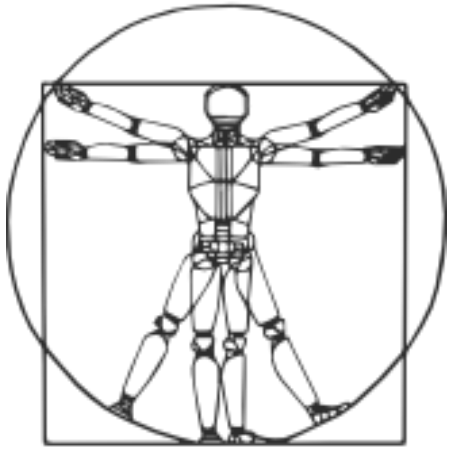
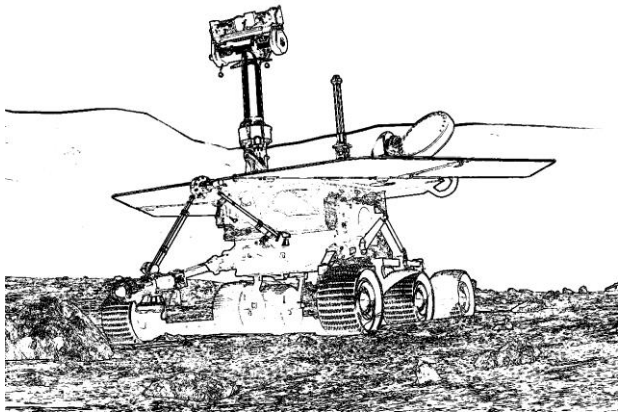




COMP417
Introduction to Robotics and Intelligent Systems

Lecture 20: Image Formation and Multi-View Geometry

Florian Shkurti

Computer Science Ph.D. student

florian@cim.mcgill.ca

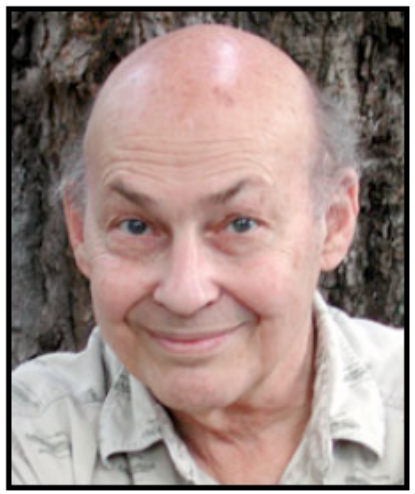
McGill

MRL Mobile Robotics Lab at McGill University

Motivation

- We have already seen quite successful SLAM methods based on laser sensors. Why bother with vision?
  - Camera technology cheap and ubiquitous
  - Camera is a passive sensor, lower energy
  - Some environments/platforms can’t support laser
  - Vision is quite a “rich” source of information

How hard is computer vision?



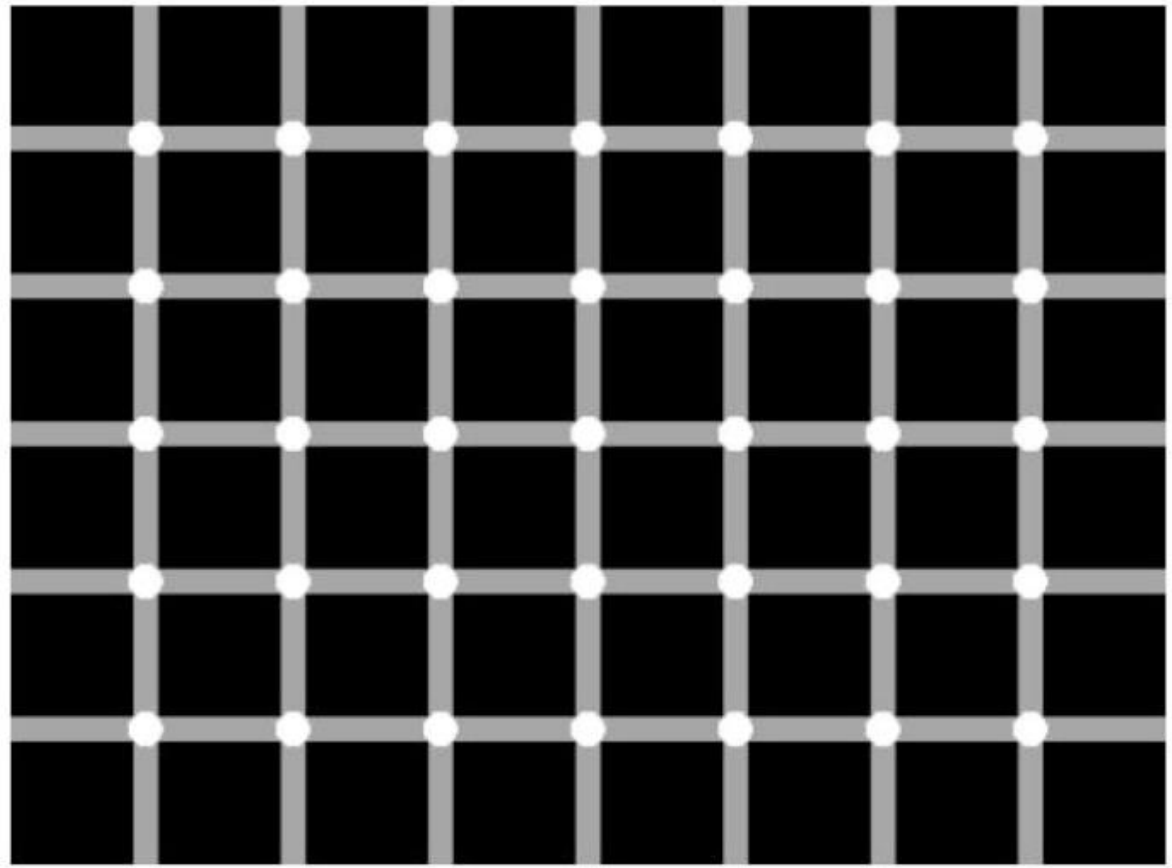
Marvin Minsky, MIT
Turing award,1969

"In 1966, Minsky hired a first-year undergraduate (JS) student and assigned him a problem to solve over the summer: connect a television camera to a computer and get the machine to describe what it sees."

Crevier 1993, pg. 88

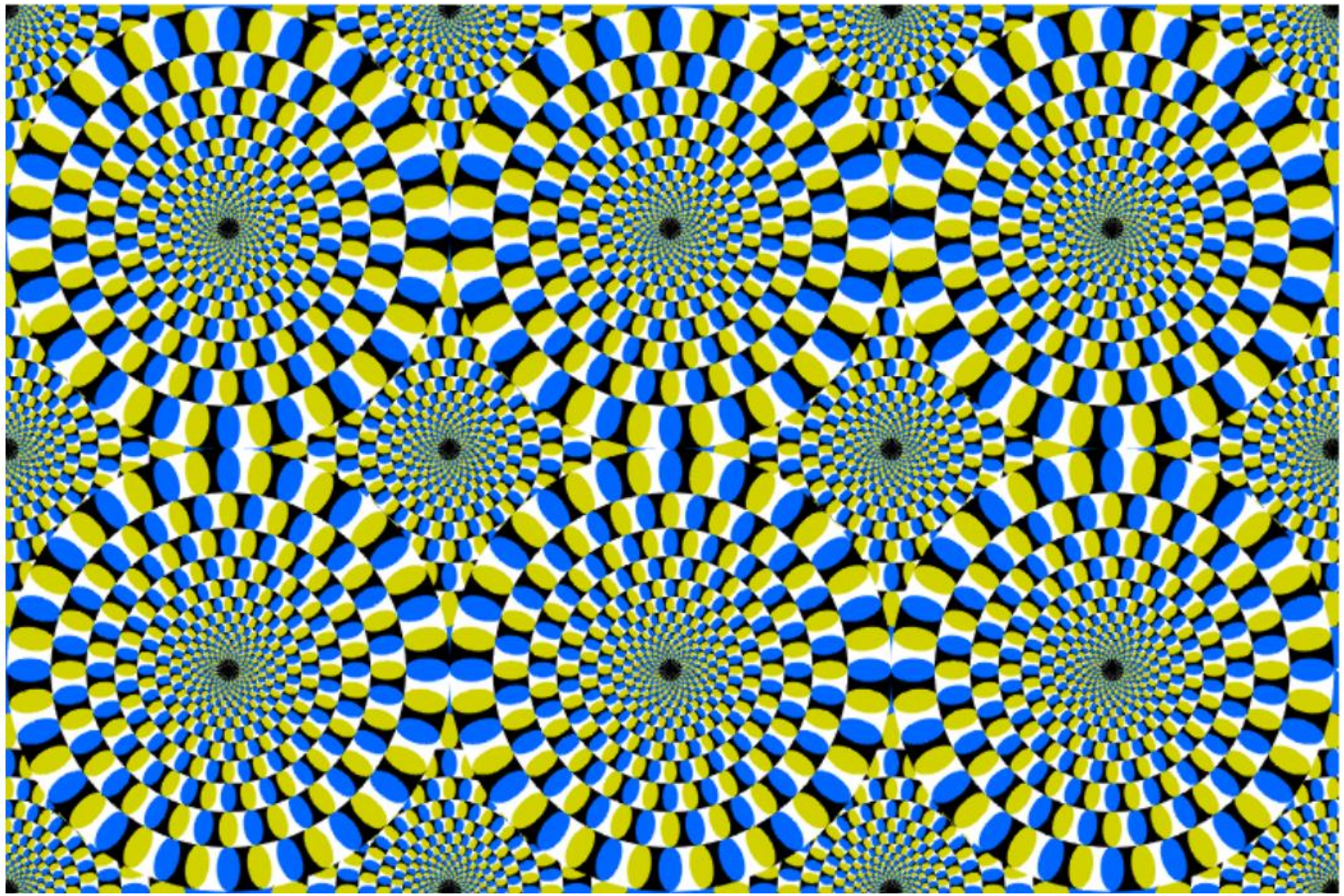
Depth perception can be ambiguous from just a single image

What do humans see?



