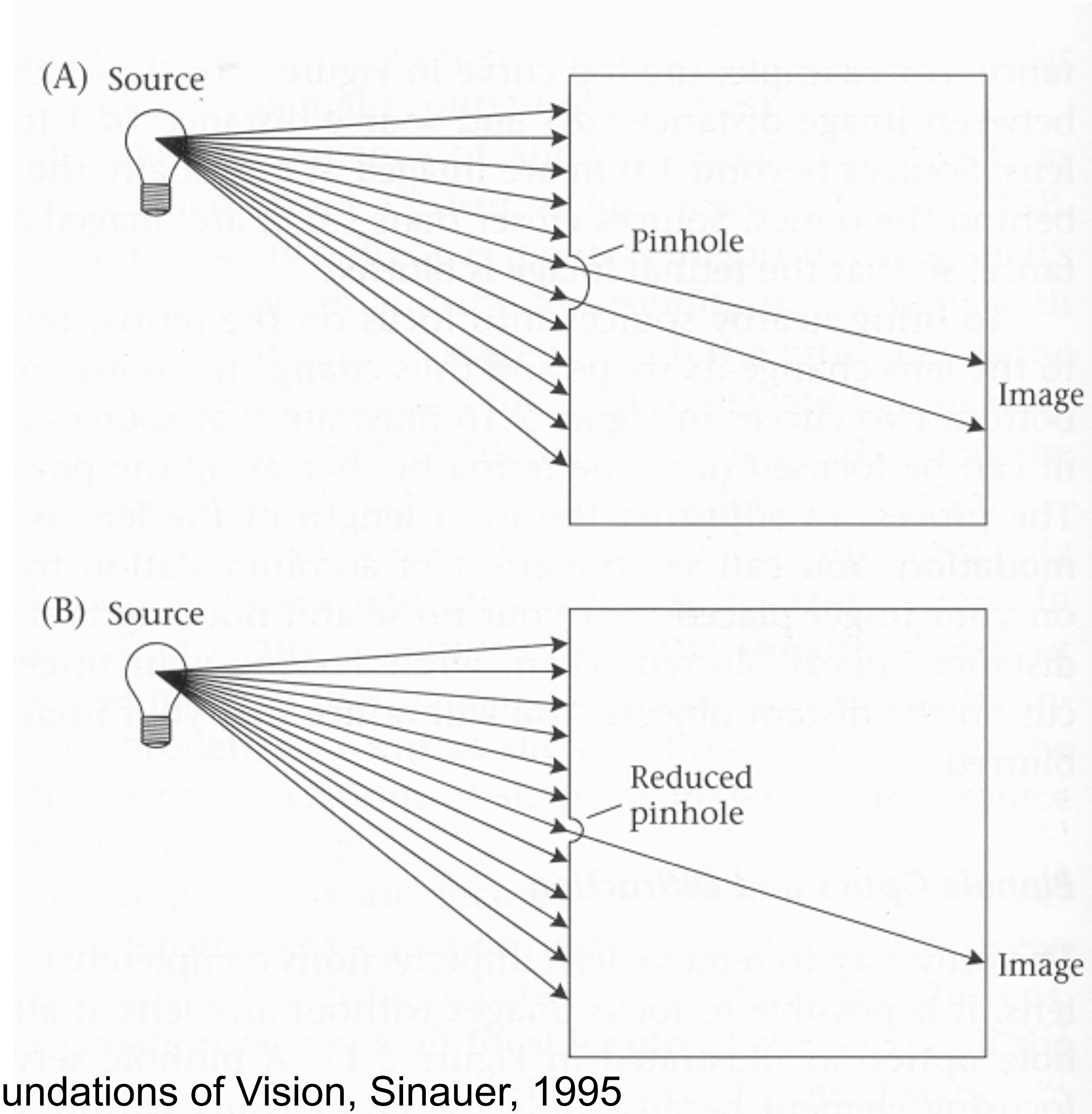


http://www.foundphotography.com/PhotoThoughts/archives/2005/04/pinhole_camera_2.html

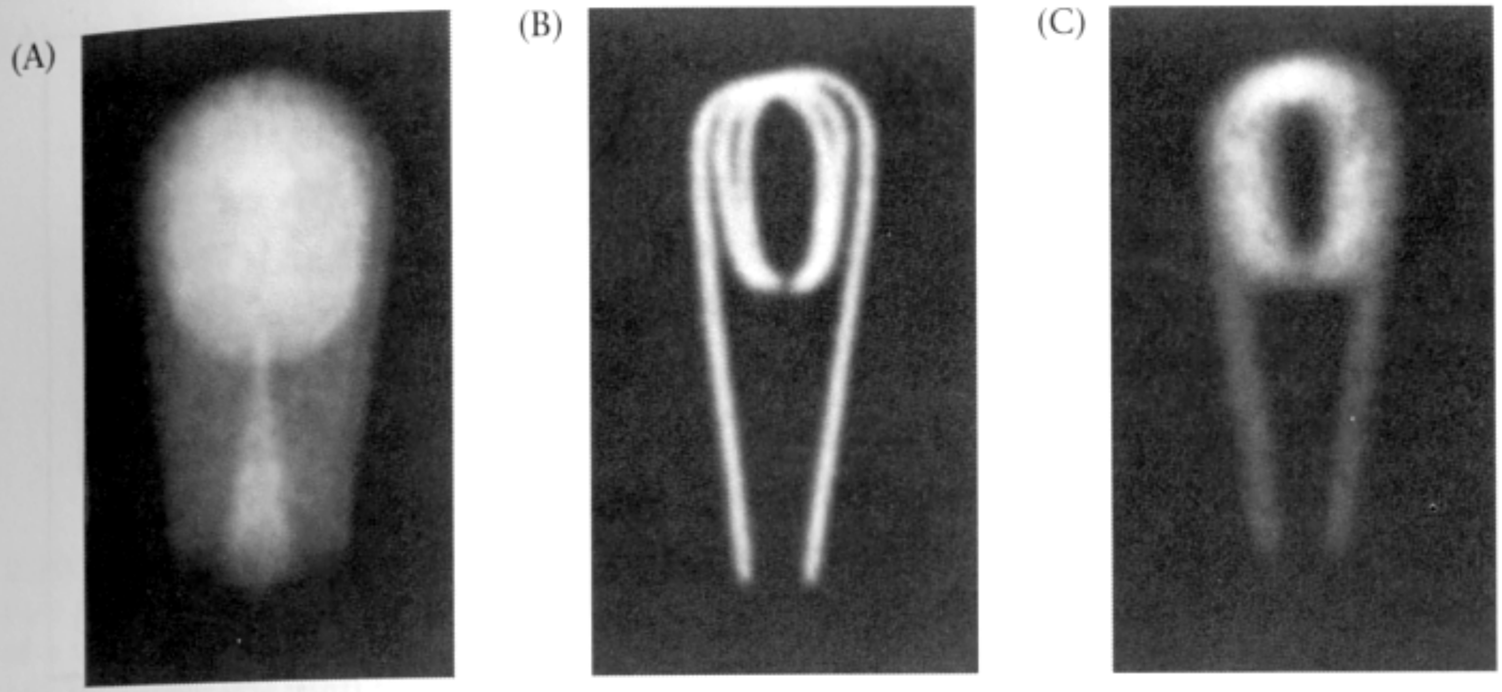
Problem Set 1



Effect of pinhole size



Wandell, Foundations of Vision, Sinauer, 1995



2.18 DIFFRACTION LIMITS THE QUALITY OF PINHOLE OPTICS. These three images of a bulb filament were made using pinholes with decreasing size. (A) When the pinhole is relatively large, the image rays are not properly converged, and the image is blurred. (B) Reducing the size of the pinhole improves the focus. (C) Reducing the size of the pinhole further worsens the focus, due to diffraction. From Ruechardt, 1958.

Wandell, Foundations of Vision, Sinauer, 1995

Animal Eyes

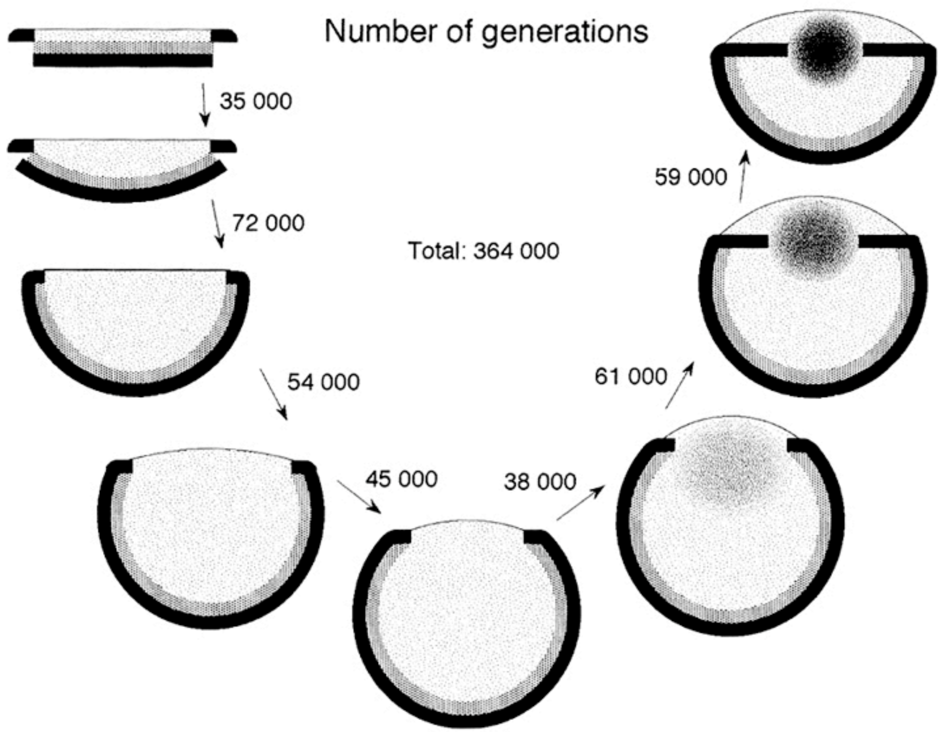
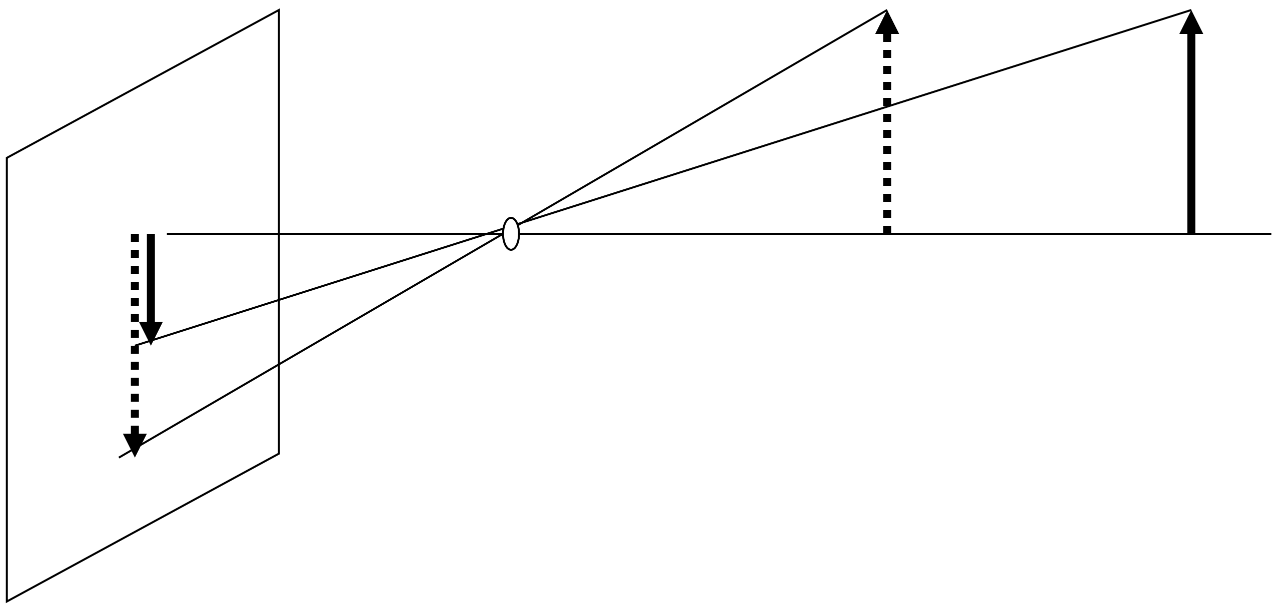


Fig. 1.6 A patch of light sensitive epithelium can be gradually turned into a perfectly focussed camera-type eye if there is a continuous selection for improved spatial vision. A theoretical model based on conservative assumptions about selection pressure and the amount of variation in natural populations suggest that the whole sequence can be accomplished amazingly fast, in less than 400 000 generations. The number of generations is also given between each of the consecutive intermediates that are drawn in the figure. The starting point is a flat piece of epithelium with an outer protective layer, an intermediate layer of receptor cells, and a bottom layer of pigment cells. The first half of the sequence is the formation of a pigment cup eye. When this principle cannot be improved any further, a lens gradually evolves. Modified from Nilsson and Pelger (1994).

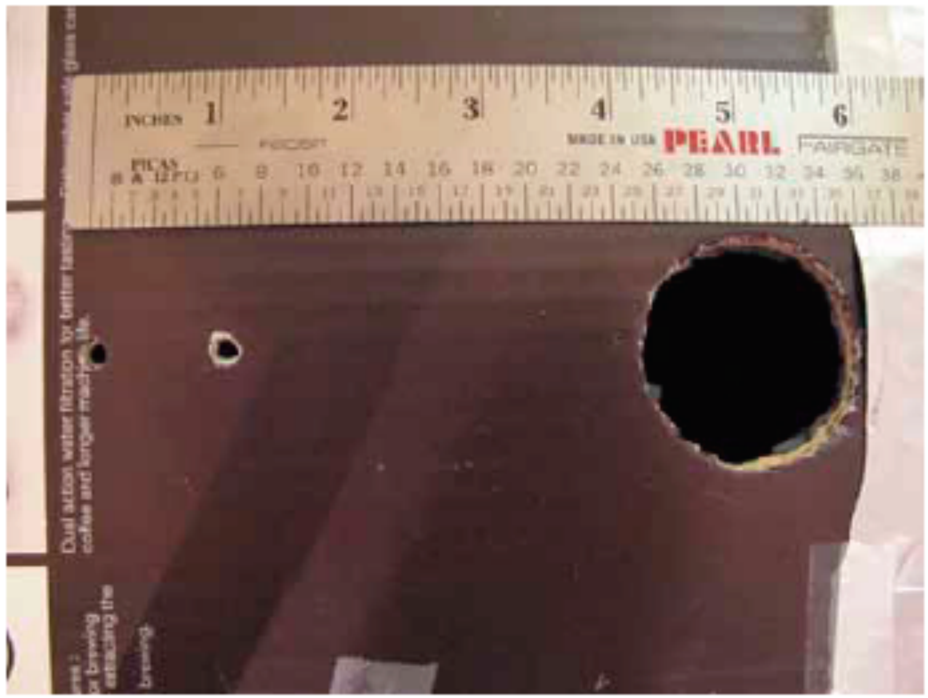
Animal Eyes. Land & Nilsson. Oxford Univ. Press

Measuring distance

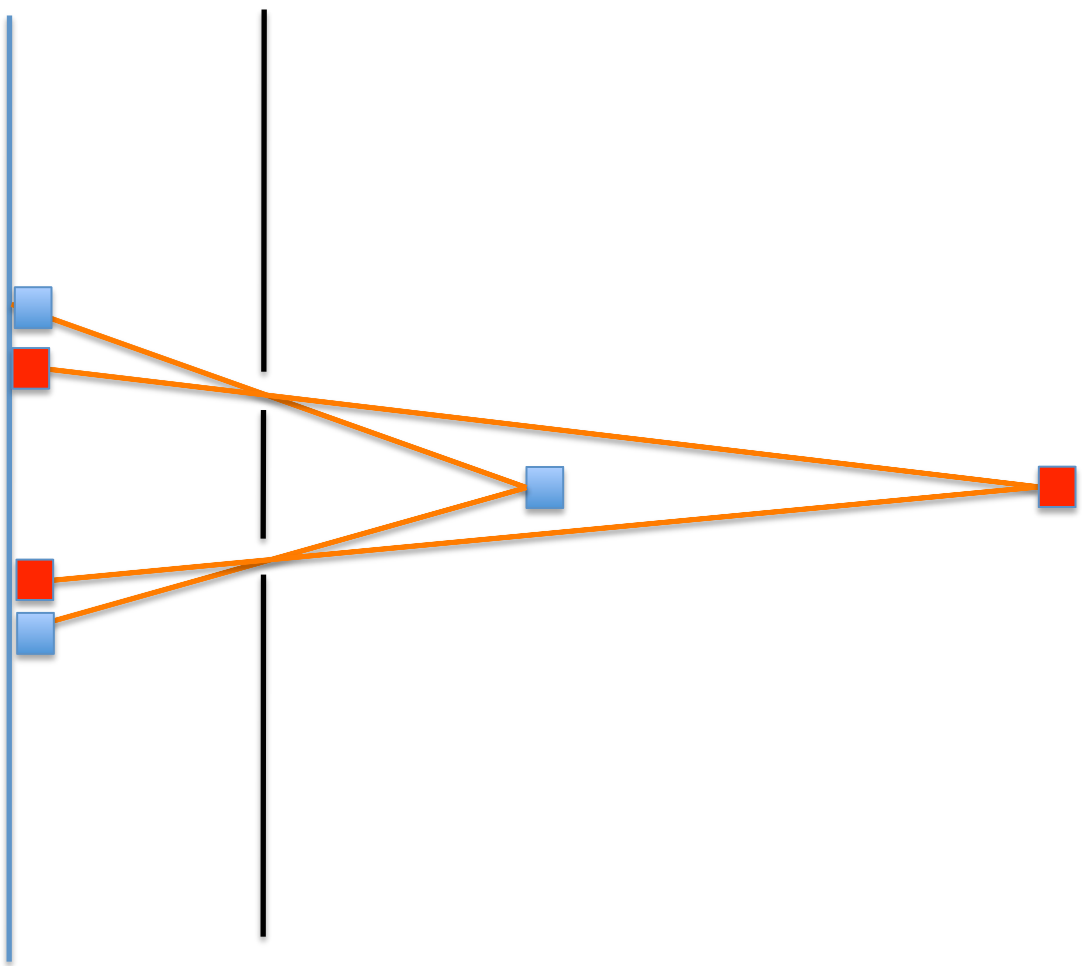


• Object size decreases with distance to the pinhole
• There, given a single projection, if we know the size of the object we can know how far it is.
• But for objects of unknown size, the 3D information seems to be lost.