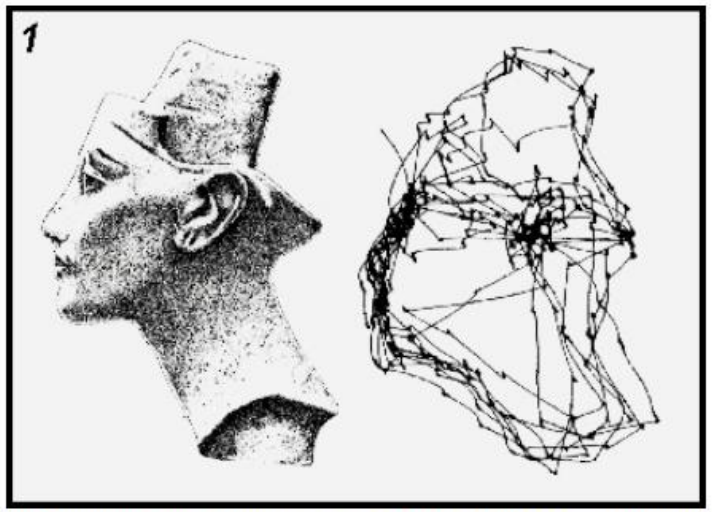
Count the black dots! :o)

Peripheral drift illusion

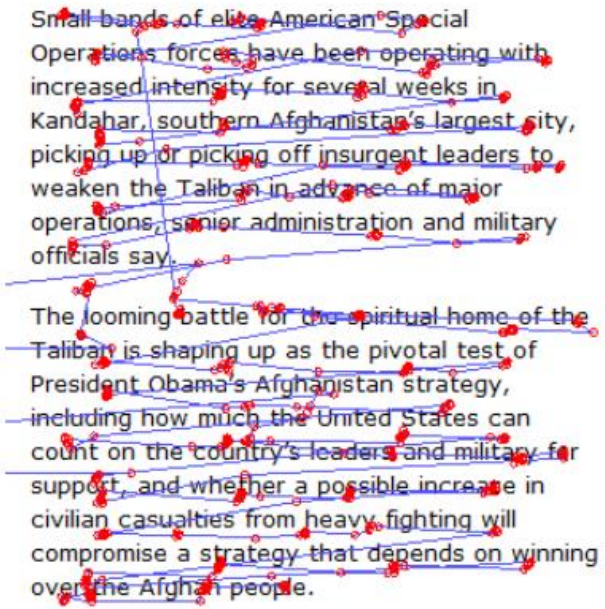


Where do humans fixate?

Top down or bottom up?



"Eye Movements and Vision" by A. L. Yarbus; Plenum Press, New York; 1967



Visual saccades

Camera obscura: dark room

- Known during classical period in China and Greece (e.g., Mo-Ti, China, 470BC to 390BC)

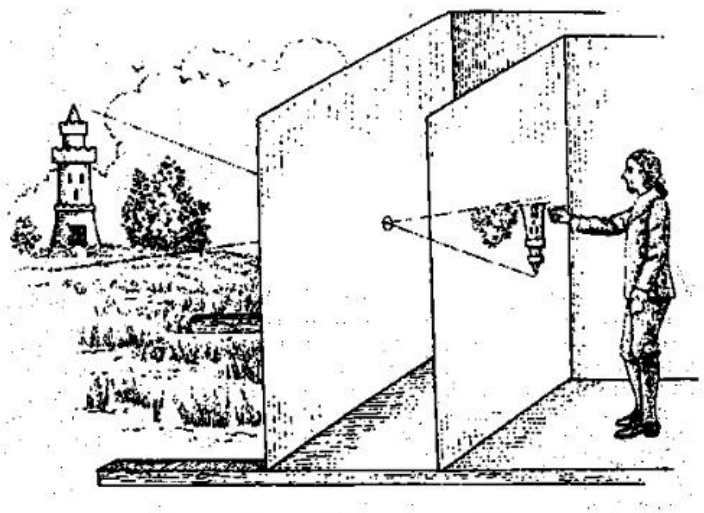
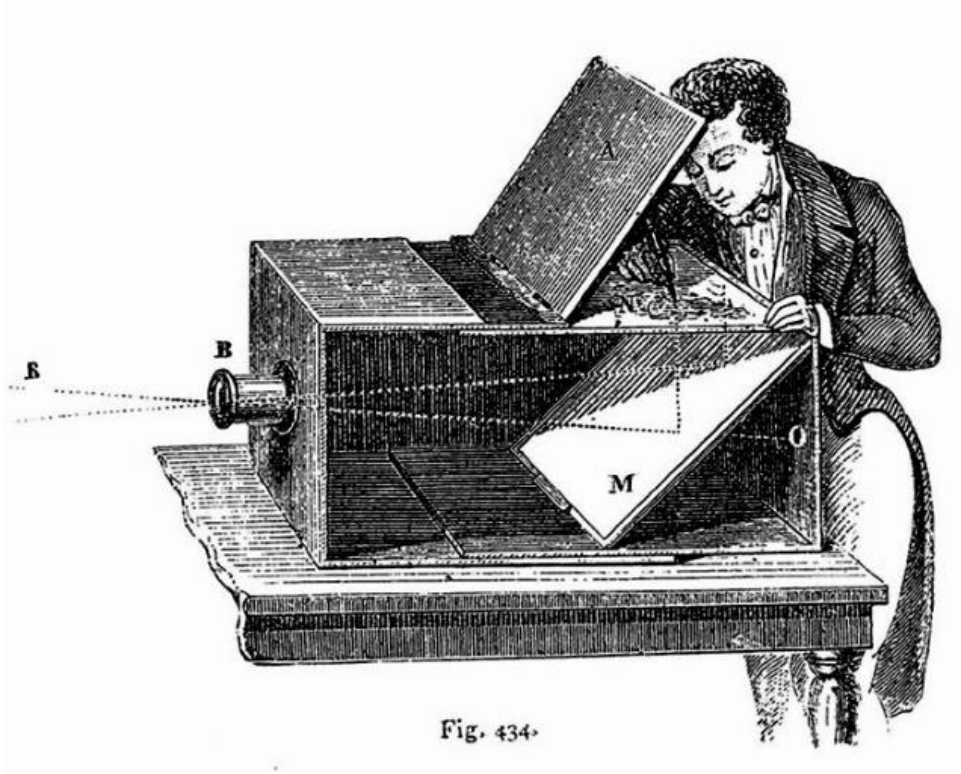


Illustration of Camera Obscura

Freestanding camera obscura at UNC Chapel Hill

Photo by Seth Ilys



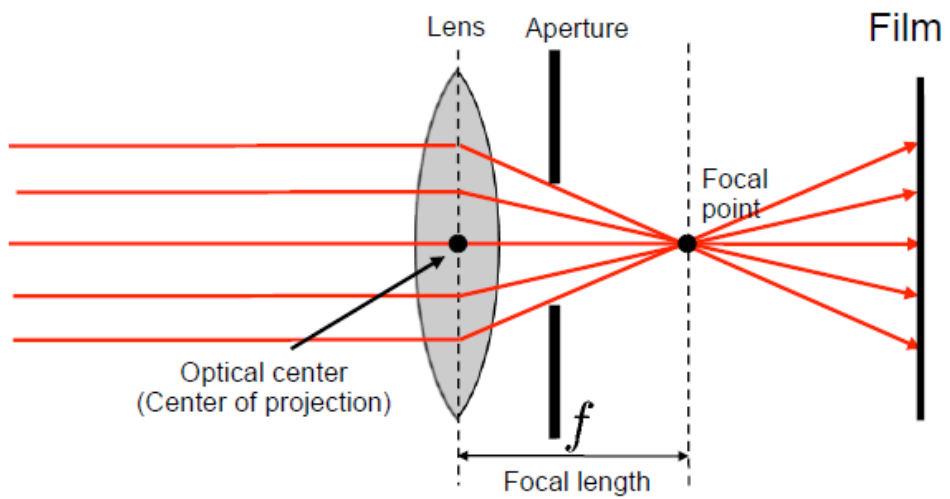
Lens Based Camera Obscura, 1568

Oldest surviving photograph

- Took 8 hours on pewter plate

Joseph Niepce, 1826

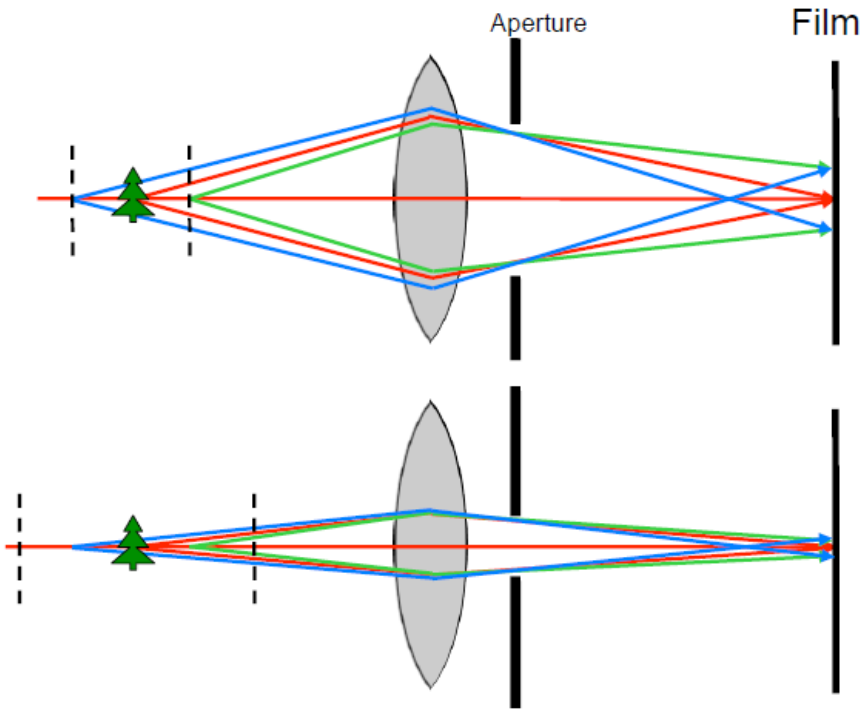
Lenses



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)

Depth of field



f / 5.6

f / 32

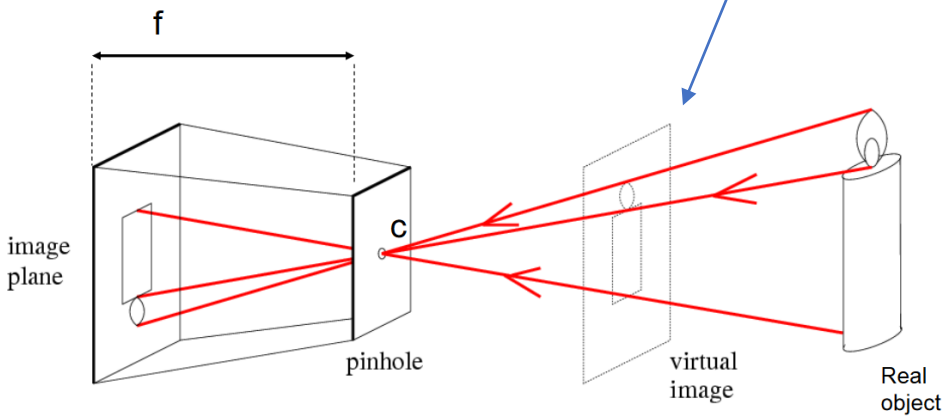
Changing the aperture size affects depth of field

- 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

To avoid thinking
about image inversion

Pinhole camera model

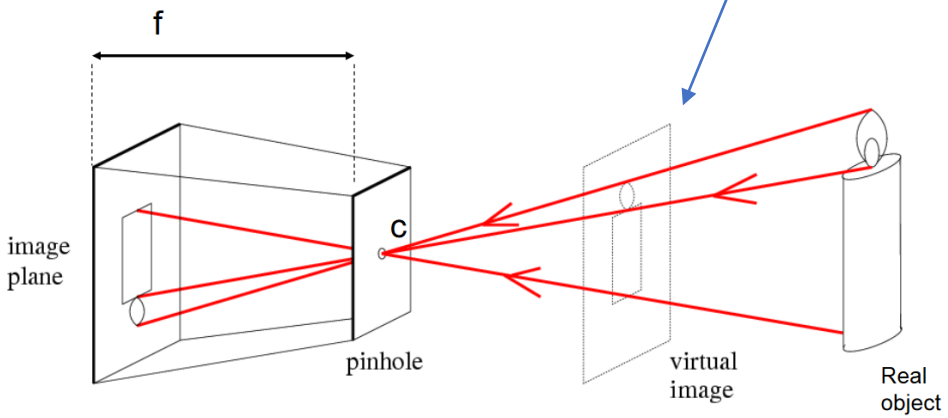


f = Focal length
c = Optical center of the camera

Point aperture → nearly every pixel in the image is in focus

To avoid thinking about image inversion

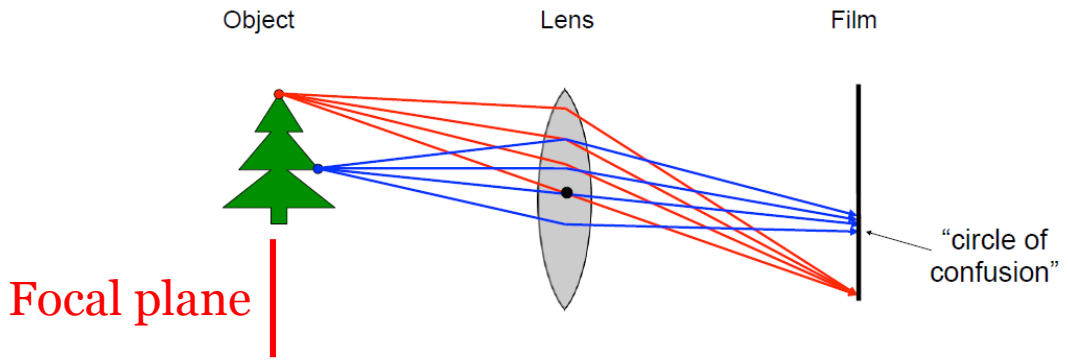
Pinhole camera model



f = Focal length
c = Optical center of the camera

Adding a lens

Some times called the thin-lens model



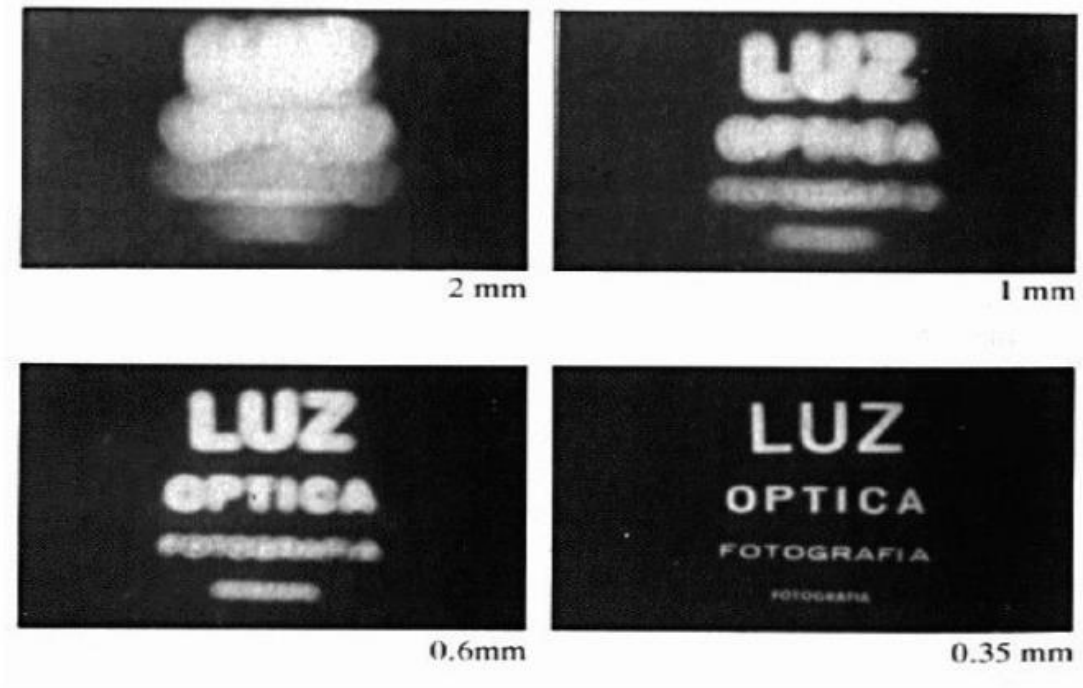
A lens focuses light onto the film

- 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

Point aperture → nearly every pixel in the image is in focus → almost infinite depth of field

Aperture of nonzero diameter → only pixels corresponding to objects on the focal plane are in focus → narrow depth of field

Shrinking the aperture

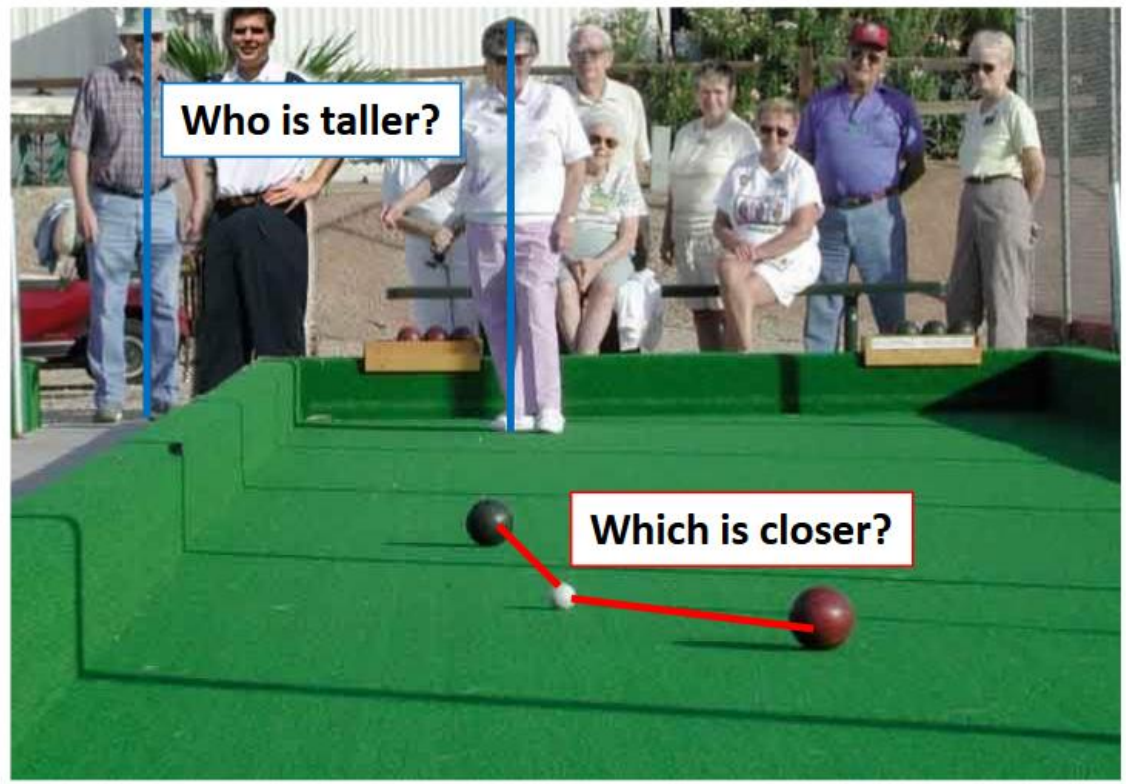


Why not make the aperture as small as possible?

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

Projective Geometry

Length (and so area) is lost.