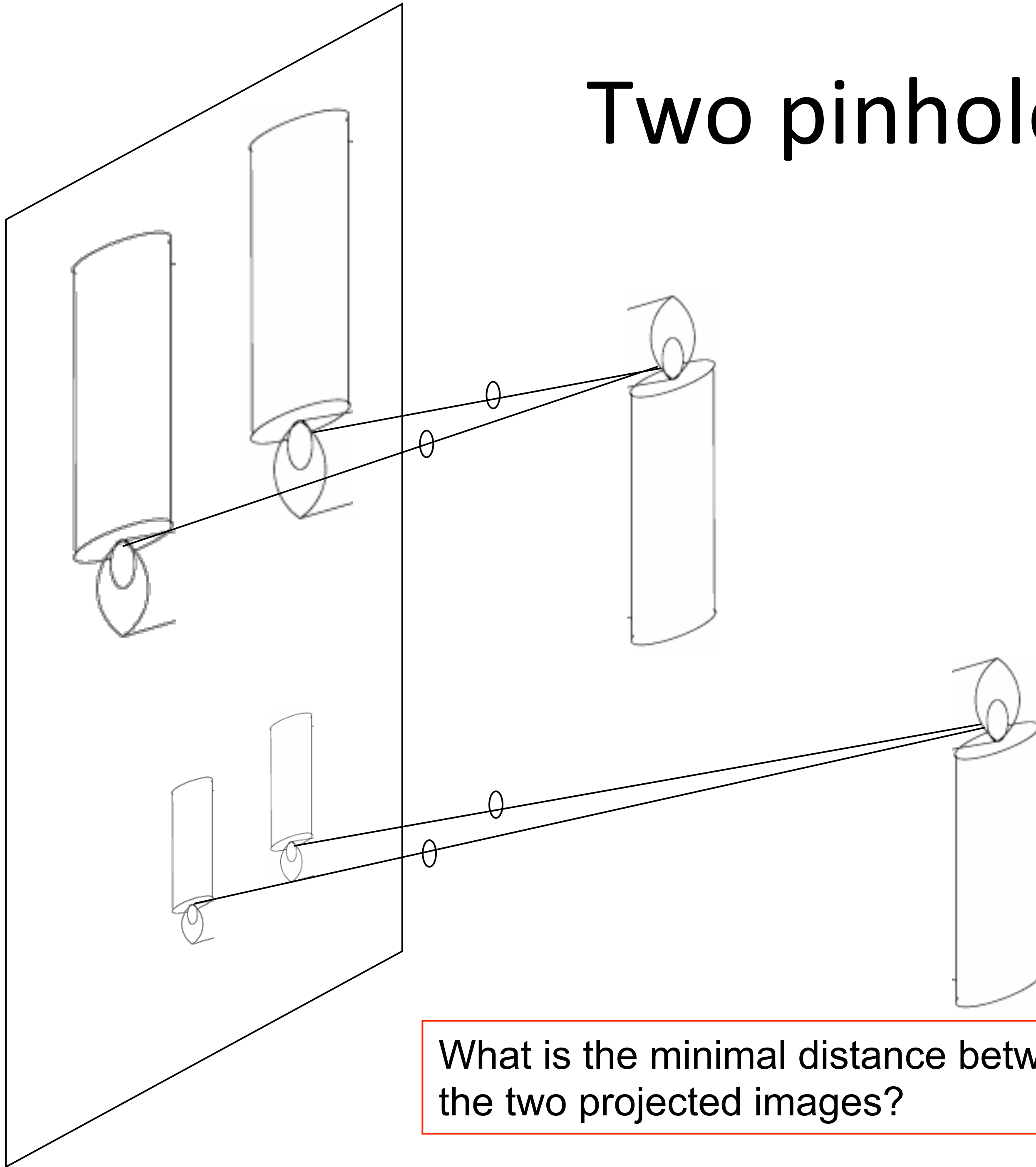
Playing with pinholes



Two pinholes

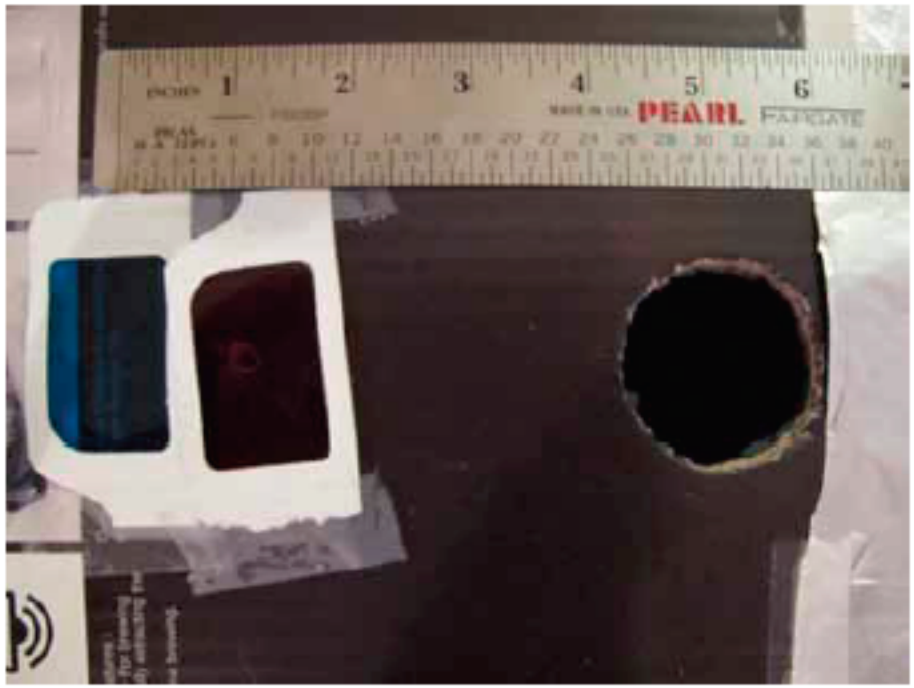


Two pinholes

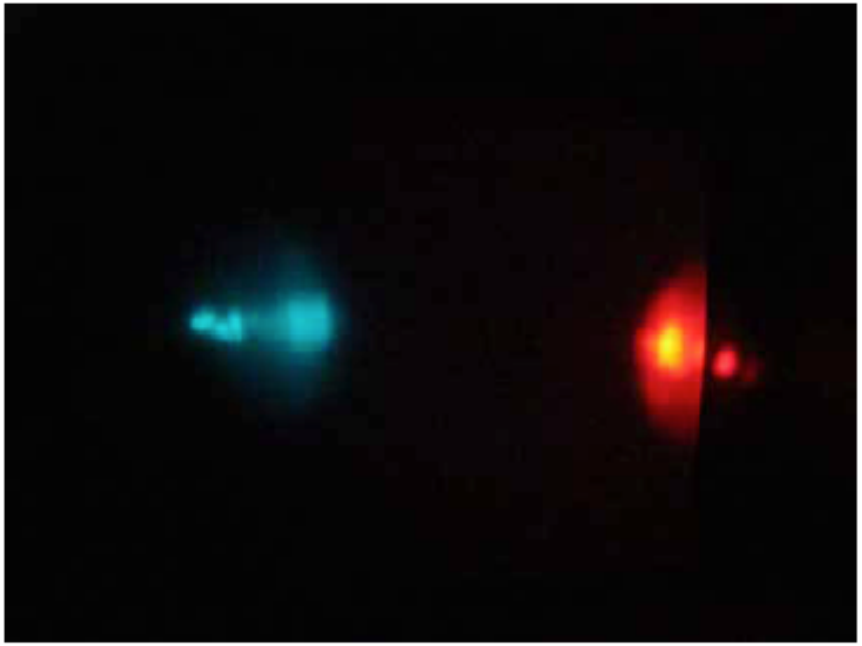


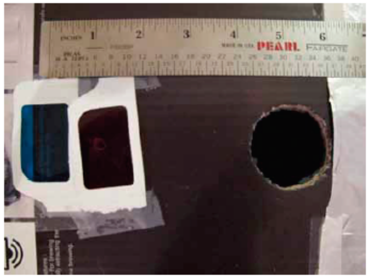
What is the minimal distance between the two projected images?