Length and area are not preserved

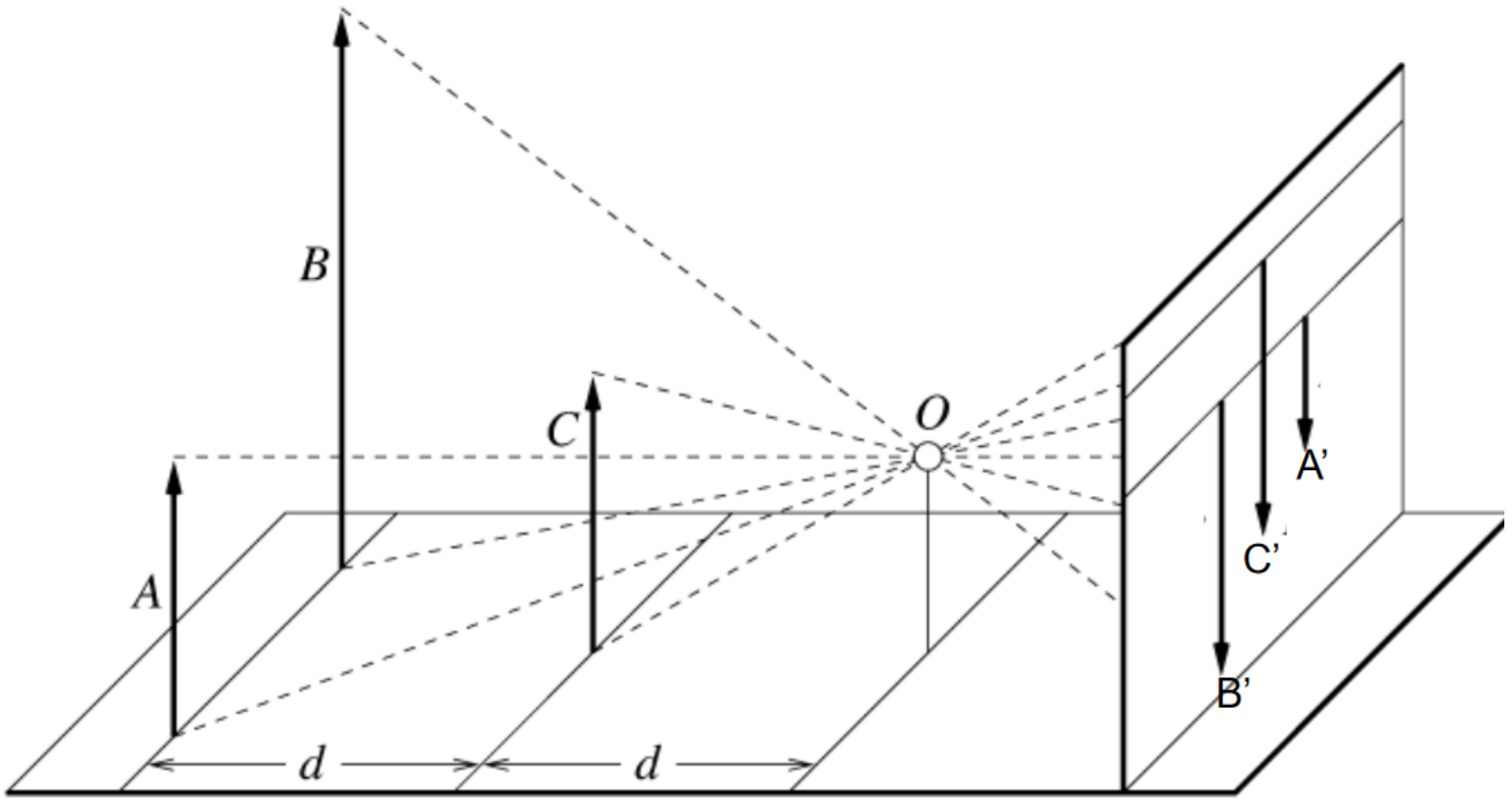
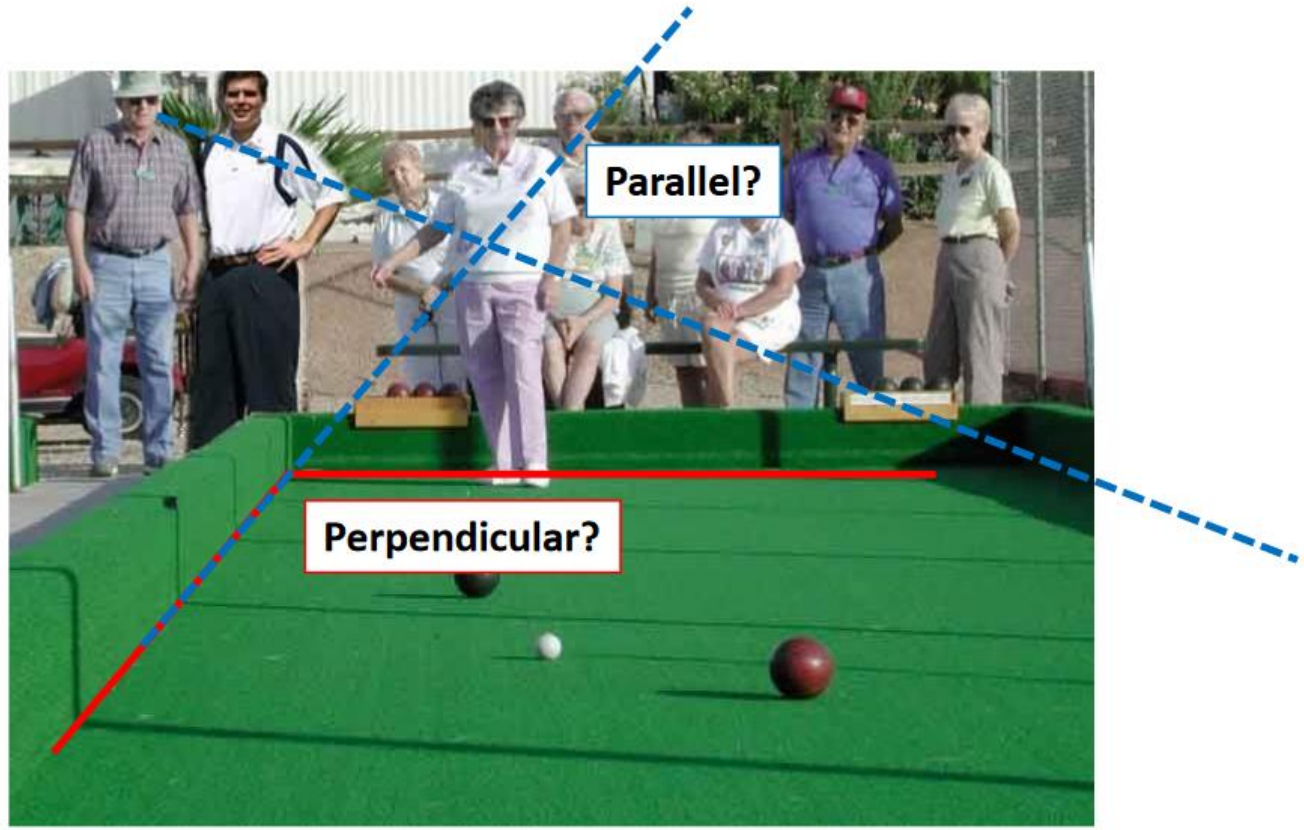


Figure by David Forsyth

Projective Geometry

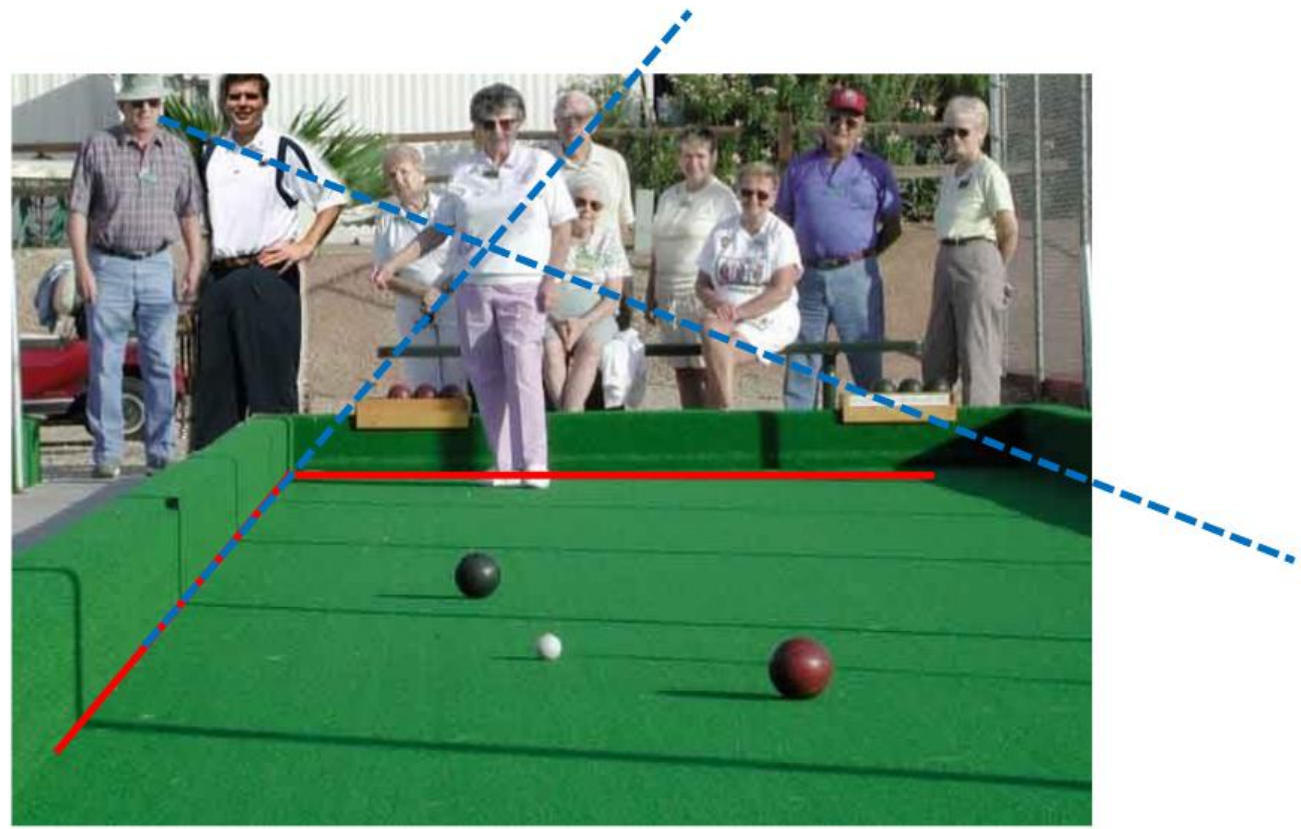
Angles are lost.



Projective Geometry

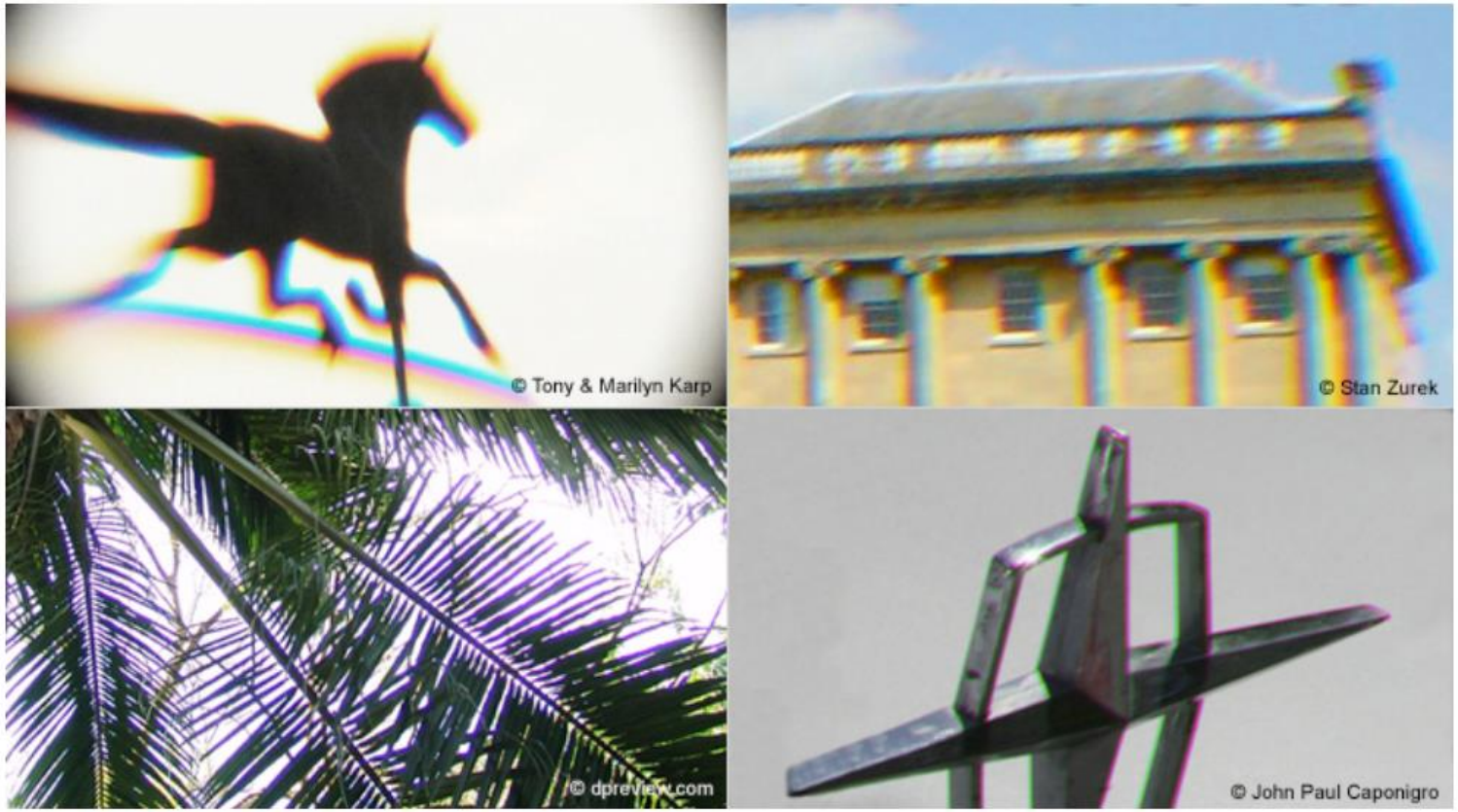
What is preserved?

- Straight lines are still straight.

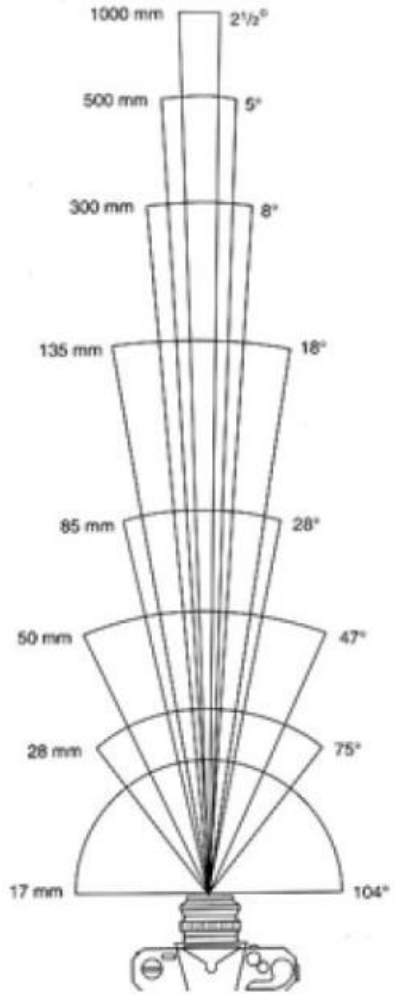


Chromatic aberration

Failure of a lens to focus all colors to the same convergence point.
Due to difference wavelengths having different refractive indeces



Field of View (Zoom, focal length)



17mm

28mm

50mm

85mm

From London and Upton

Camera parameters

Focus – Shifts the depth that is in focus.

Focal length – Adjusts the zoom, i.e., wide angle or telephoto lens.

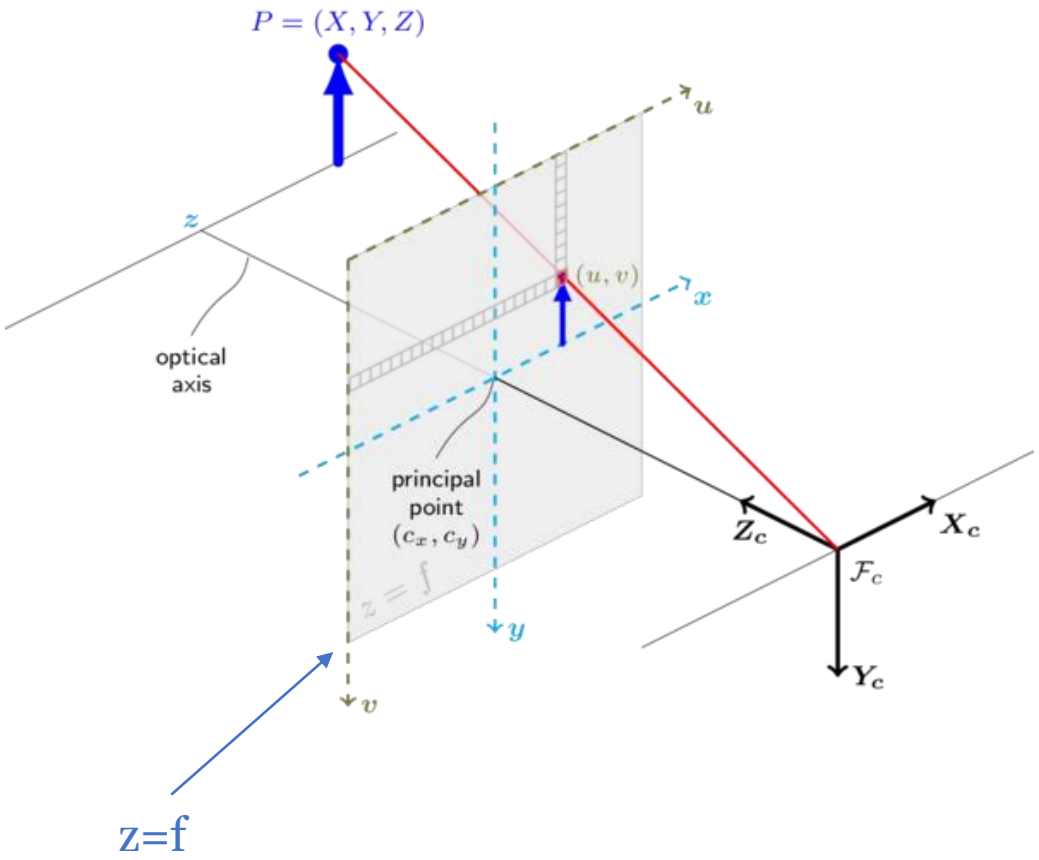
Aperture – Adjusts the depth of field and amount of light let into the sensor.

Exposure time – How long an image is exposed. The longer an image is exposed the more light, but could result in motion blur.

ISO – Adjusts the sensitivity of the “film”. Basically a gain function for digital cameras. Increasing ISO also increases noise.

How do we project 3D points to pixels?
What is the measurement model?

From 3D points to pixels: pinhole camera



(1) Perspective projection $\begin{bmatrix} x \\ y \end{bmatrix} = \pi(X, Y, Z)$

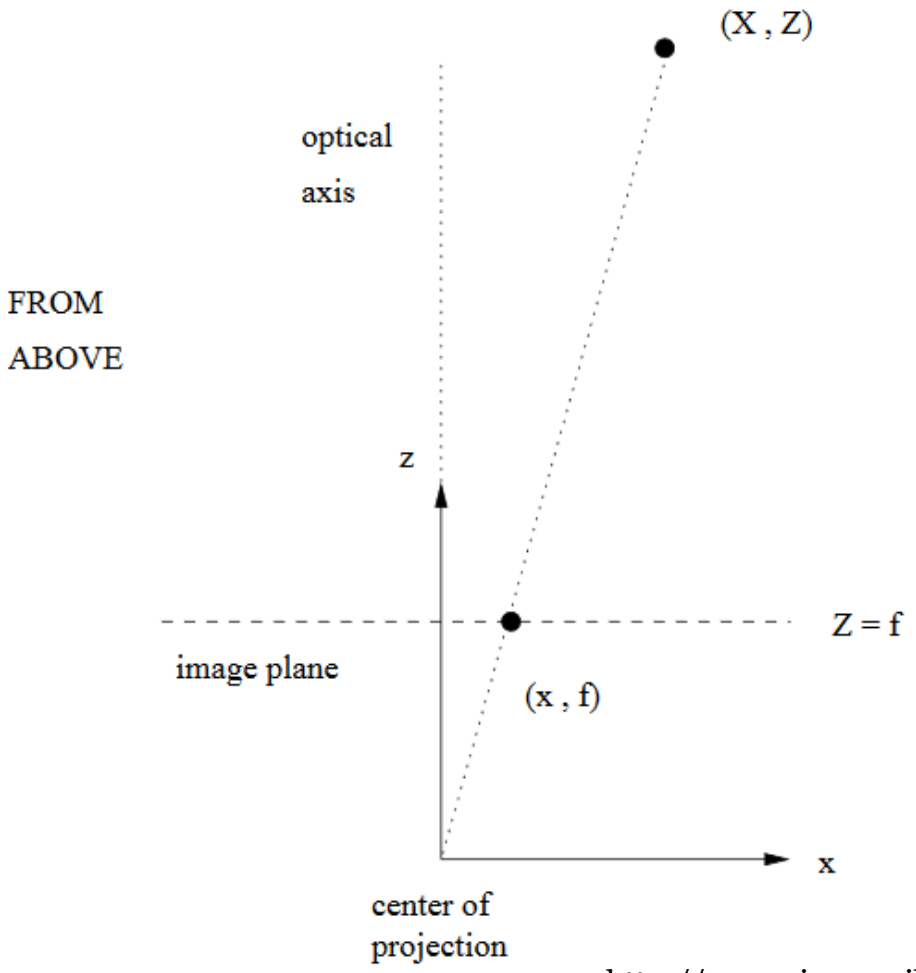
(2) Conversion from metric to pixel coordinates