Anaglyph pinhole camera

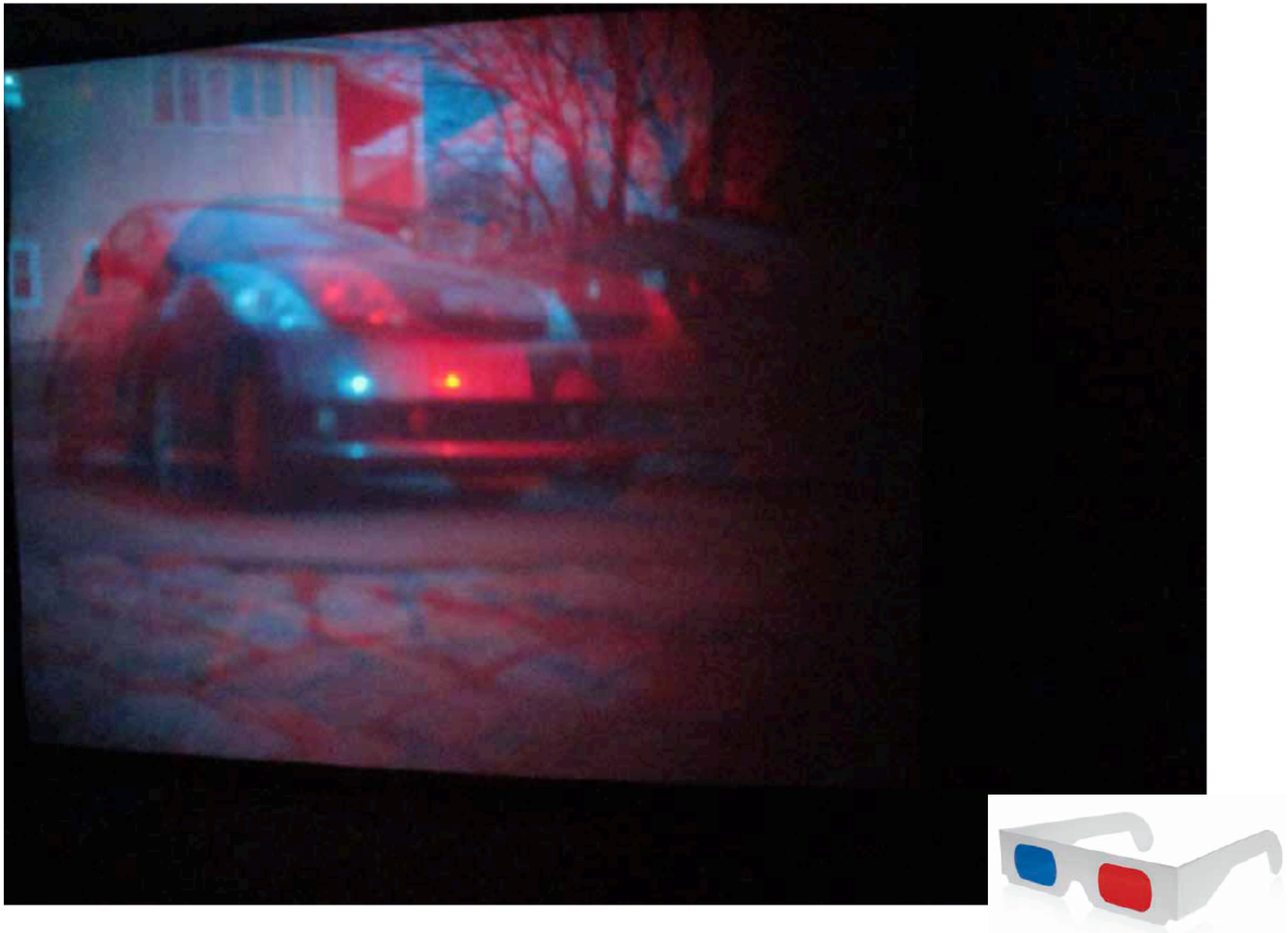


Anaglyph pinhole camera

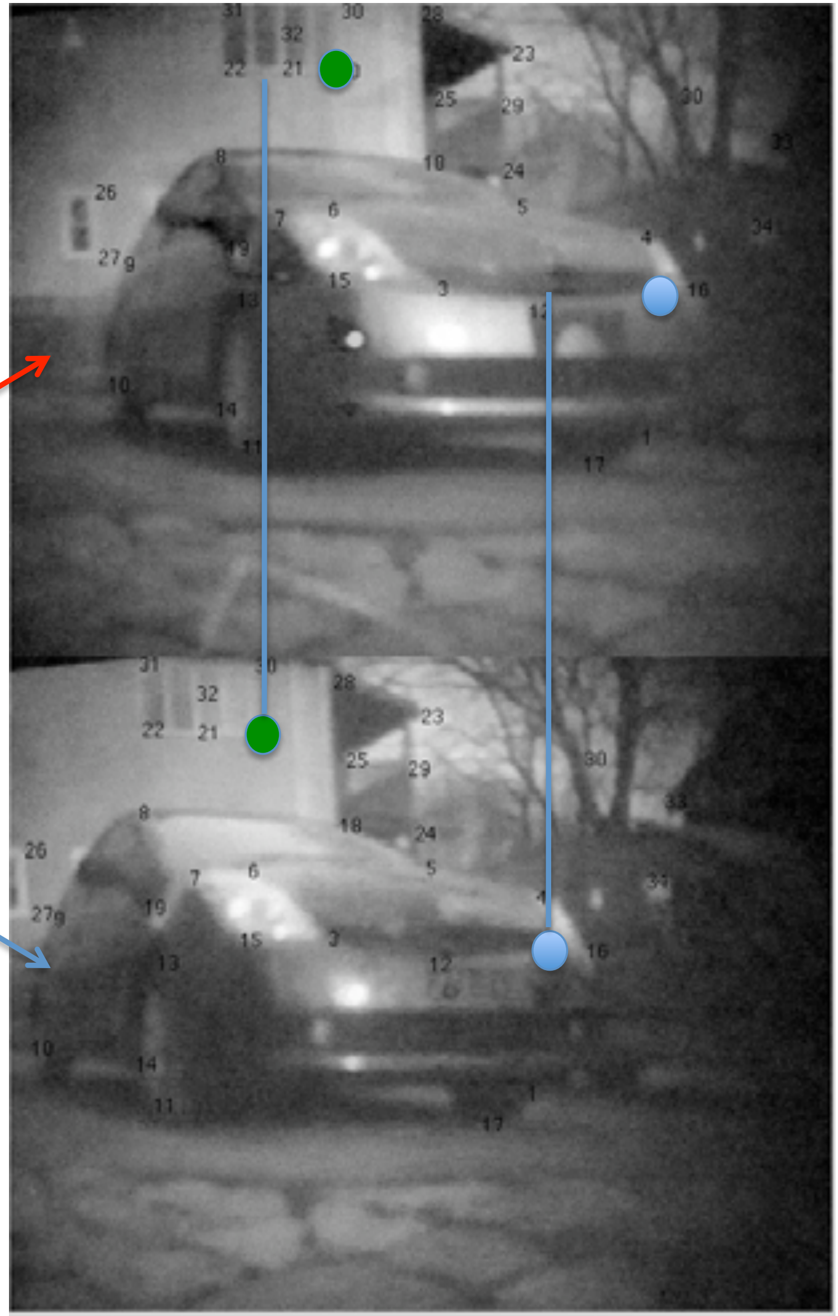




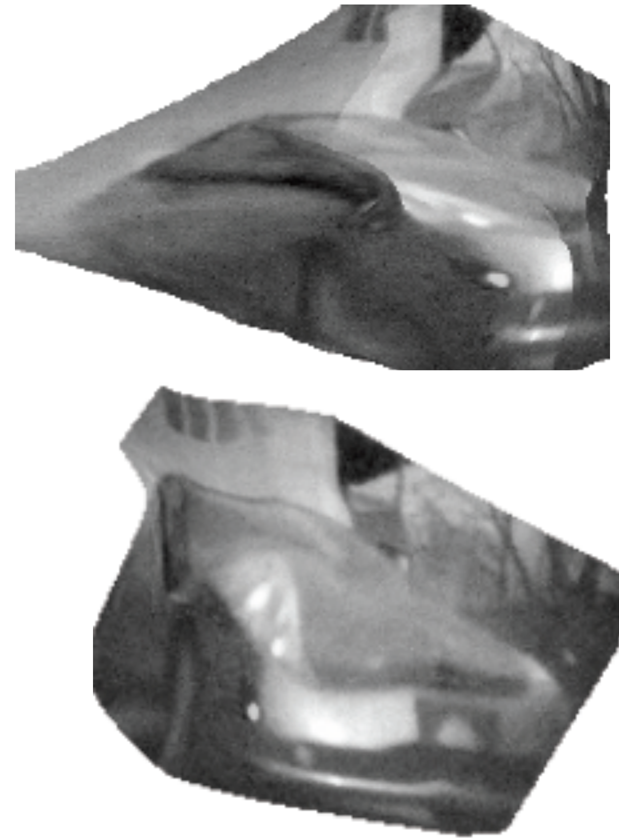
Anaglyph pinhole camera



Anaglyph



Synthesis of new views



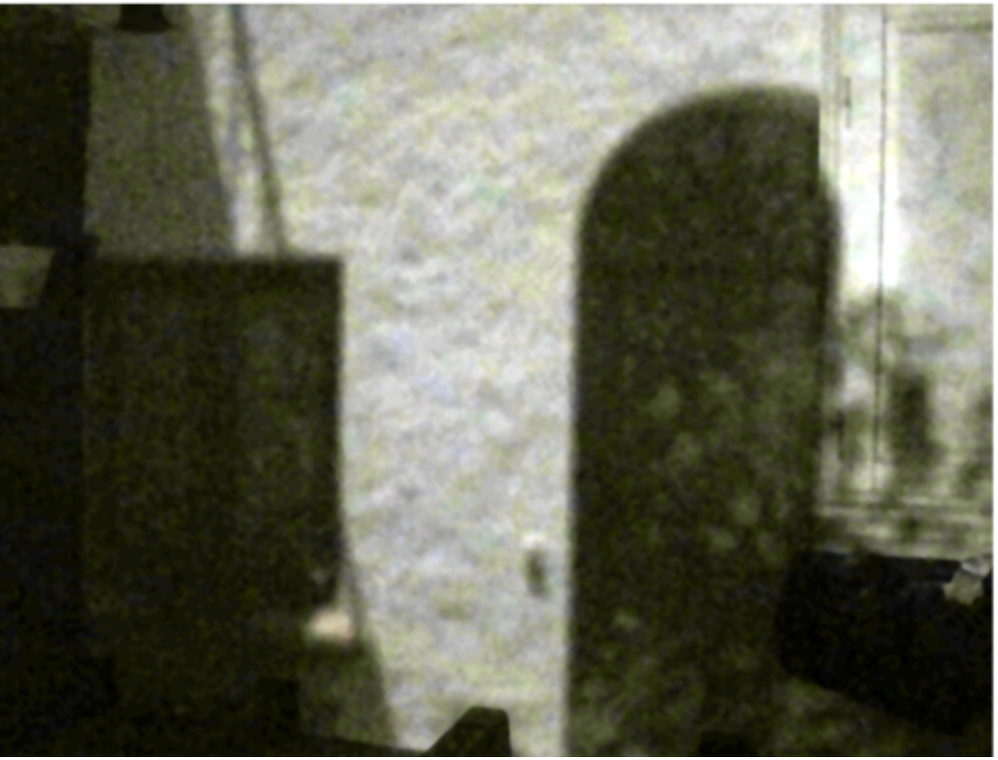
Problem set

- 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


Shadows?

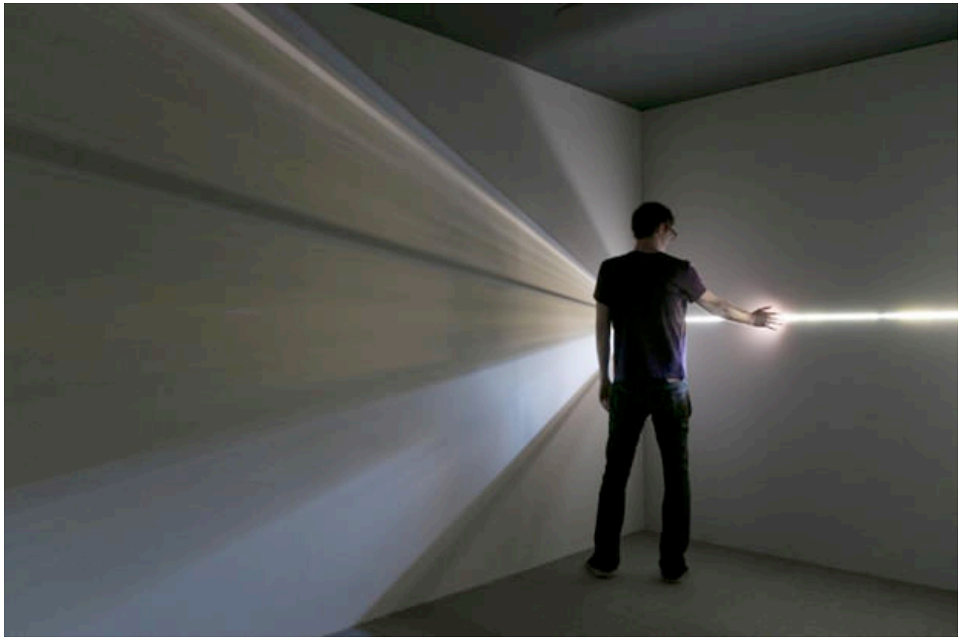
Accidental pinhole camera

Window turned into a pinhole



View outside

Source: wikipedia



Chris Fraser

”a camera obscura has been used ... to bring images from the outside into a darkened room”

Aberlado Morell

Window open

Window turned into a pinhole

Making a pinhole with home materials

Making a pinhole with home materials

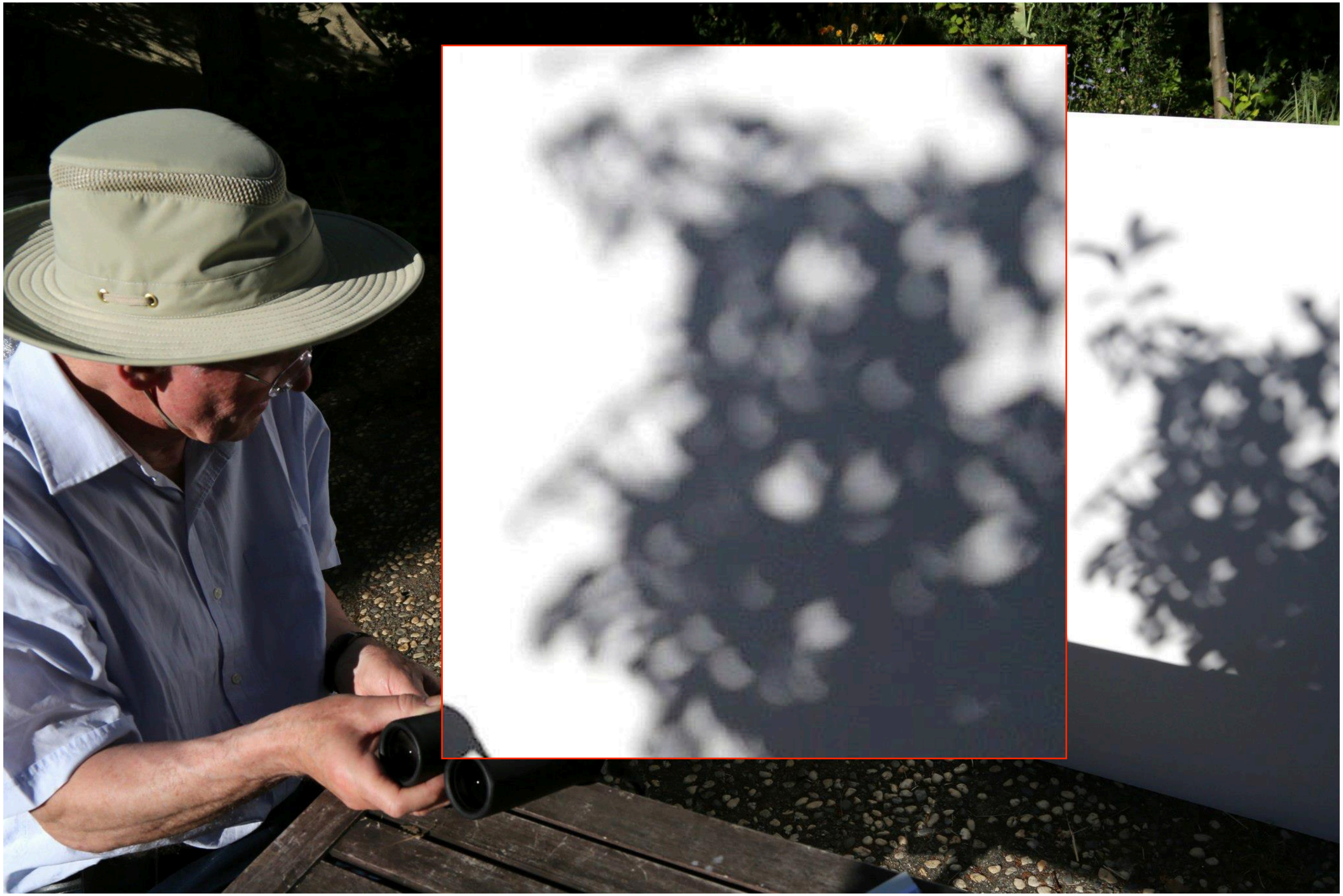
An hotel room, contrast enhanced.

The view from my window

Accidental pinholes produce images that are unnoticed or misinterpreted as shadows

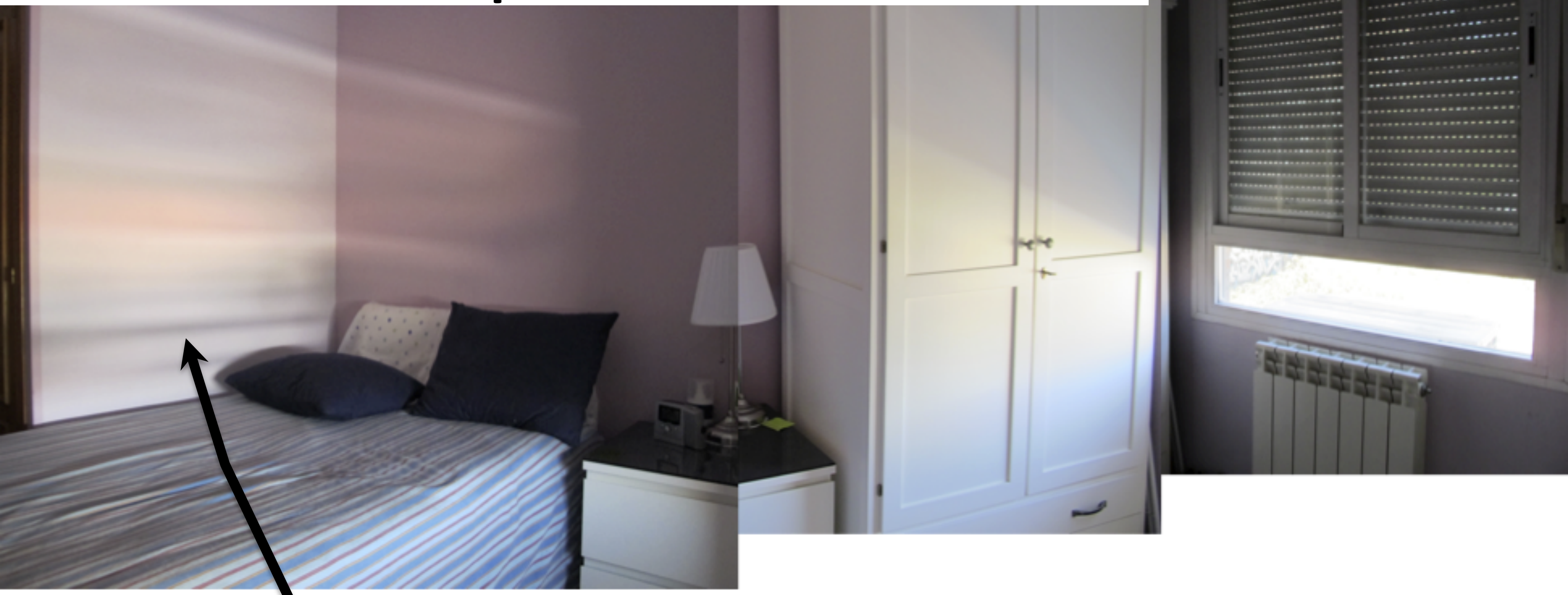
Another hotel room


Accidental pinholes in outdoor scenes



Pierre Moreels father (source: facebook)

Accidental pinhole camera



Outside scene

*

Aperture

See Zomet, A.; Nayar, S.K. CVPR 2006 for a detailed analysis.

Visualizing the convolution

Anti-pinhole or Pinspeck cameras

Adam L. Cohen, 1982

OPTICA ACTA, 1982, VOL. 29, NO. 1, 63–67

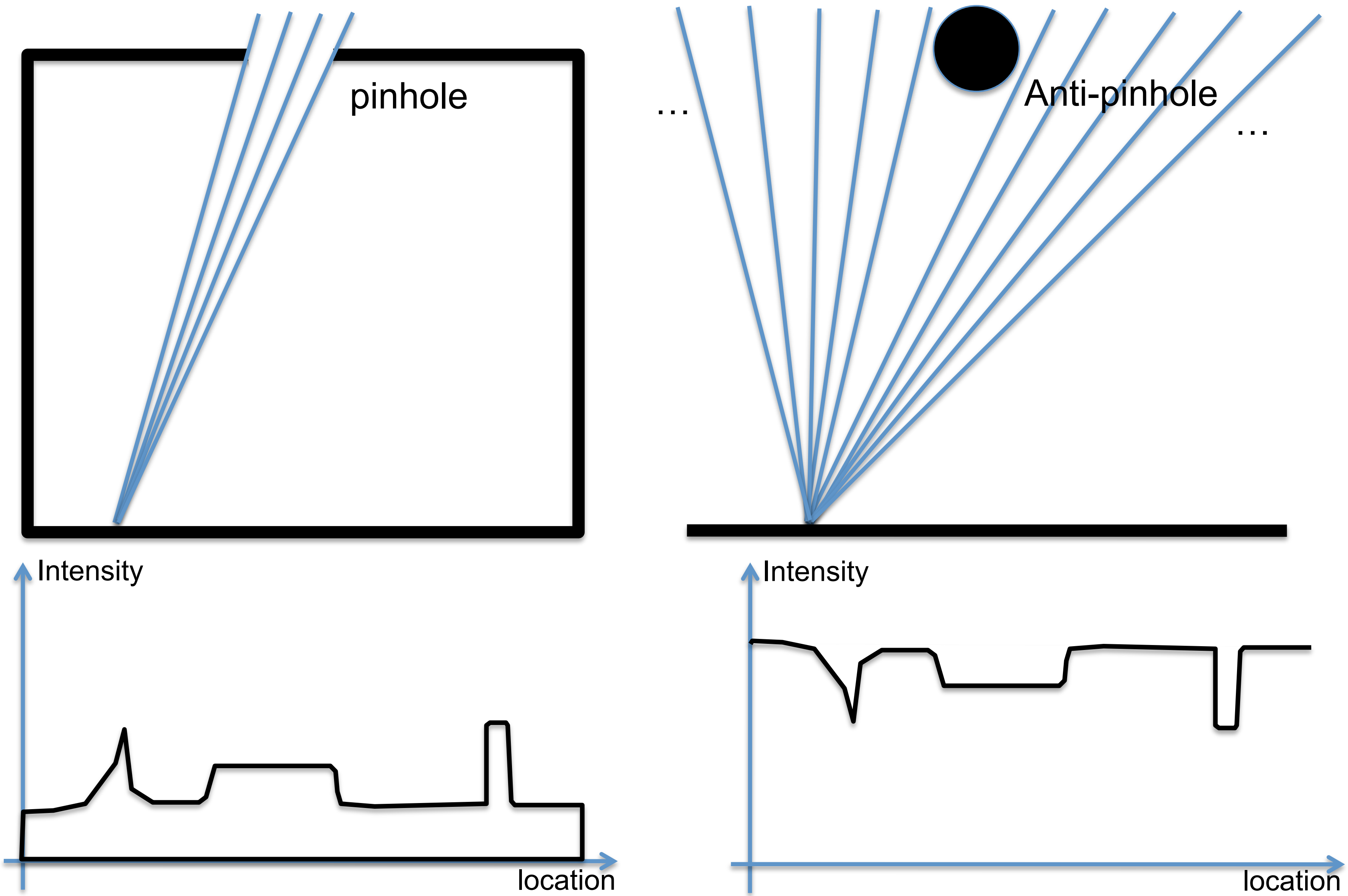
Anti-pinhole imaging

ADAM LLOYD COHEN
Parmly Research Institute, Loyola University of Chicago,
Chicago, Illinois 60626, U.S.A.