$$u = m_x x + c_x$$

$$v = m_y y + c_y$$

m_x, m_y represent number of pixels per mm for the two axes

Perspective projection
$[x,y] = \pi(X,Y,Z)$



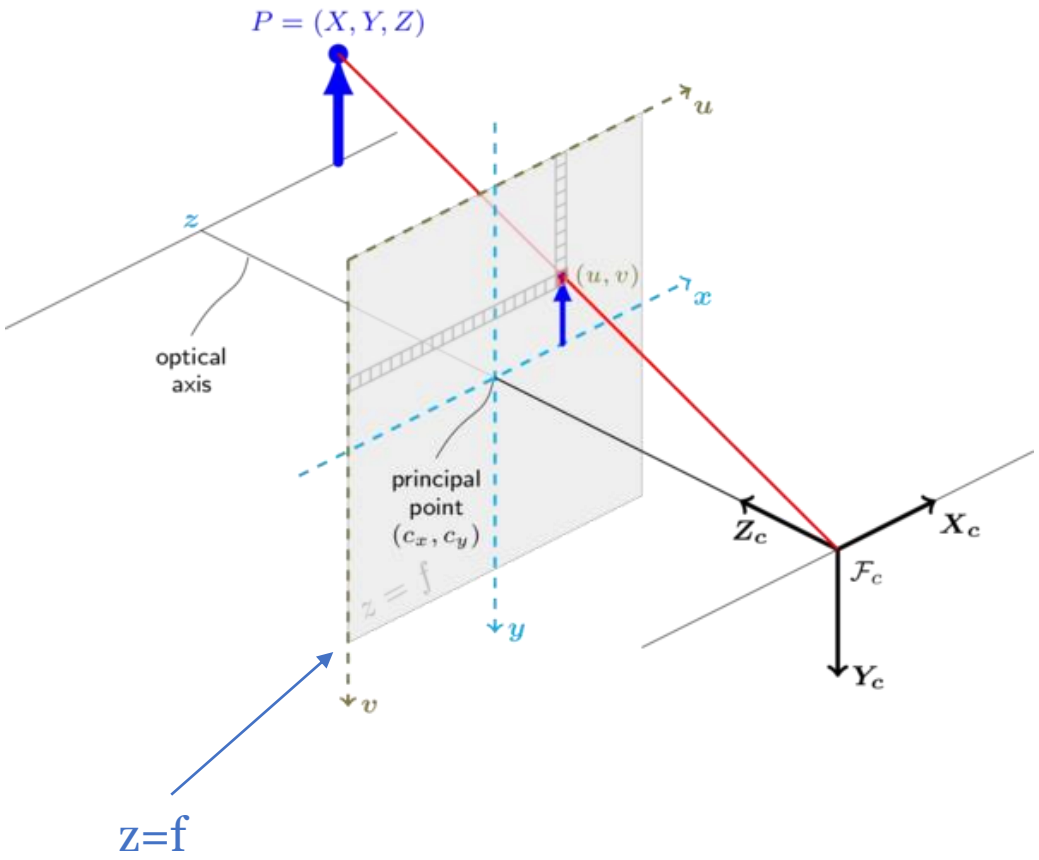
By similar triangles: $x/f = X/Z$

So, $x = f * X/Z$ and similarly $y = f * Y/Z$

Problem: we just lost depth (Z) information by doing this projection, i.e. depth is now uncertain.

http://www.cim.mcgill.ca/%7Elanger/558.html

From 3D points to pixels: pinhole camera



(1) Perspective projection $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} fX/Z \\ fY/Z \end{bmatrix} = \pi(X, Y, Z)$

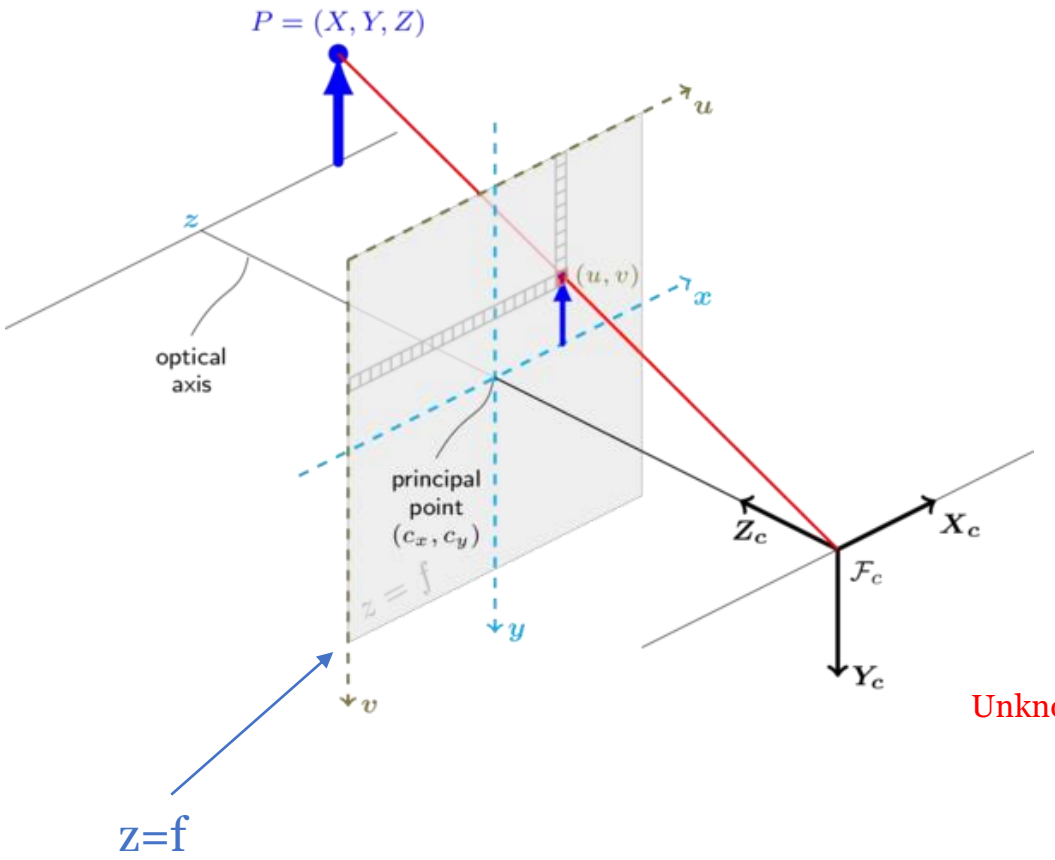
(2) Conversion from metric to pixel coordinates

$$u = m_x x + c_x$$

$$v = m_y y + c_y$$

$$h_{\text{pinhole}}(X, Y, Z) = \begin{bmatrix} \frac{f m_x X}{Z} + c_x \\ \frac{f m_y Y}{Z} + c_y \end{bmatrix} + \text{noise in pixels}$$

From 3D points to pixels: pinhole camera



(1) Perspective projection $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} fX/Z \\ fY/Z \end{bmatrix} = \pi(X, Y, Z)$

(2) Conversion from metric to pixel coordinates

$$u = m_x x + c_x$$

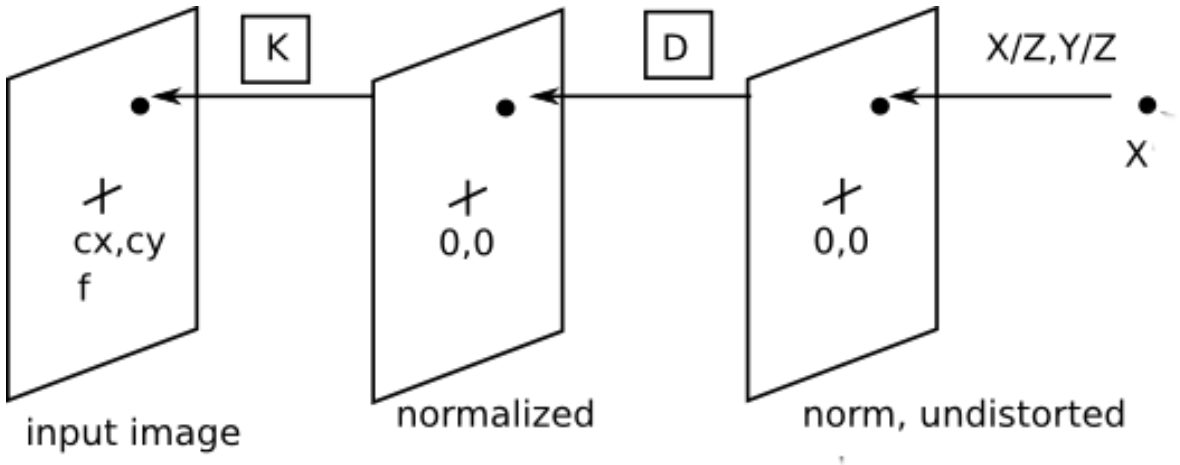
$$v = m_y y + c_y$$

Unknown depth/scale

Camera calibration matrix

Usually presented as $s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} fm_x & 0 & c_x \\ 0 & fm_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$

From 3D points to pixels: thin lens camera



(1) Perspective projection $\begin{bmatrix} x \\ y \end{bmatrix} = \pi(X, Y, Z)$

(2) Lens distortion

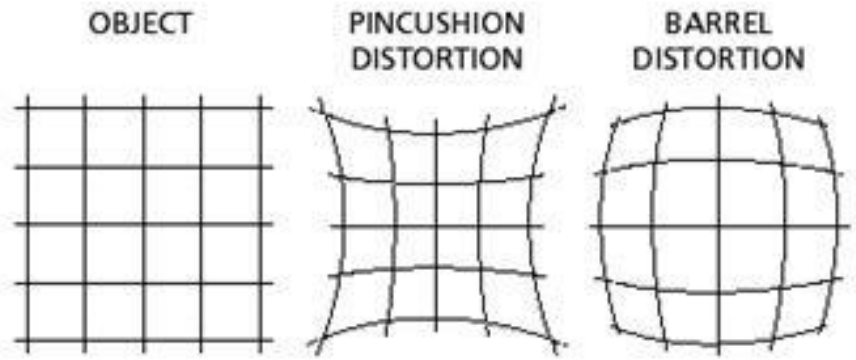
$$[x^*, y^*] = D(x, y)$$

(3) Conversion from metric to pixel coordinates

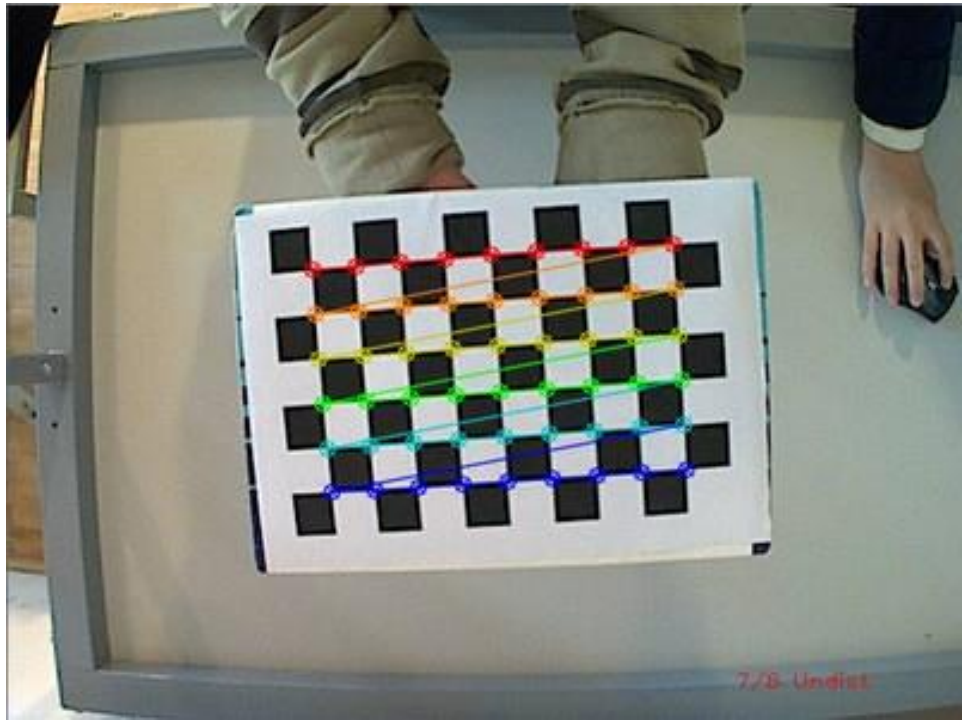
$$u = m_x x^* + c_x$$
$$v = m_y y^* + c_y$$

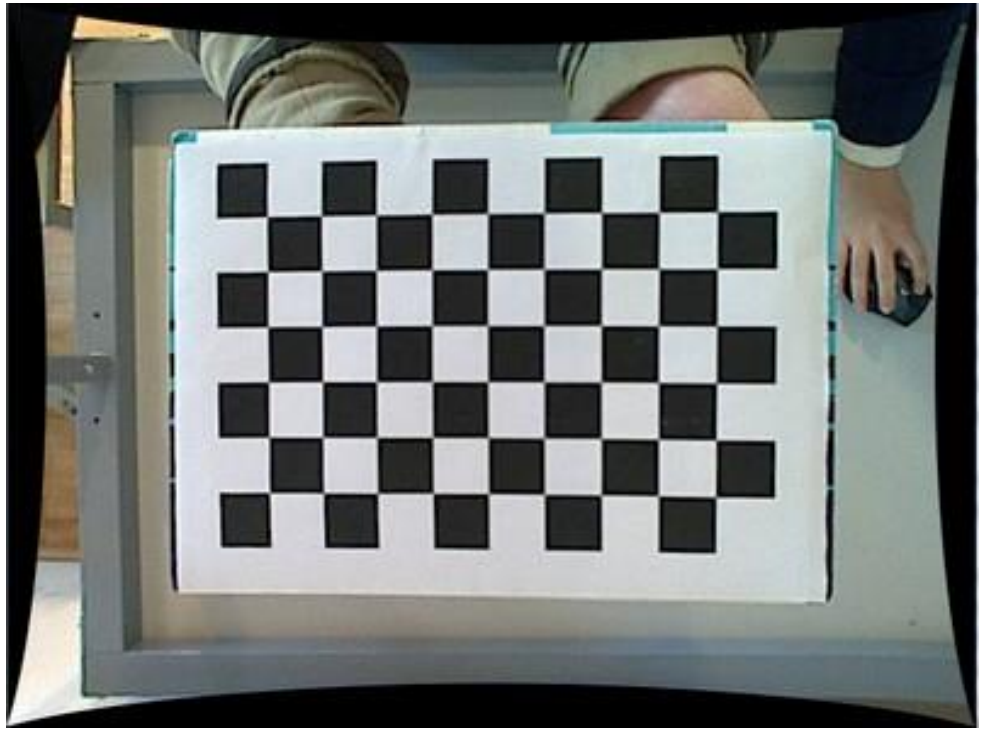
(2) Lens distortion

$$[x^*, y^*] = D(x,y)$$



(2) Estimating parameters of lens distortion: [x*, y*] = D(x,y)





$$x^* = x\frac{1 + k_1r^2 + k_2r^4 + k_3r^6}{1 + k_4r^2 + k_5r^4 + k_6r^6} \quad \text{where} \quad r = x^2 + y^2$$

$$y^* = y\frac{1 + k_1r^2 + k_2r^4 + k_3r^6}{1 + k_4r^2 + k_5r^4 + k_6r^6} \quad \text{where} \quad r = x^2 + y^2$$

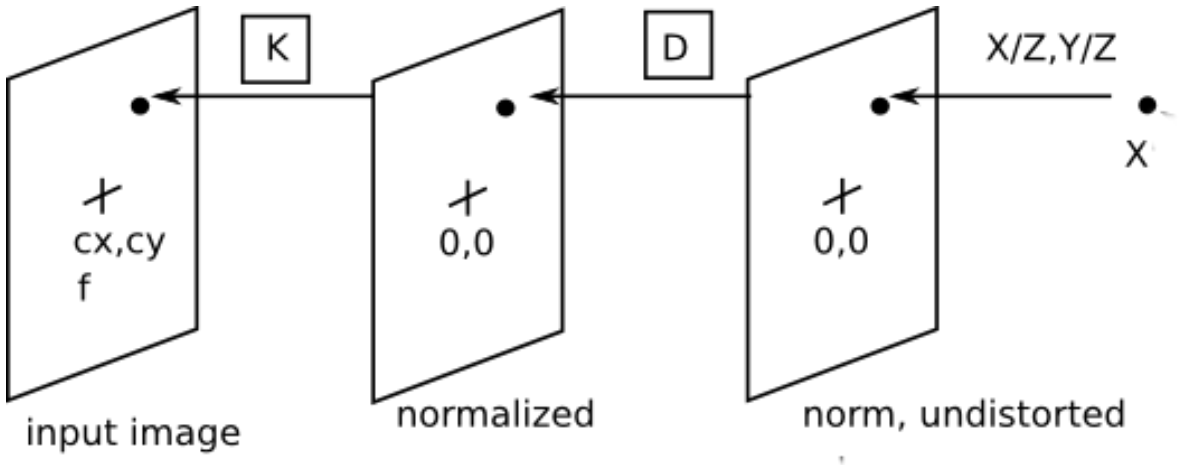
Correcting radial distortion

Barrel distortion

Corrected

from Helmut Dersch

From 3D points to pixels: thin lens camera



(1) Perspective projection $\begin{bmatrix} x \\ y \end{bmatrix} = \pi(X, Y, Z)$

(2) Lens distortion

$$[x^*, y^*] = D(x, y)$$

(3) Conversion from metric to pixel coordinates

$$u = m_x x^* + c_x$$

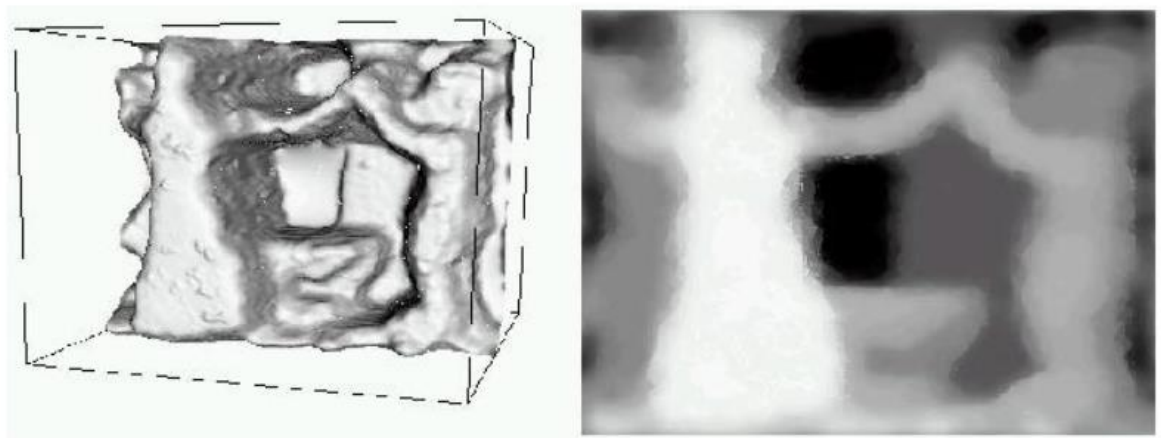
$$v = m_y y^* + c_y$$

If we have access to camera calibration parameters
we can undo the lens distortion, and treat the measurement
model as in the pinhole camera → single-camera image rectification

What visual or physiological cues help us to perceive 3D shape and depth?

Focus/defocus

Images from same point of view, different camera parameters



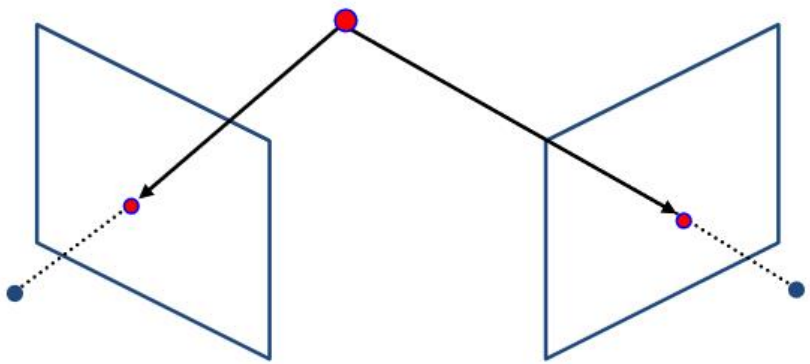
3d shape / depth estimates

[figs from H. Jin and P. Favaro, 2002]

Perspective effects

Image credit: S. Seitz

Stereo



Slides: James Hays and Kristen Grauman

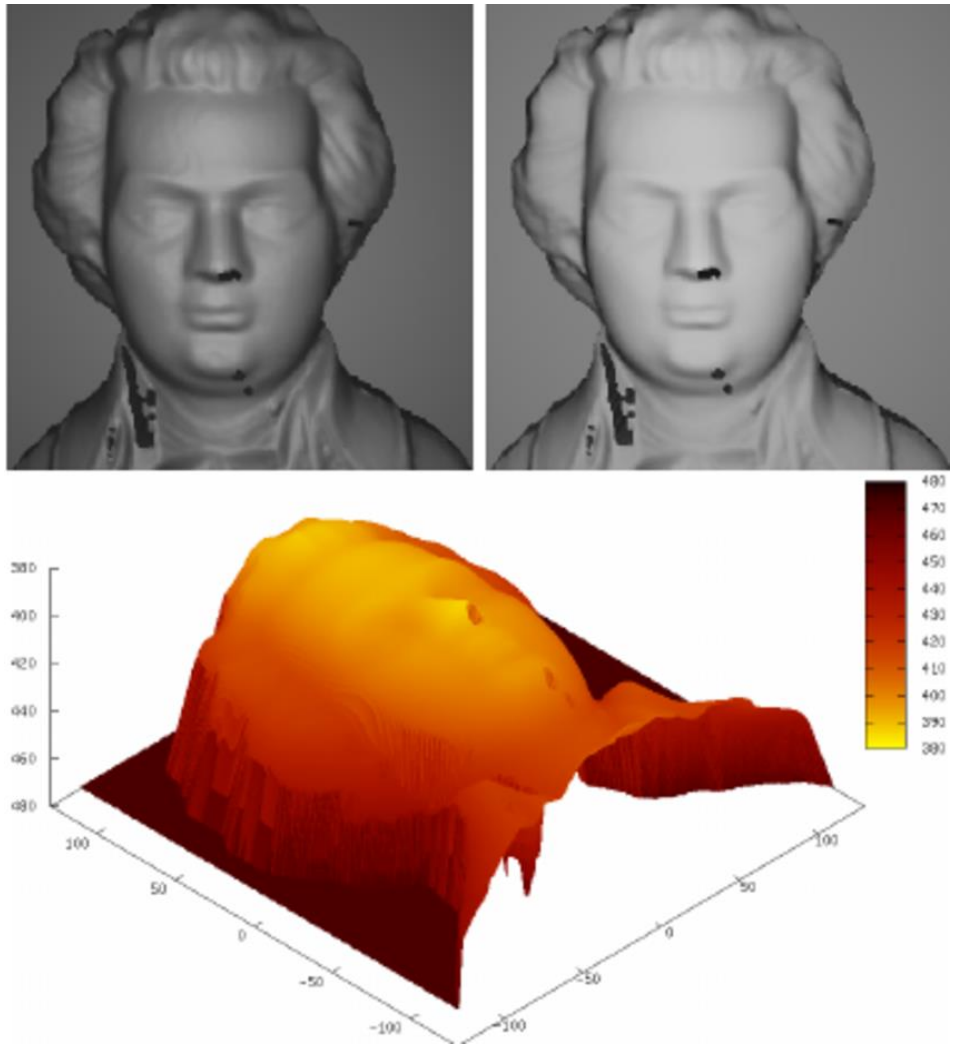
Why multiple views?