(*Received 16 April 1981; revision received 8 July 1981*)

Abstract. By complementing a pinhole to produce an isolated opaque spot, the light ordinarily blocked from the pinhole image is transmitted, and the light ordinarily transmitted is blocked. A negative geometrical image is formed, distinct from the familiar 'bright-spot' diffraction image. Anti-pinhole, or 'pinspeck' images are visible during a solar eclipse, when the shadows of objects appear crescent-shaped. Pinspecks demonstrate unlimited depth of field, freedom from distortion and large angular field. Images of different magnification may be formed simultaneously. Contrast is poor, but is improvable by averaging to remove noise and subtraction of a d.c. bias. Pinspecks may have application in X-ray space optics, and might be employed in the eyes of simple organisms.

Pinhole and Anti-pinhole cameras



Adam L. Cohen, 1982

Natural eyes

Lenses

Pinholes

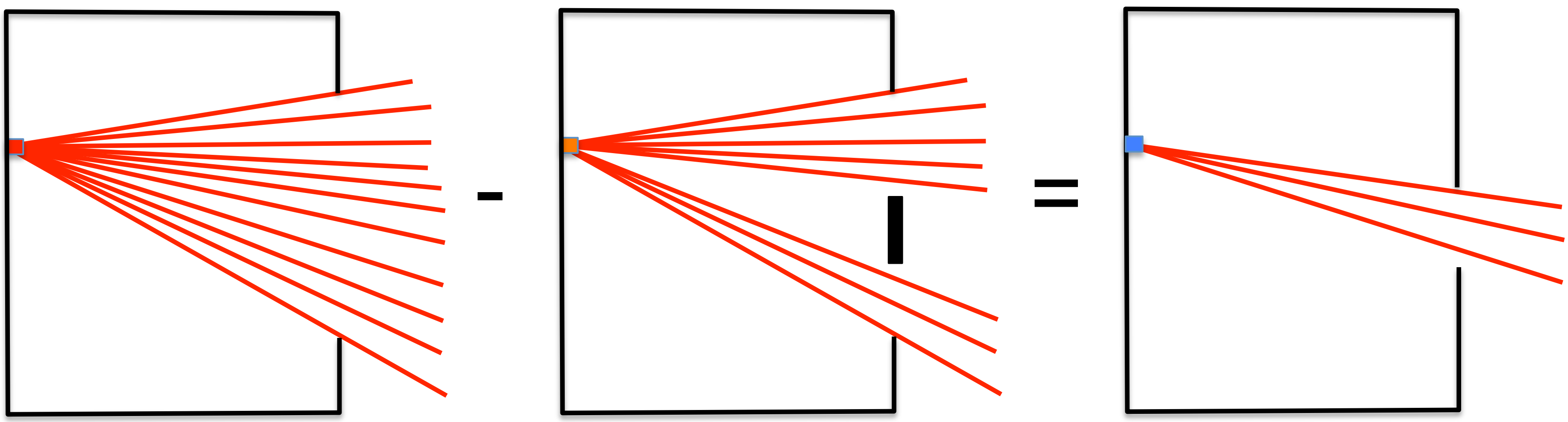
Anti-pinholes

Mixed accidental pinhole and anti-pinhole cameras

Mixed accidental pinhole and anti-pinhole cameras

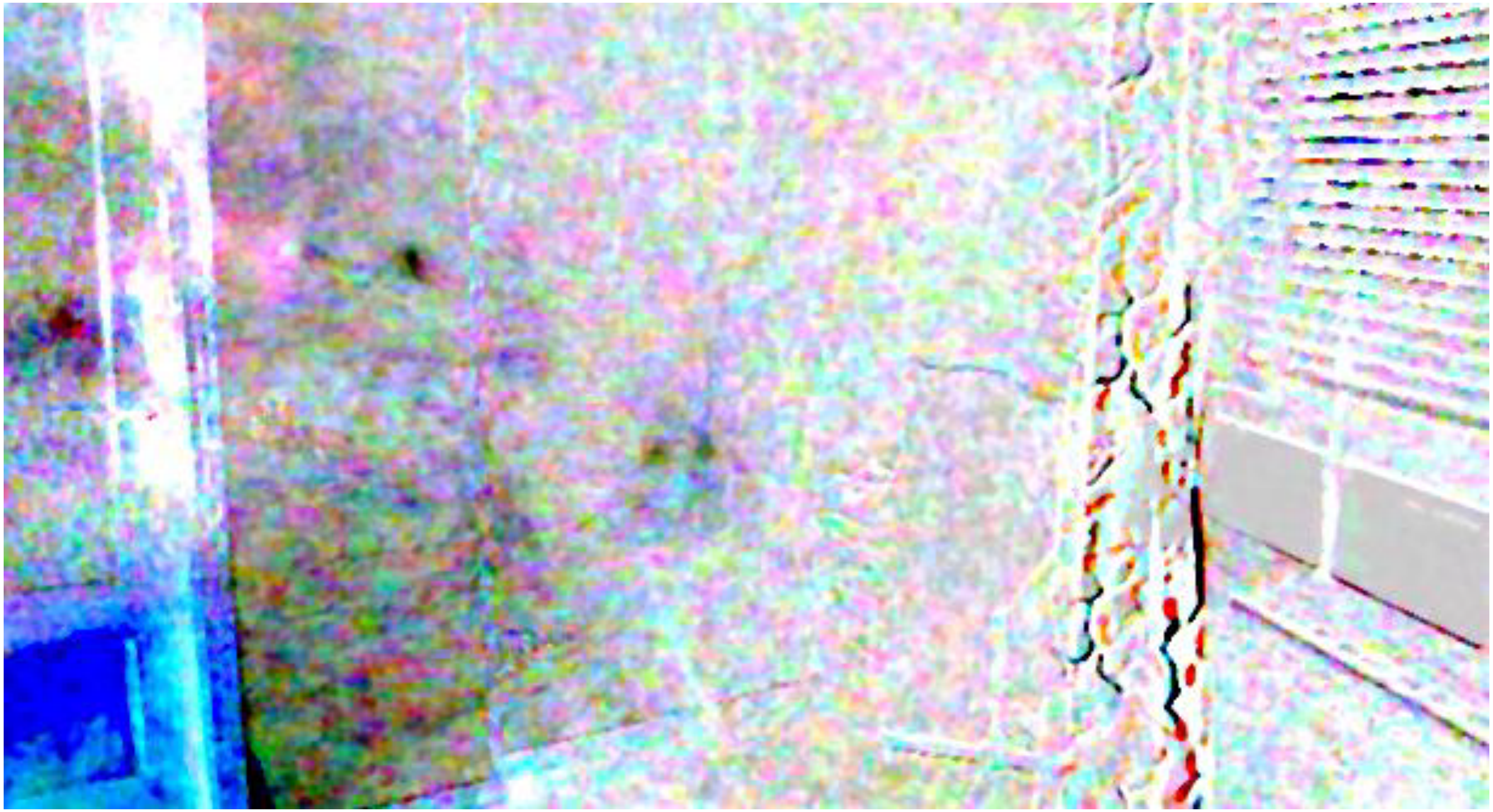
Mixed accidental pinhole and anti-pinhole cameras

Room with a window | Person in front of the window | Difference image



= ?

Mixed accidental pinhole and anti-pinhole cameras

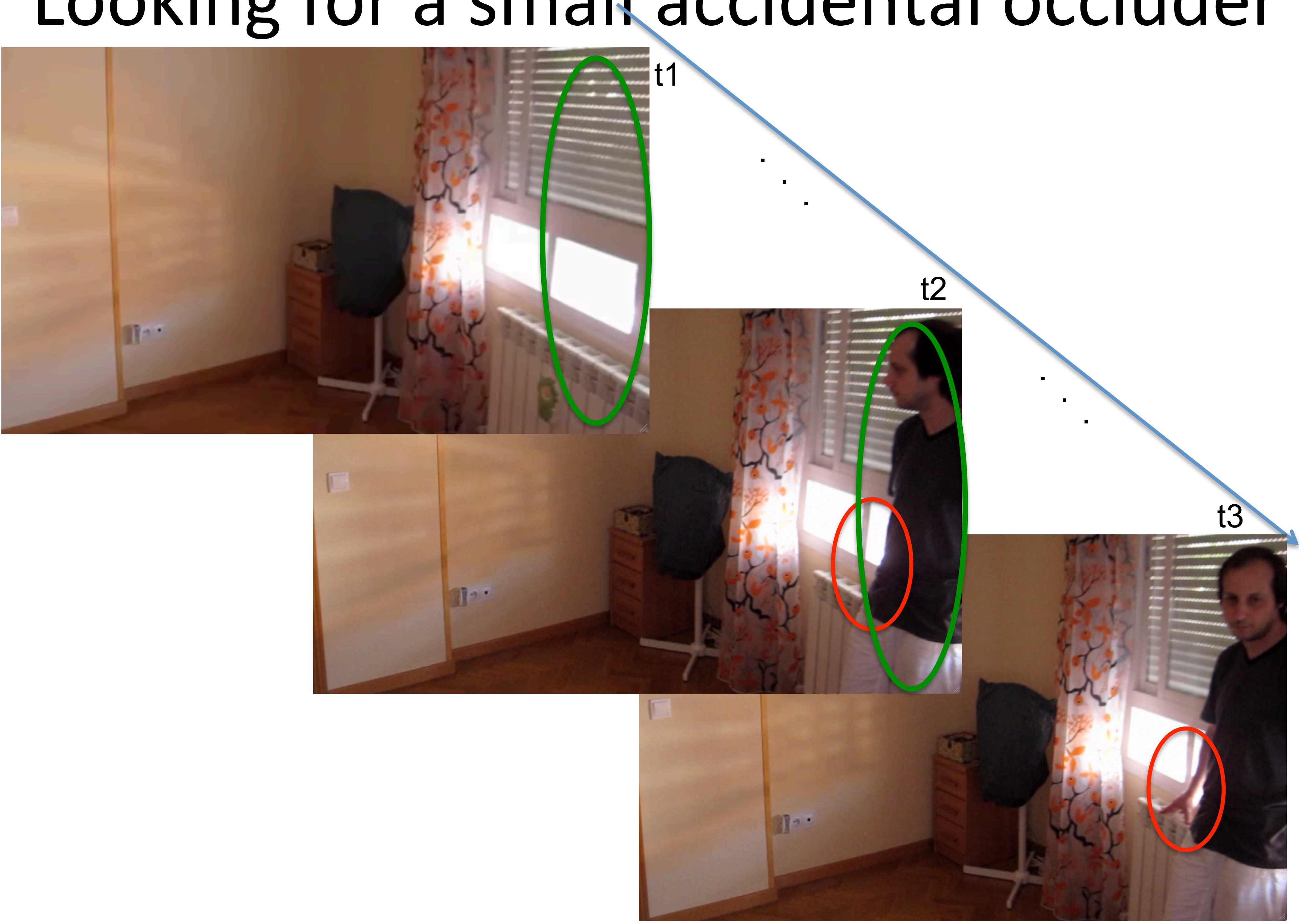


Mixed accidental pinhole and anti-pinhole cameras

Body as the occluder

View outside the window

Looking for a small accidental occluder



Reference Video

-

=

Looking for a small accidental occluder

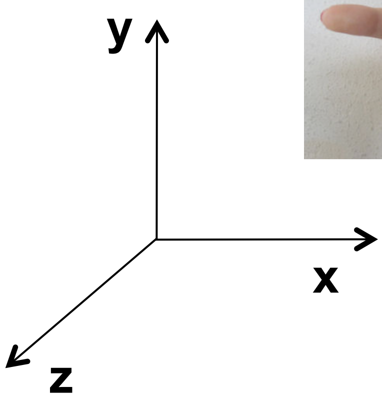
Body as the occluder

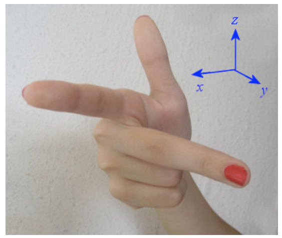
Hand as the occluder

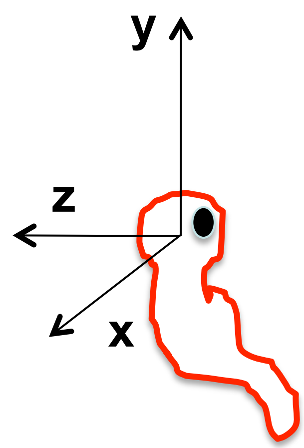
View outside the window

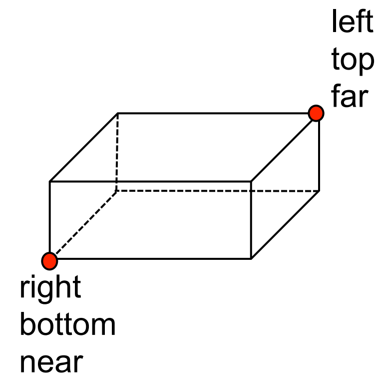
Camera Models

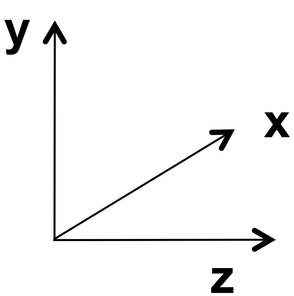
Right - handed system

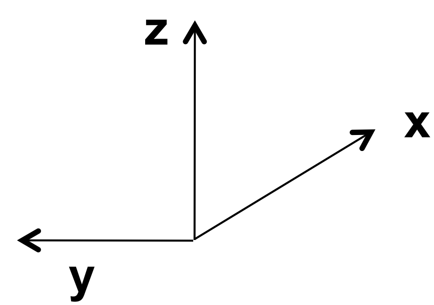




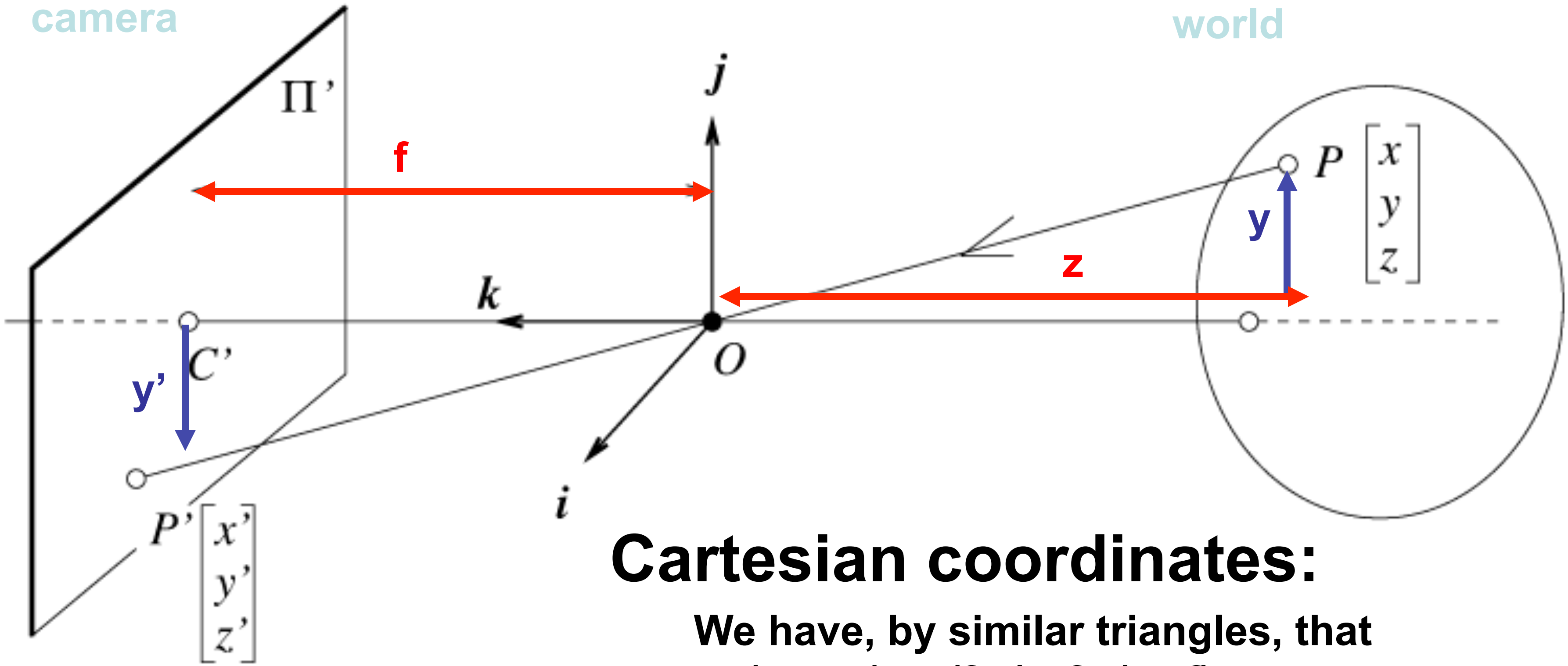








Perspective projection



Cartesian coordinates:

We have, by similar triangles, that

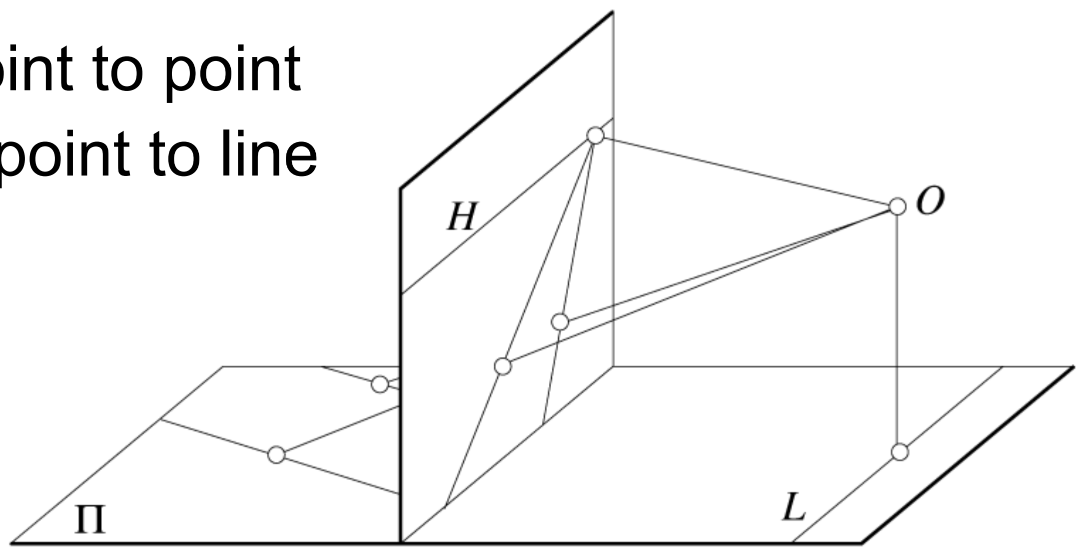
(x, y, z) -> (f x/z, f y/z, -f)

Ignore the third coordinate, and get

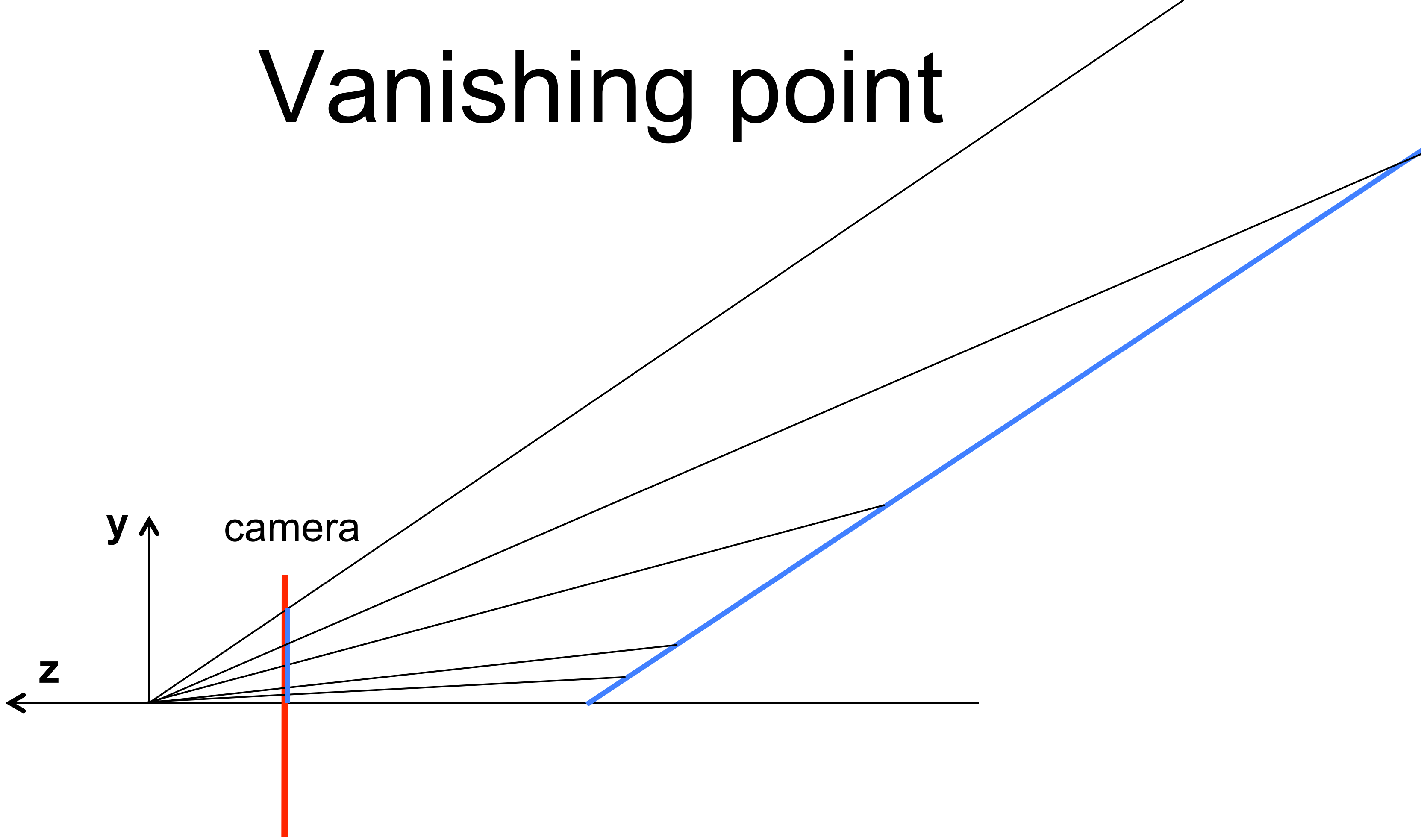
$$(x, y, z) \rightarrow (f\frac{x}{z}, f\frac{y}{z})$$

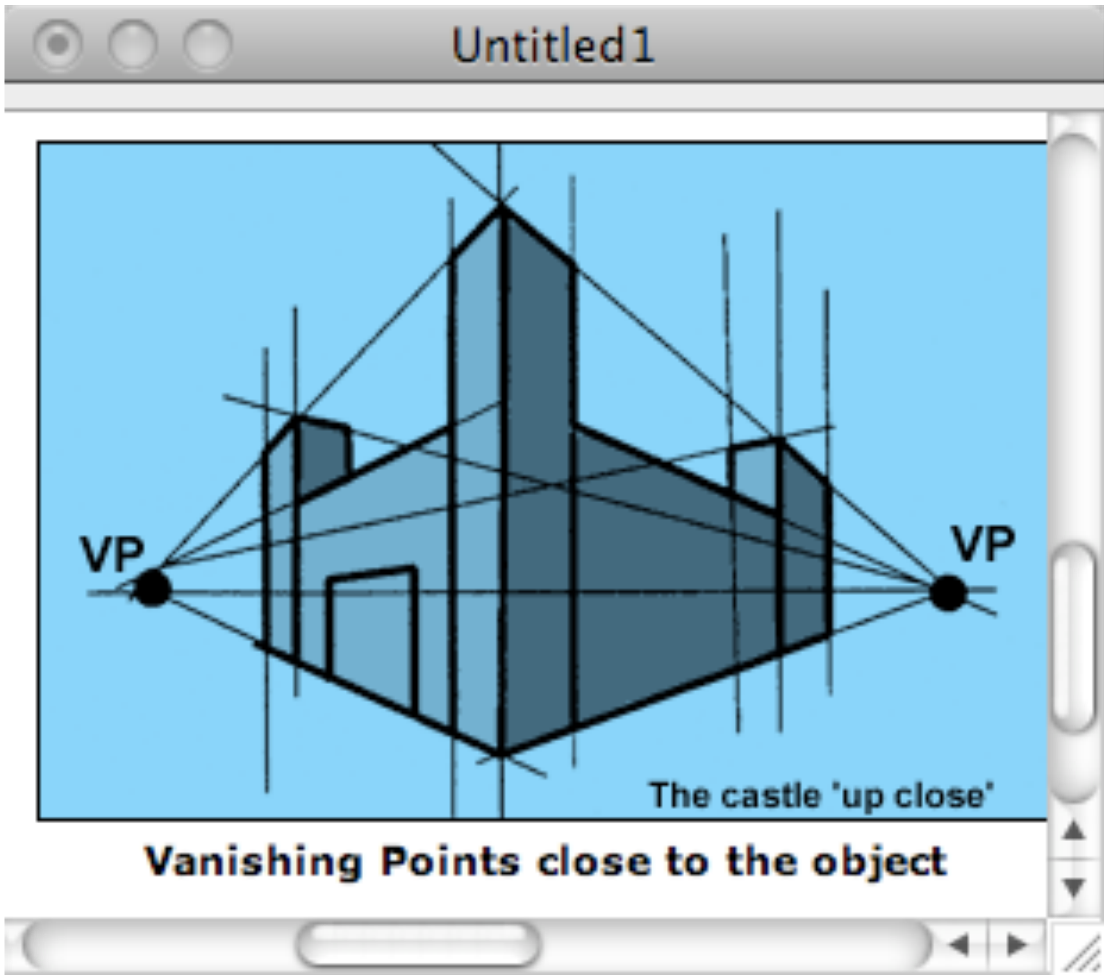
Geometric properties of projection

- 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



Vanishing point

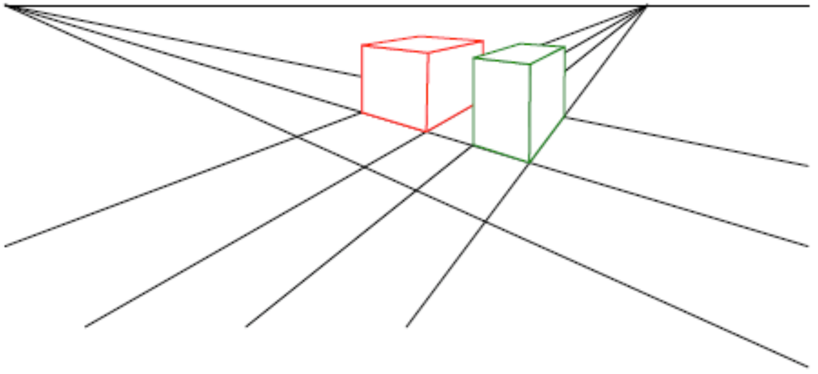




http://www.ider.herts.ac.uk/school/courseware/graphics/two_point_perspective.html

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



Line in 3-space

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

Perspective projection of that line

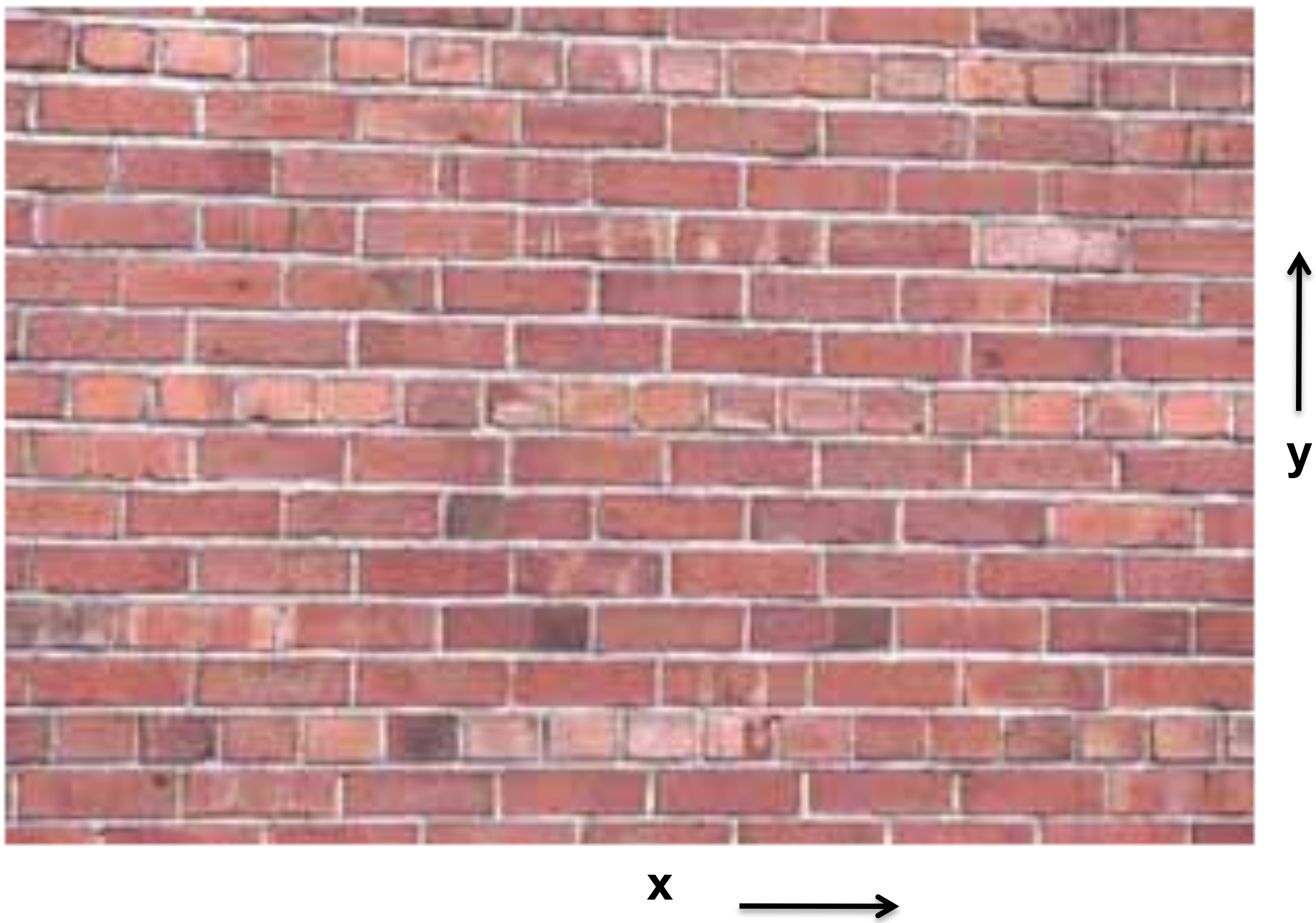
$$x'(t) = \frac{fx}{z} = \frac{f(x_0 + at)}{z_0 + ct}$$
$$y'(t) = \frac{fy}{z} = \frac{f(y_0 + bt)}{z_0 + ct}$$

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).

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

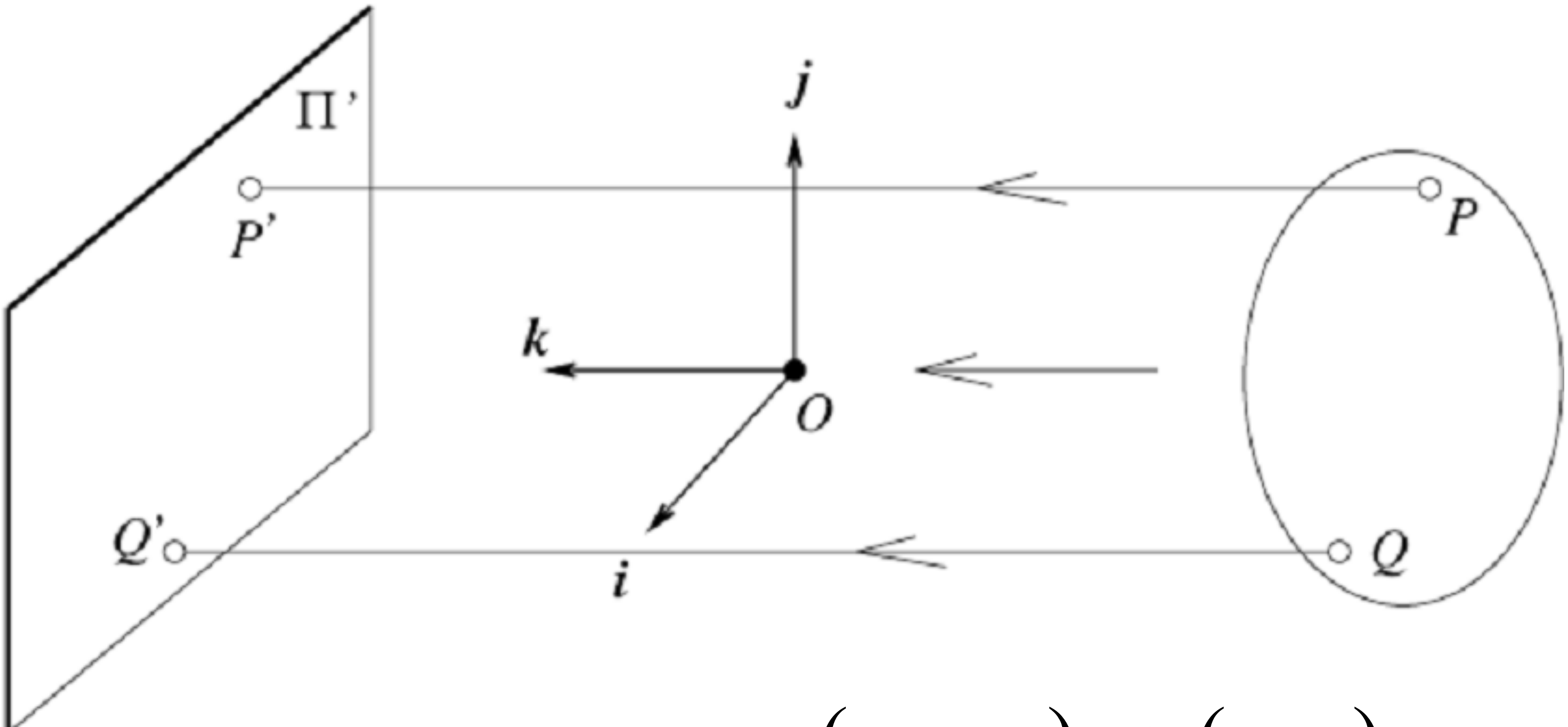
$$x'(t) = \frac{f \cdot (x_0 + at)}{z_0}$$

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

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.