Structure and depth can be ambiguous from single views...

Images from Lana Lazebnik

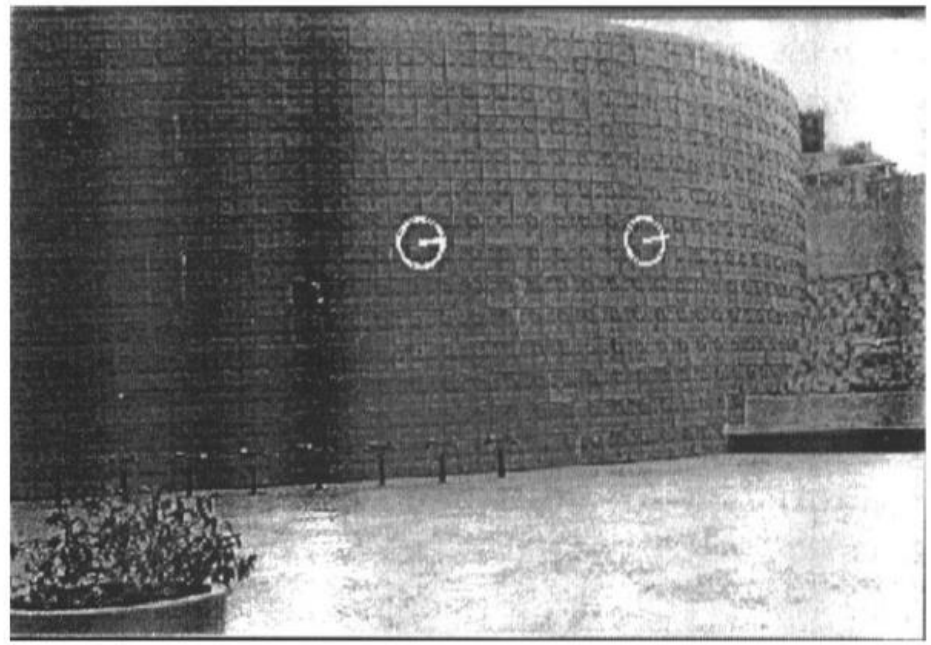
If stereo were critical for depth perception, navigation, recognition, etc., then rabbits would never have evolved.

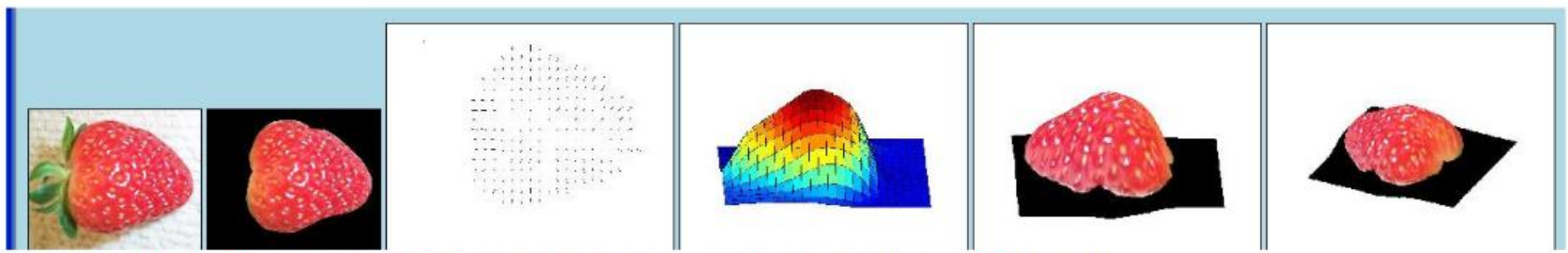
Shape from shading



"Numerical schemes for advanced reflectance models for Shape from Shading", Vogel, Cristiani

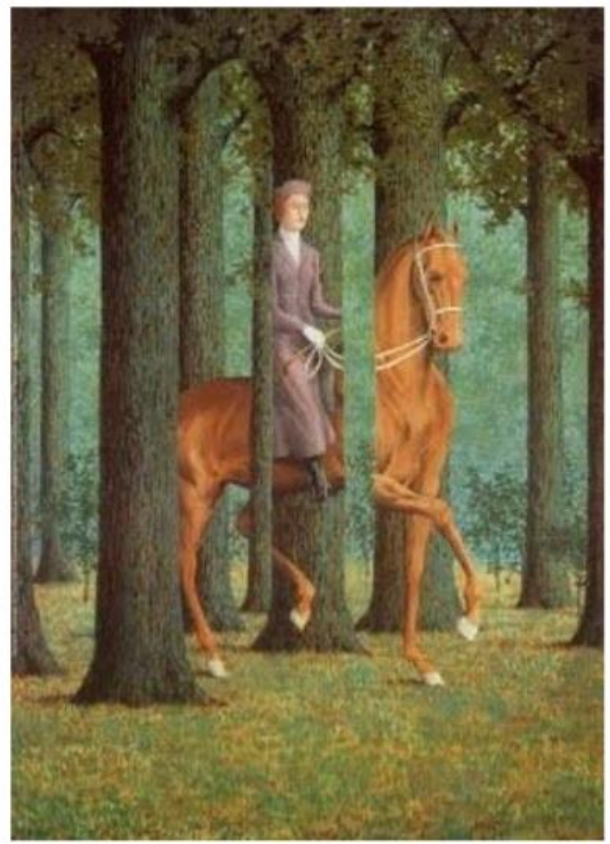
Texture





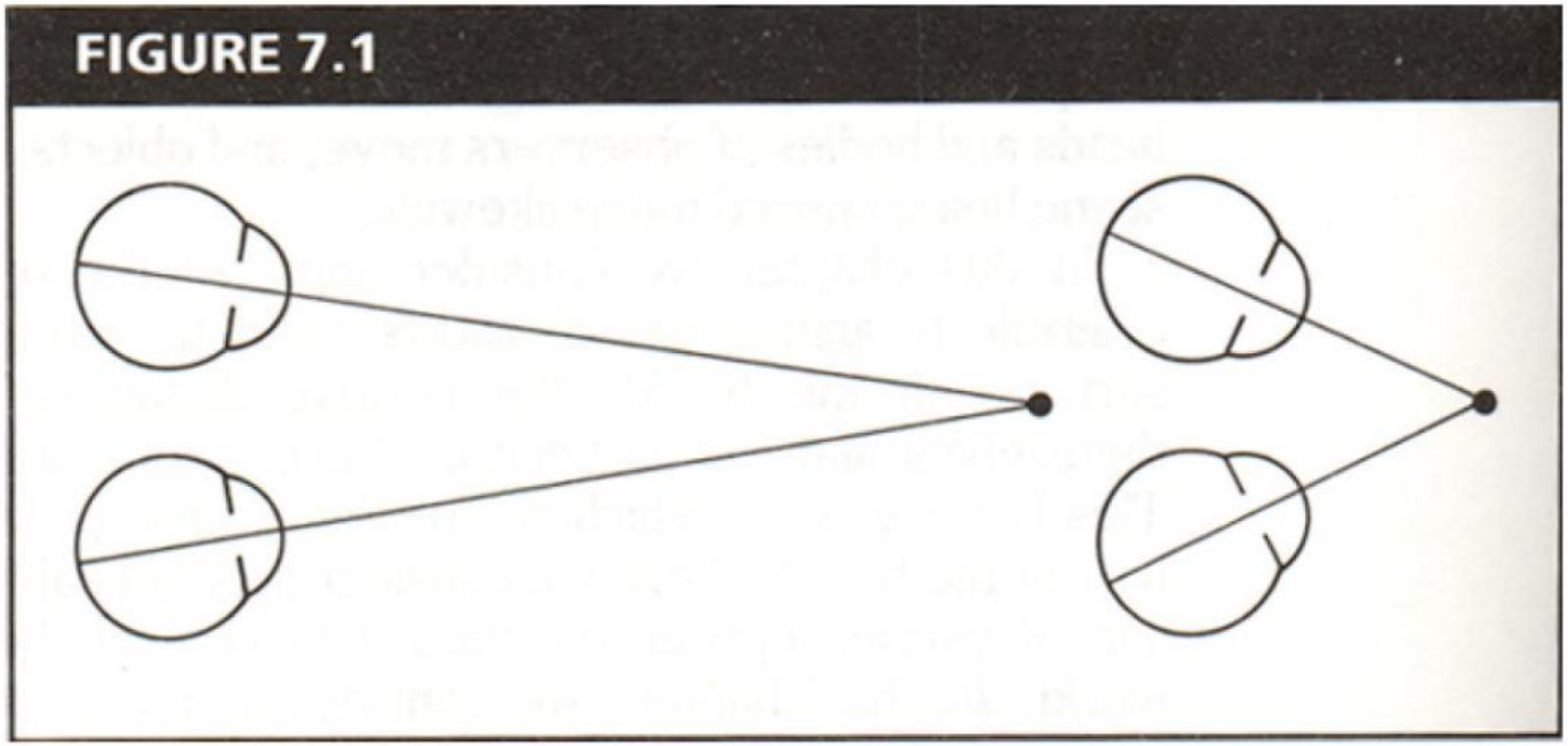
[From A.M. Loh. The recovery of 3-D structure using visual texture patterns. PhD thesis]

Occlusion



Rene Magritt'e famous painting *Le Blanc-Seing* (literal translation: "The Blank Signature") roughly translates as "free hand" or "free rein".

Human stereopsis



From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Human eyes **fixate** on point in space – rotate so that corresponding images form in centers of fovea.

Structure from Motion

"SFMedu: A Structure from Motion System for Education", Jianxiong Xiao

Many depth from X methods. We are going to focus on structure from motion and stereo → part of multiple-view geometry