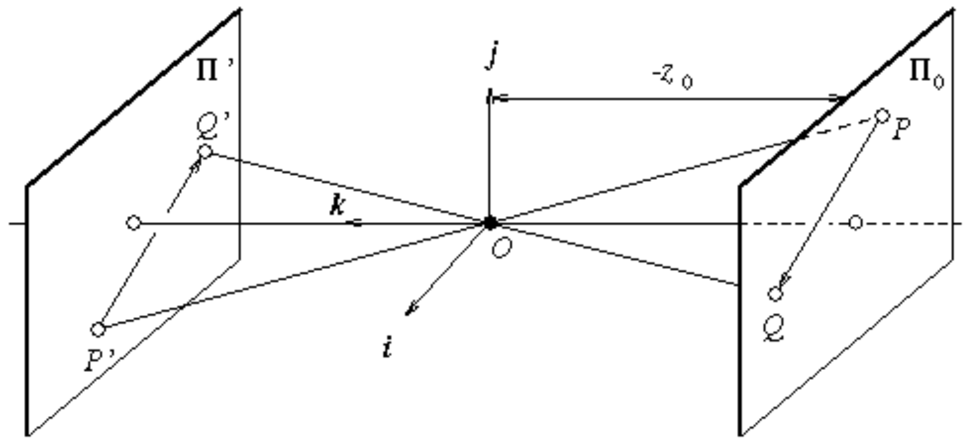
Other projection models: Orthographic projection



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

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

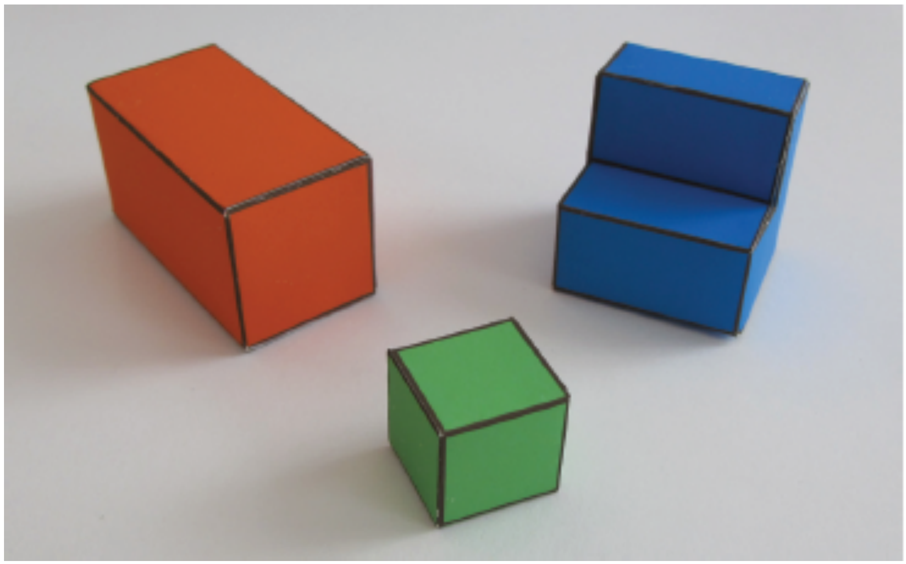
3-d point → **2-d image position**

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

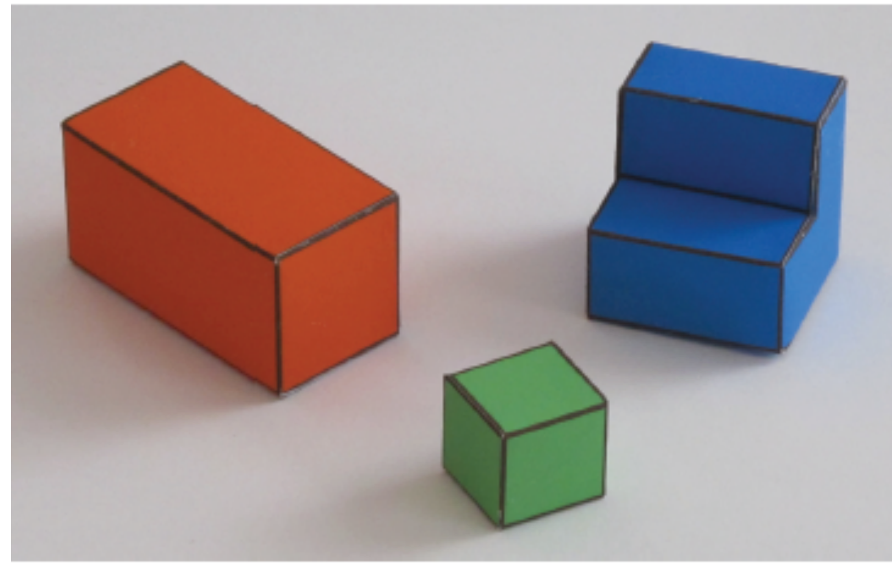
(2) Weak perspective: $$(x, y, z) \rightarrow \left(\frac{fx}{z_0}, \frac{fy}{z_0} \right)$$

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

Three camera projections



Perspective projection



Parallel (orthographic) projection

Weak perspective?

Homogeneous coordinates

- Is this a linear transformation?
  - **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

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

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

Slide by Steve Seitz

Perspective Projection

- Projection is a matrix multiply using homogeneous coordinates:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z/f \end{bmatrix} \Rightarrow \left(f\frac{x}{z}, f\frac{y}{z} \right)$$

This is known as perspective projection

- **The matrix is the projection matrix**

Slide by Steve Seitz

Perspective Projection

How does scaling the projection matrix change the transformation?

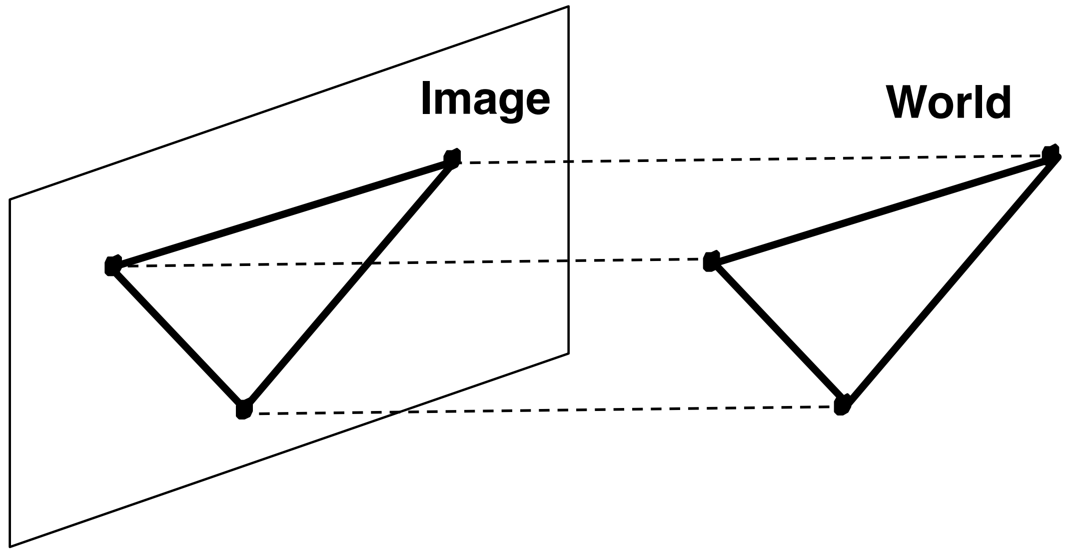
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z/f \end{bmatrix} \Rightarrow \left(f\frac{x}{z}, f\frac{y}{z} \right)$$

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} fx \\ fy \\ z \end{bmatrix} \Rightarrow \left(f\frac{x}{z}, f\frac{y}{z} \right)$$

Orthographic Projection

Special case of perspective projection

- Distance from the COP to the PP is infinite



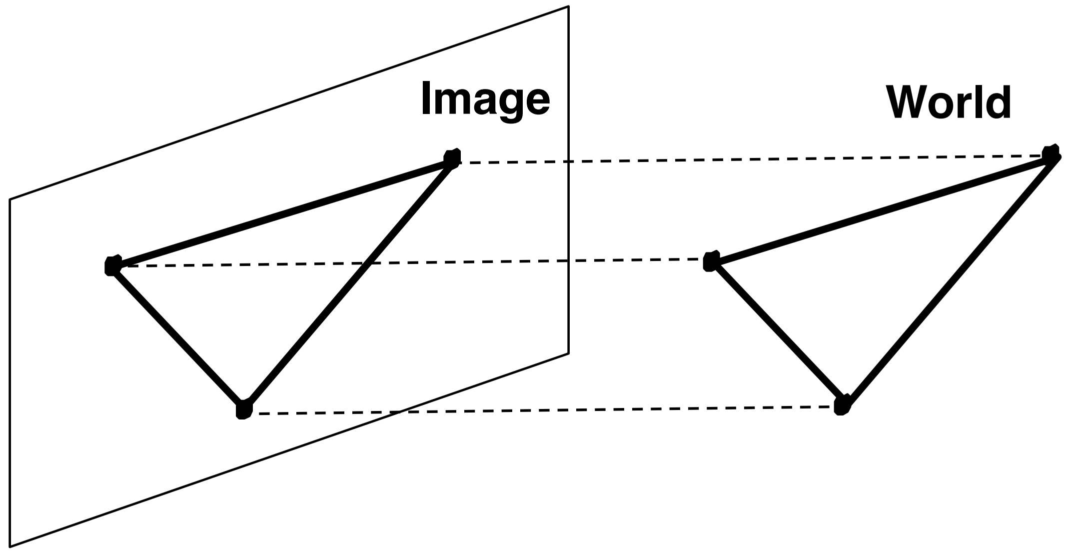
- Also called “parallel projection”
- What’s the projection matrix?

$$\boxed{\quad ? \quad} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Rightarrow (x, y)$$

Orthographic Projection

Special case of perspective projection

- Distance from the COP to the PP is infinite



- Also called “parallel projection”
- What’s the projection matrix?

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

Slide by Steve Seitz

Homogeneous coordinates

2D Points:

$$p = \begin{bmatrix} x \\ y \end{bmatrix} \longrightarrow p' = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

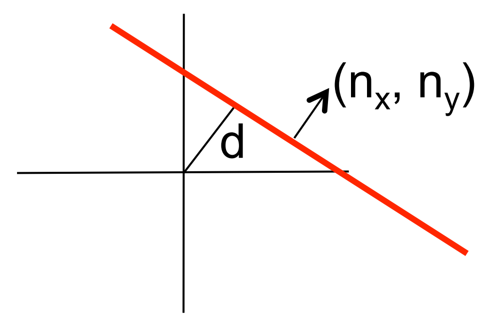
$$p' = \begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} \longrightarrow p = \begin{bmatrix} x'/w' \\ y'/w' \end{bmatrix}$$

2D Lines:

$$ax + by + c = 0$$

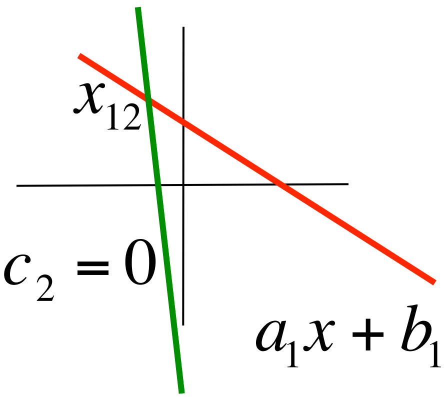
$$\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

$$l = \begin{bmatrix} a & b & c \end{bmatrix} \Rightarrow \begin{bmatrix} n_x & n_y & d \end{bmatrix}$$



Homogeneous coordinates

Intersection between two lines:



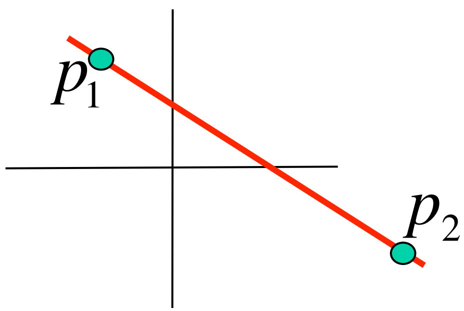
$$a_2 x + b_2 y + c_2 = 0$$

$$a_1 x + b_1 y + c_1 = 0$$

$$\left.\begin{array}{l} l_1 = \begin{bmatrix} a_1 & b_1 & c_1 \end{bmatrix} \\ \\ l_2 = \begin{bmatrix} a_2 & b_2 & c_2 \end{bmatrix} \end{array}\right\} \quad x_{12} = l_1 \times l_2$$

Homogeneous coordinates

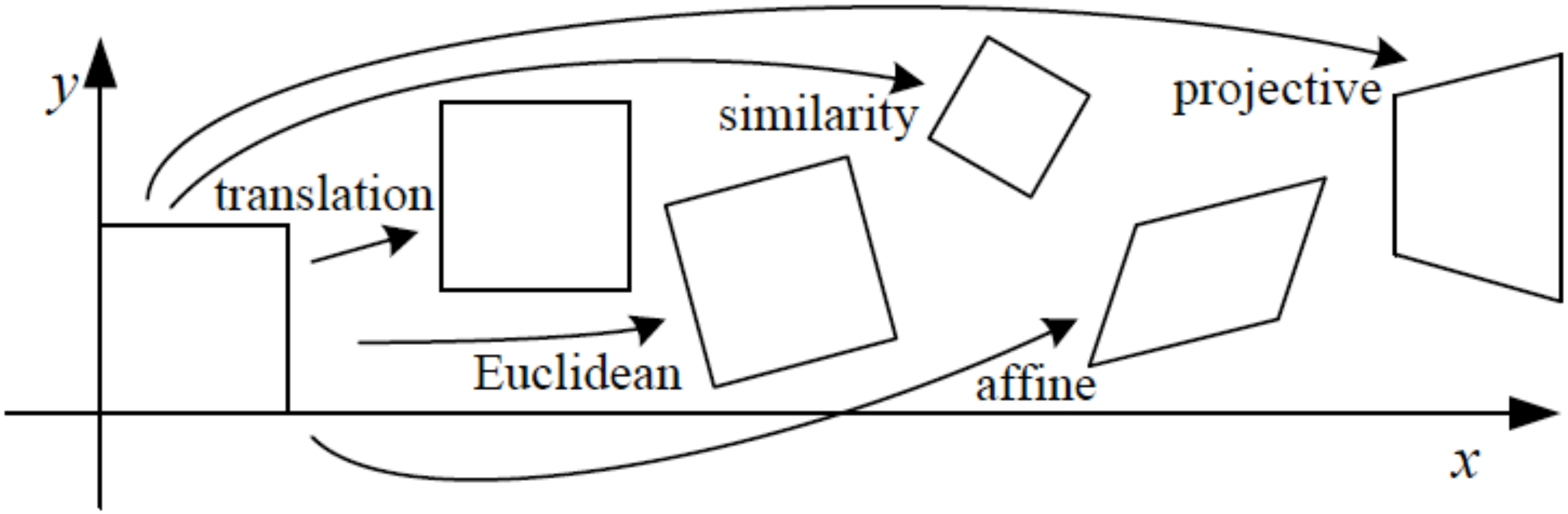
Line joining two points:



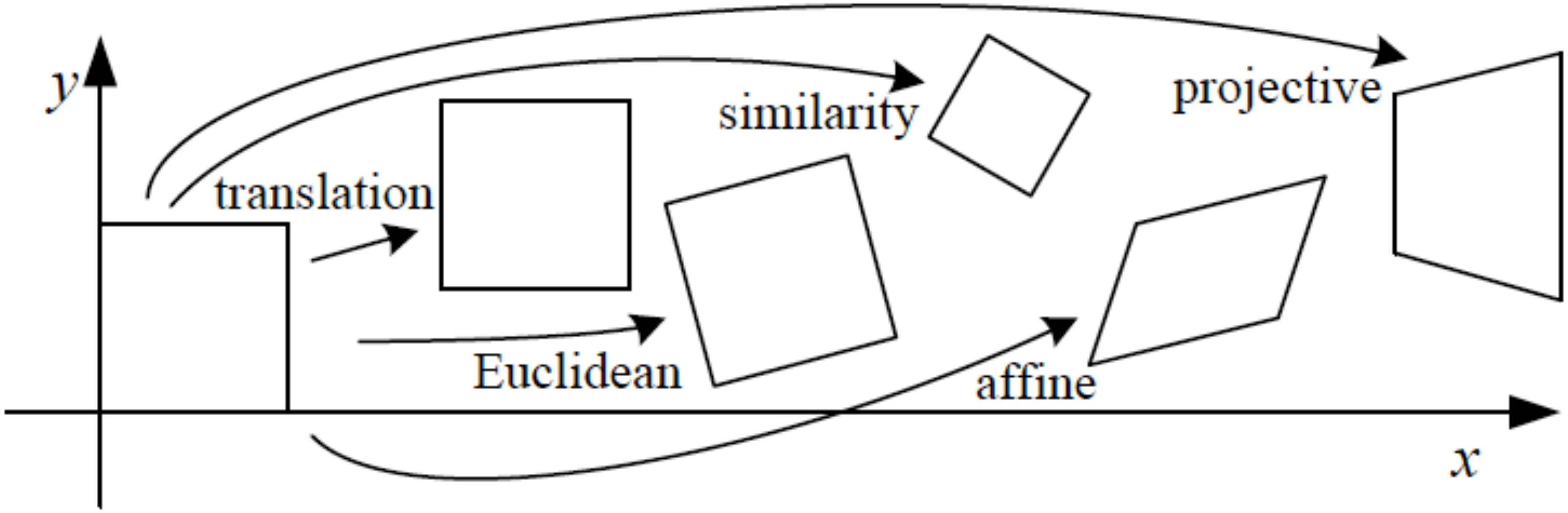
$$ax + by + c = 0$$

$$\left.\begin{array}{l} p_1 = \begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} \\ \\ p_2 = \begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \end{array}\right\} \; l = p_1 \times p_2$$

2D Transformations

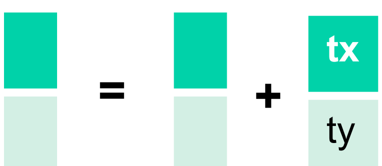


2D Transformations

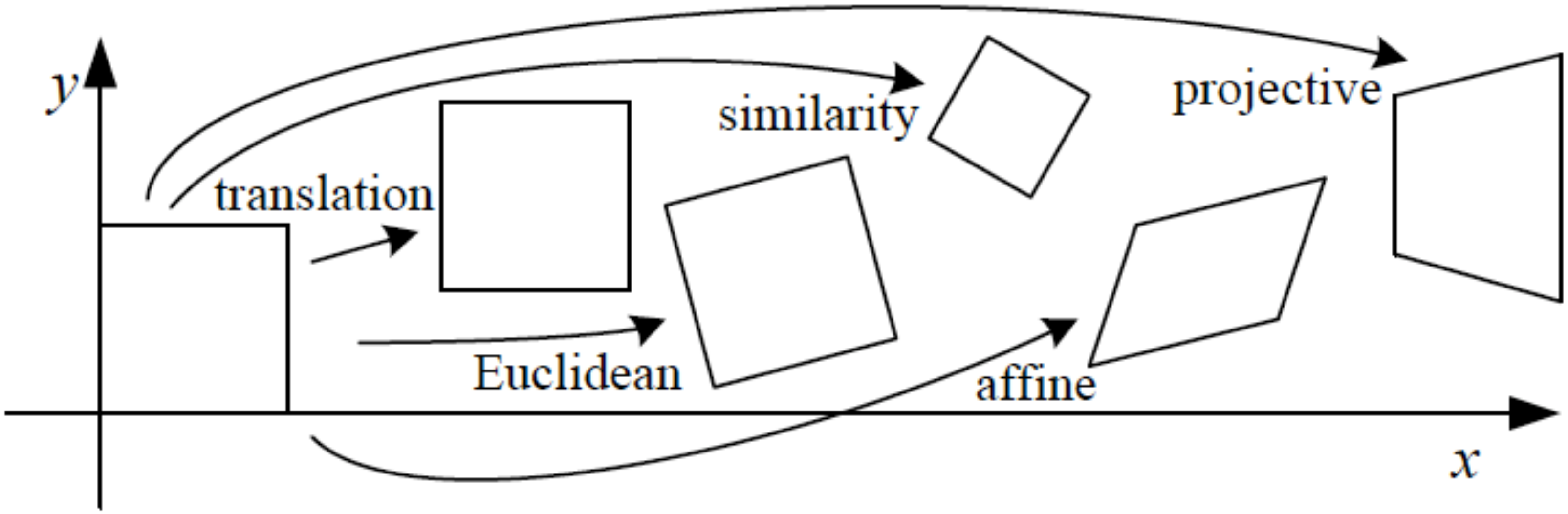


Example: translation

$$x' = x + t$$

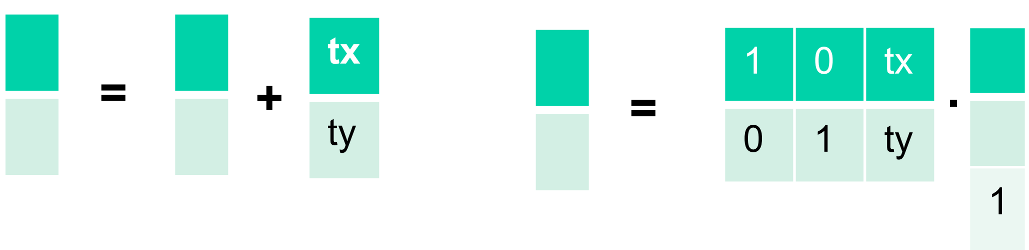


2D Transformations

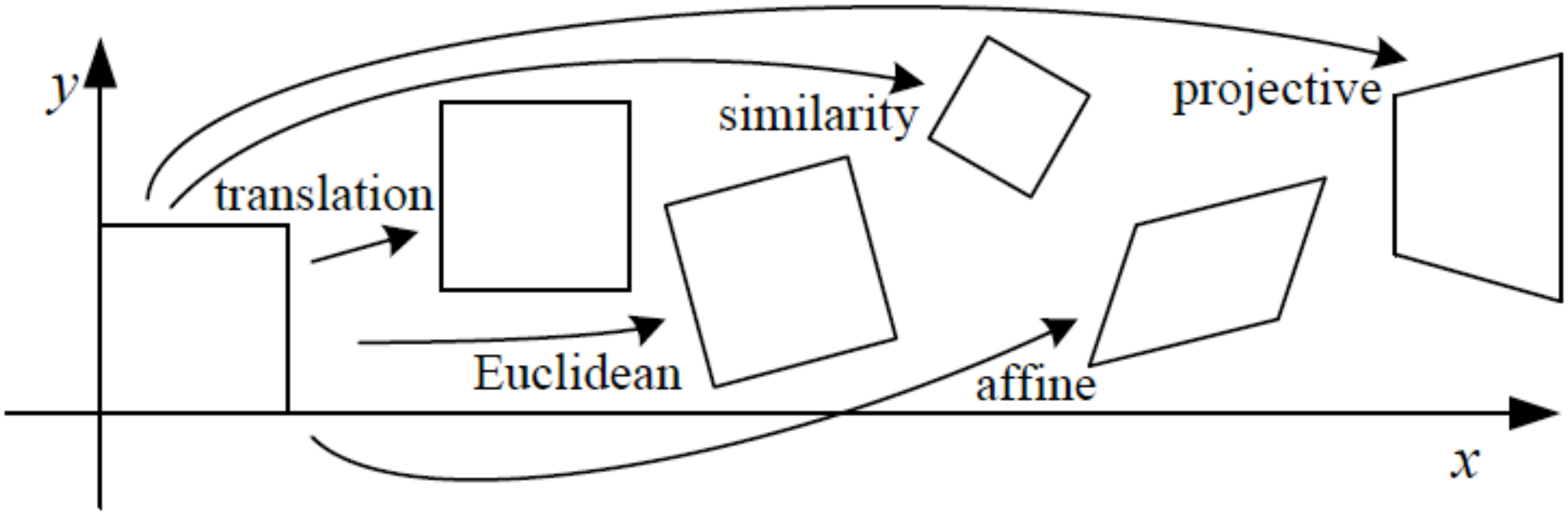


Example: translation

$$\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{t} \qquad \boldsymbol{x}' = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{t} \end{bmatrix} \bar{\boldsymbol{x}}$$

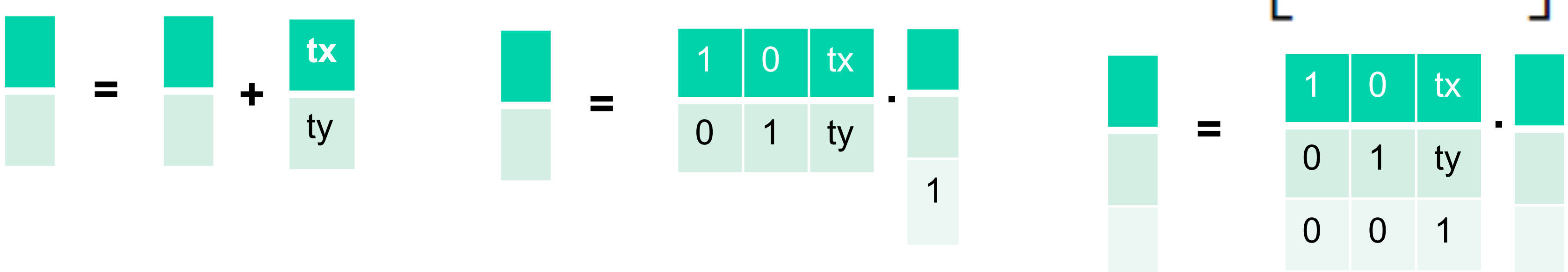


2D Transformations



Example: translation

$$x' = x + t \qquad x' = \begin{bmatrix} I & t \end{bmatrix} \bar{x} \qquad \bar{x}' = \begin{bmatrix} I & t \\ 0^T & 1 \end{bmatrix} \bar{x}$$



Now we can chain transformations

Translation and rotation, written in each set of coordinates

Non-homogeneous coordinates

$$ {}^{B}\vec{p} = {}^{B}_{A}R \; {}^{A}\vec{p} + {}^{B}_{A}\vec{t} $$

Homogeneous coordinates

$$ {}^{B}\vec{p} = {}^{B}_{A}C \; {}^{A}\vec{p} $$

where

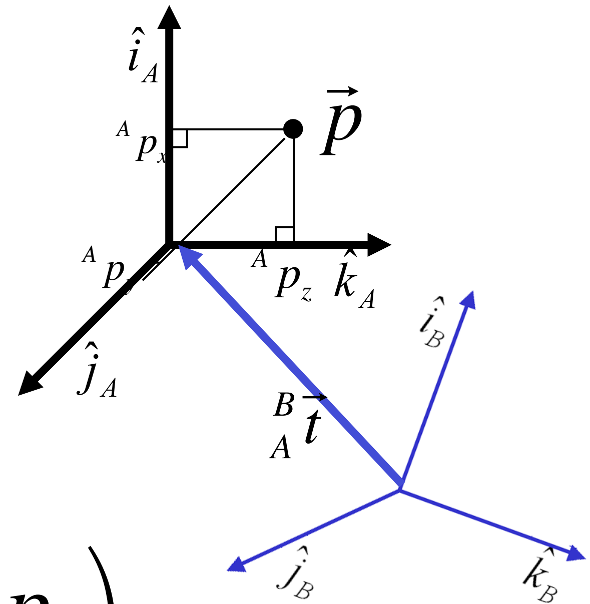
$$ {}^{B}_{A}C = \begin{pmatrix} - & - & - & | \\ - & {}^{B}_{A}R & - & {}^{B}_{A}\vec{t} \\ - & - & - & | \\ 0 & 0 & 0 & 1 \end{pmatrix} $$

Translation and rotation

"as described in the coordinates of frame B"

Let's write

$$ {}^{B}\vec{p} = {}^{B}_{A}R \; {}^{A}\vec{p} + {}^{B}_{A}\vec{t} $$



as a single matrix equation:

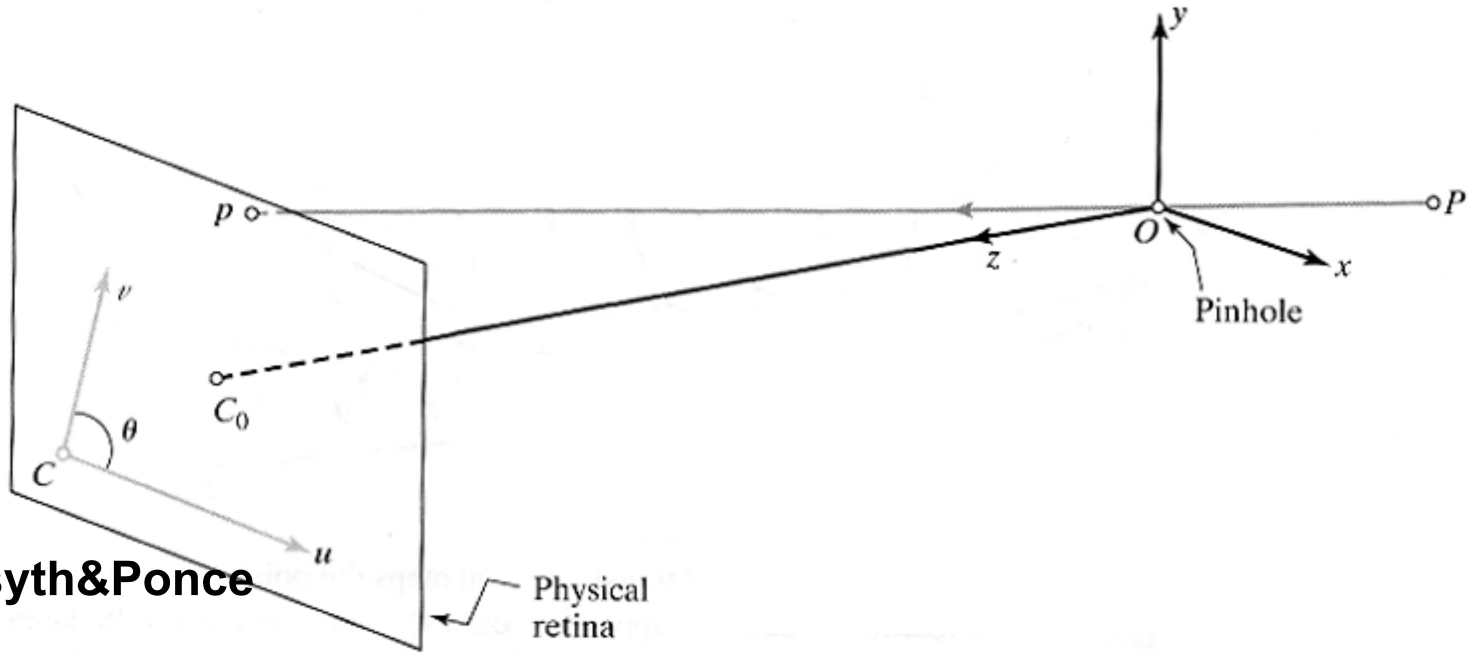
$$ \begin{pmatrix} {}^{B}p_x \\ {}^{B}p_y \\ {}^{B}p_z \\ 1 \end{pmatrix} = \begin{pmatrix} - & - & - & | \\ - & {}^{B}_{A}R & - & {}^{B}_{A}t \\ - & - & - & | \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} {}^{A}p_x \\ {}^{A}p_y \\ {}^{A}p_z \\ 1 \end{pmatrix} $$

Camera calibration

Use the camera to tell you things about the world:

- Relationship between coordinates in the world and coordinates in the image: *geometric camera calibration, see* Szeliski, section 5.2, 5.3 for references
- (Relationship between intensities in the world and intensities in the image: *photometric image formation*, see Szeliski, sect. 2.2.)

Intrinsic parameters: from idealized world coordinates to pixel values



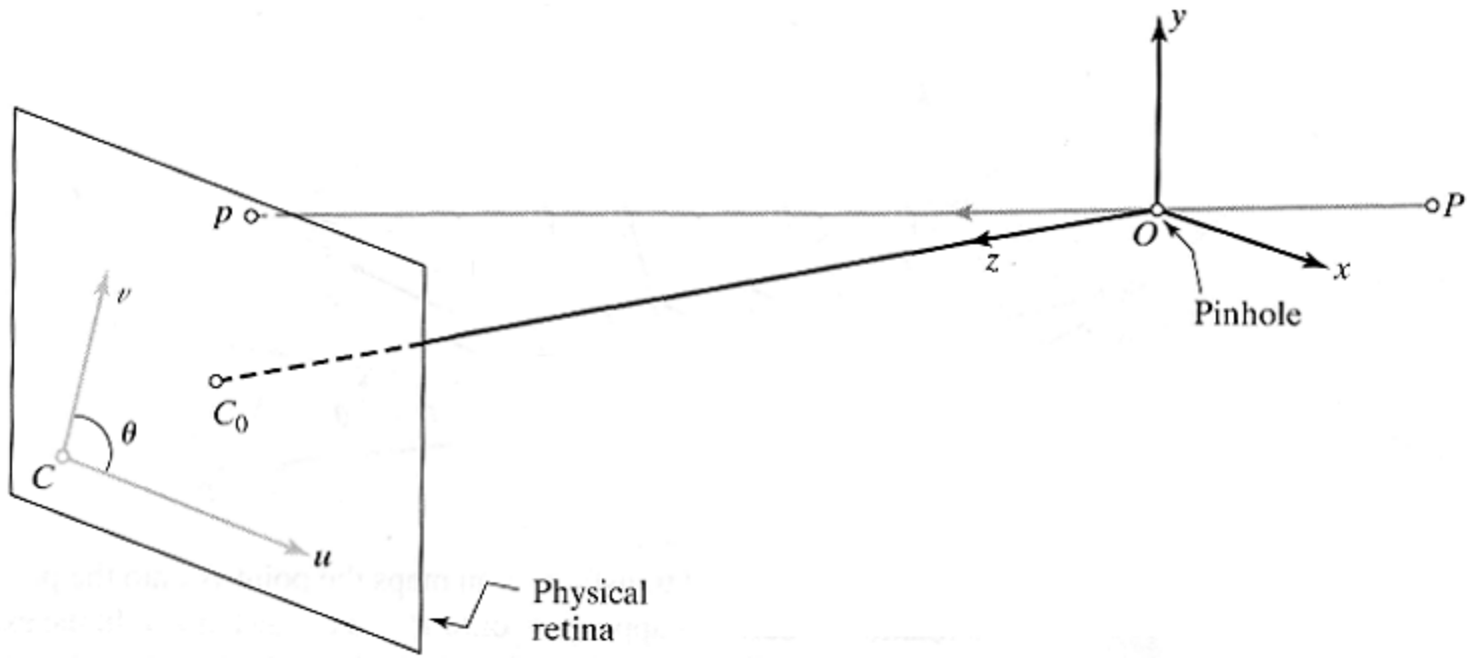
Forsyth&Ponce

Perspective projection

$$u = f\frac{x}{z}$$

$$v = f\frac{y}{z}$$

Intrinsic parameters

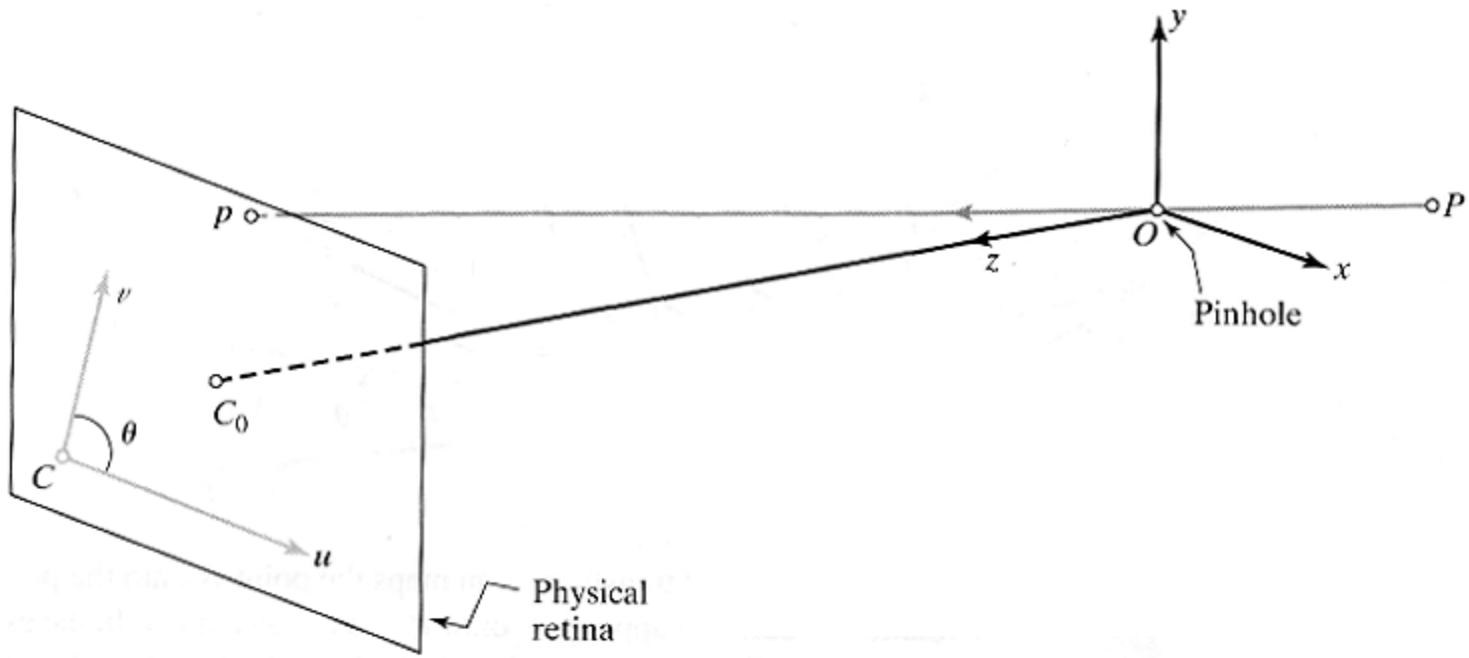


But "pixels" are in some arbitrary spatial units

$$u = \alpha \frac{x}{z}$$

$$v = \alpha \frac{y}{z}$$

Intrinsic parameters

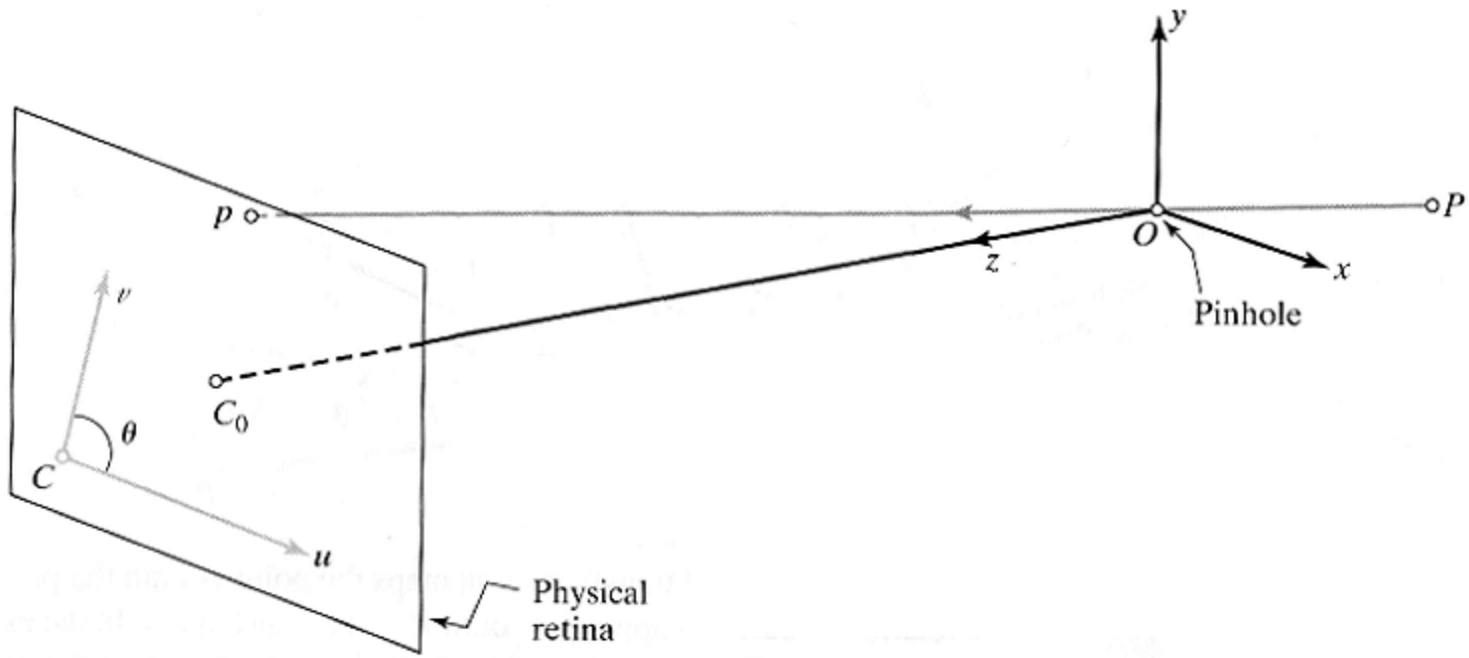


Maybe pixels are not square

$$u = \alpha \frac{x}{z}$$

$$v = \beta \frac{y}{z}$$

Intrinsic parameters

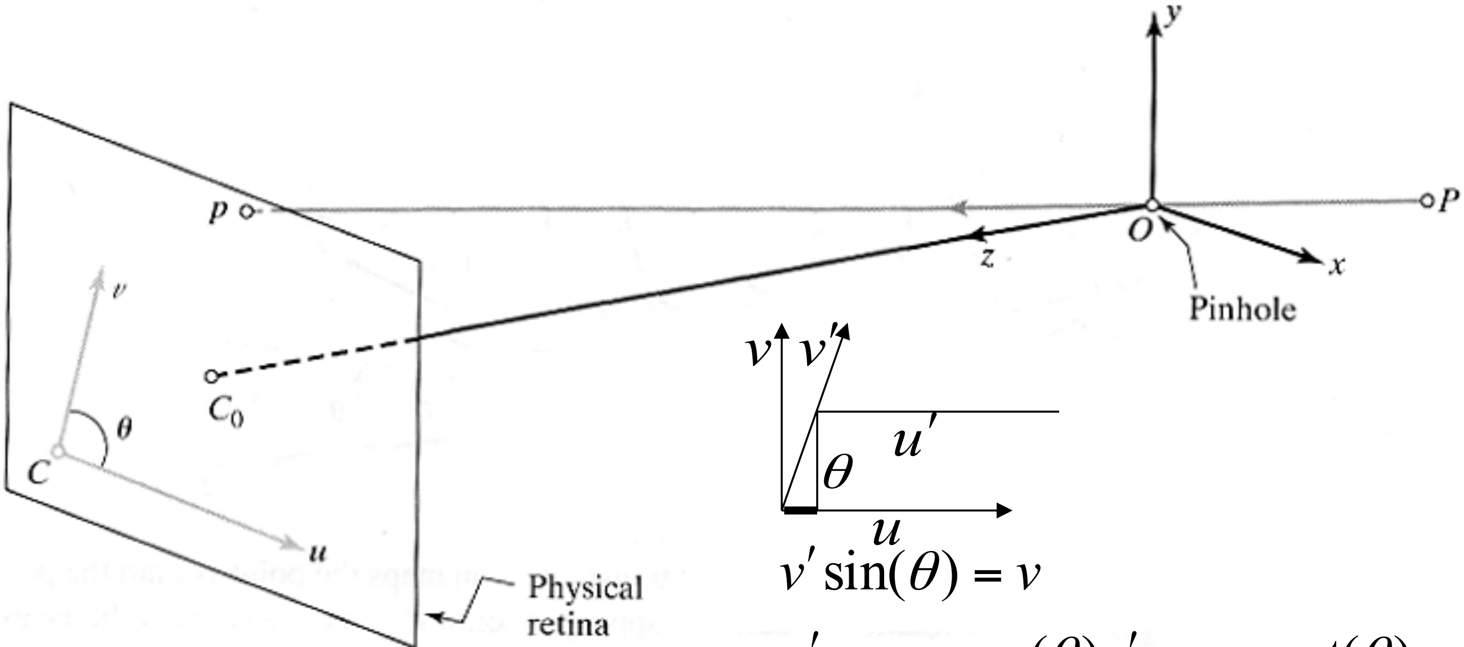


We don't know the origin of our camera pixel coordinates

$$u = \alpha \frac{x}{z} + u_0$$

$$v = \beta \frac{y}{z} + v_0$$

Intrinsic parameters



$$v' \sin(\theta) = v$$

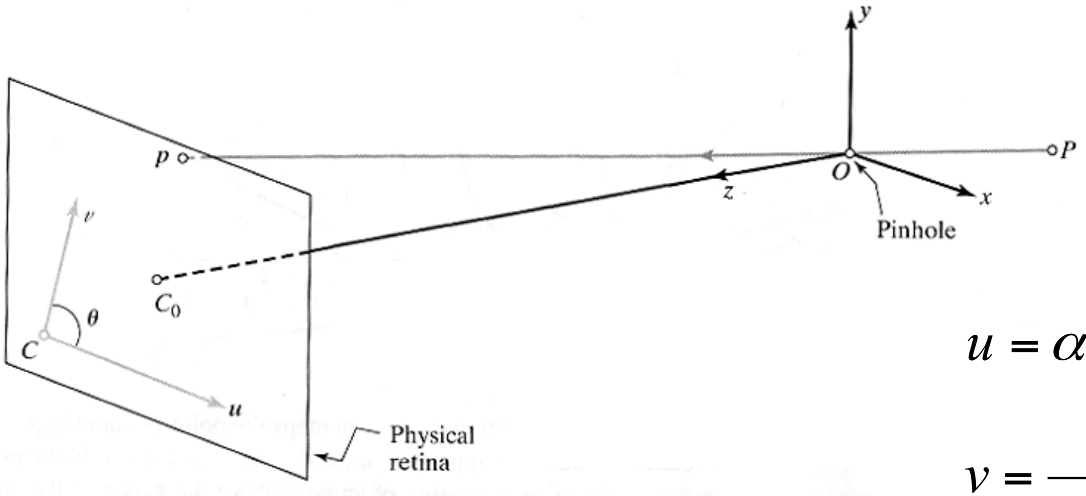
$$u' = u - \cos(\theta)v' = u - \cot(\theta)v$$

May be skew between camera pixel axes

$$u = \alpha \frac{x}{z} - \alpha \cot(\theta) \frac{y}{z} + u_0$$

$$v = \frac{\beta}{\sin(\theta)} \frac{y}{z} + v_0$$

Intrinsic parameters, homogeneous coordinates



$$u = \alpha \frac{x}{z} - \alpha \cot(\theta) \frac{y}{z} + u_0$$

$$v = \frac{\beta}{\sin(\theta)} \frac{y}{z} + v_0$$

Using homogenous coordinates, we can write this as:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & -\alpha\cot(\theta) & u_0 & 0 \\ 0 & \dfrac{\beta}{\sin(\theta)} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

or:

In pixels →

$$\vec{p} = \mathrm{K} \; {}^{C}\vec{p}$$

In camera-based coords

Extrinsic parameters: translation and rotation of camera frame

$$^{C}\vec{p} = {}^{C}_{W}R \; {}^{W}\vec{p} + {}^{C}_{W}\vec{t}$$

Non-homogeneous coordinates

$$\begin{pmatrix} \\ ^{C}\vec{p} \\ \\ \\ \end{pmatrix} = \begin{pmatrix} - & - & - & | \\ - & {}^{C}_{W}R & - & {}^{C}_{W}\vec{t} \\ - & - & - & | \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \\ ^{W}\vec{p} \\ \\ \\ \end{pmatrix}$$

Homogeneous coordinates

Combining extrinsic and intrinsic calibration parameters, in homogeneous coordinates

pixels

Intrinsic

$$\vec{p} = K \; {}^{C}\vec{p}$$

World coordinates

Camera coordinates

$$\begin{pmatrix} \\ {}^{C}\vec{p} \\ \\ \end{pmatrix} = \begin{pmatrix} - & - & - & | \\ - & {}^{C}_{W}R & - & {}^{C}_{W}\vec{t} \\ - & - & - & | \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \\ {}^{W}\vec{p} \\ \\ \end{pmatrix}$$

Extrinsic

$$\vec{p} = K \begin{pmatrix} {}^{C}_{W}R & {}^{C}_{W}\vec{t} \\ 0\,0\,0 & 1 \end{pmatrix} {}^{W}\vec{p}$$

$$\vec{p} = M \; {}^{W}\vec{p}$$

Forsyth&Ponce

Other ways to write the same equation

pixel coordinates

world coordinates

$$\vec{p} = M \; {}^{W}\vec{p}$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} . & m_1^T & . & . \\ . & m_2^T & . & . \\ . & m_3^T & . & . \end{pmatrix} \begin{pmatrix} {}^{W}p_x \\ {}^{W}p_y \\ {}^{W}p_z \\ 1 \end{pmatrix}$$

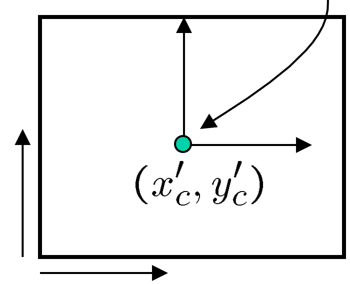
Conversion back from homogeneous coordinates leads to:

$$\begin{cases} u = \dfrac{m_1 \cdot \vec{P}}{m_3 \cdot \vec{P}} \\[2ex] v = \dfrac{m_2 \cdot \vec{P}}{m_3 \cdot \vec{P}} \end{cases}$$

Camera parameters

A camera is described by several parameters

- Translation T of the optical center from the origin of world coords
- Rotation R of the image plane
- focal length f, principle point (x'_c, y'_c), pixel size (s_x, s_y)
- blue parameters are called "extrinsics," red are "intrinsics"



Projection equation

$$\mathbf{X} = \begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{\Pi X}$$

- The projection matrix models the cumulative effect of all parameters
- Useful to decompose into a series of operations

identity matrix

$$\Pi = \begin{bmatrix} -fs_x & 0 & x'_c \\ 0 & -fs_y & y'_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{3x3} & \mathbf{0}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3x3} & \mathbf{T}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix}$$

intrinsics projection rotation translation

- The definitions of these parameters are not completely standardized
  - especially intrinsics—varies from one book to another