CS 558: Computer Vision
1st Set of Notes

Instructor: Philippos Mordohai
Webpage: www.cs.stevens.edu/~mordohai
E-mail: Philippos.Mordohai@stevens.edu
Office: Lieb 215

Important Points

- This is an elective course. You chose to be here.
- Expect to work and to be challenged.
- Exams won't be based on recall. They will be open book and you will be expected to solve new problems.

Logistics

- Office hours: Tuesday 5-6 and by email
- Evaluation:
  - 4 homework sets (40%)
  - At least 10 quizzes and participation (25%)
  - Final exam (35%)

Textbook

- Richard Szeliski, Computer Vision: Algorithms and Applications, Springer, 2010
- Available online at http://szeliski.org/Book/

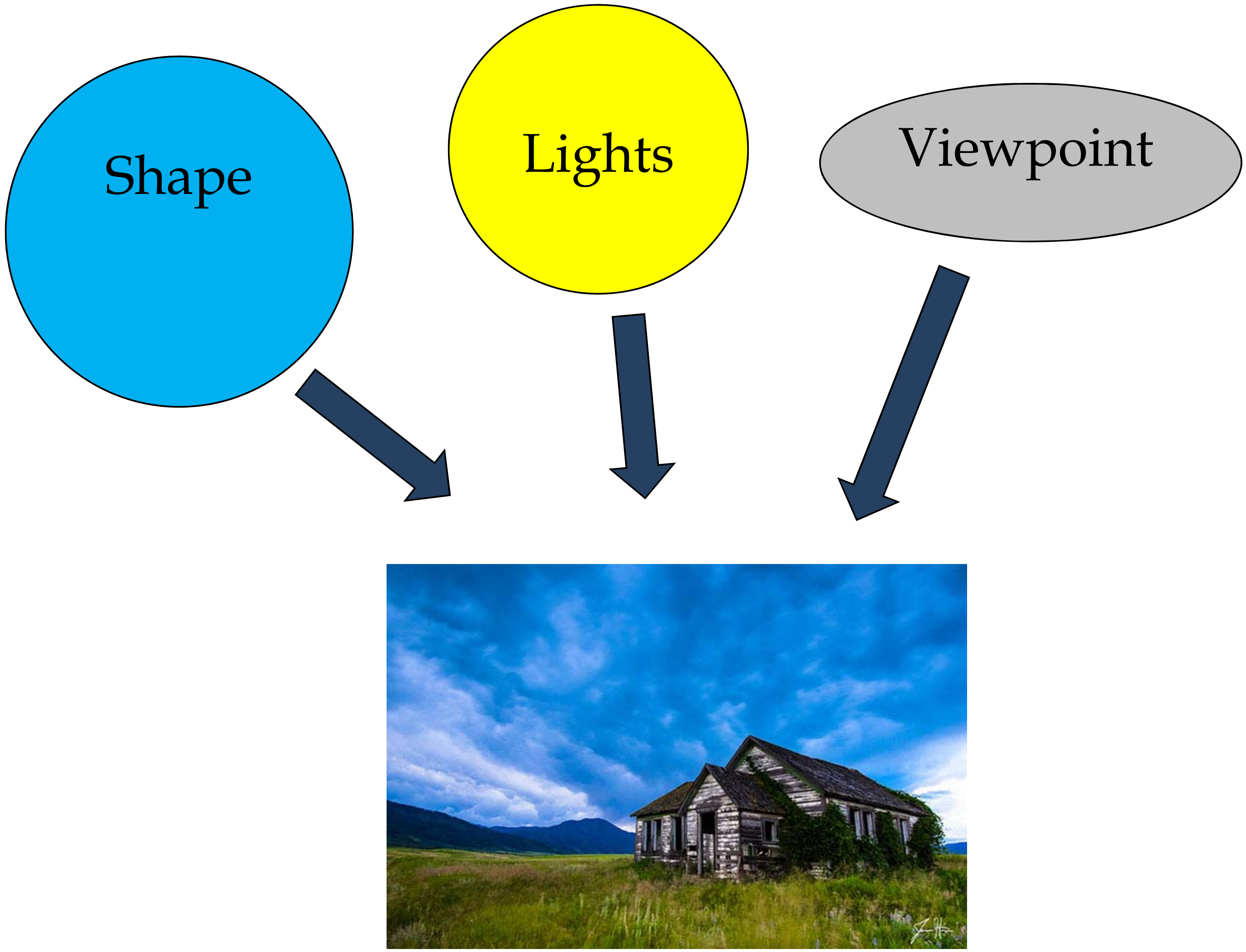
Tentative Schedule

Week Starting	**Topic(s)**	**Readings**	**Assignment**
January 20	Introduction and cameras	Szeliski Ch. 2.3	
January 27	Image formation, convolution and filtering	Szeliski Ch. 2.2, 2.3 and 3.2	
February 3	Edge, corner and feature detection	Szeliski Ch. 4.1 and 4.2	Homework 1, due 2/10
February 10	Fitting, alignment and tracking	Szeliski Ch.4.3 and 6.1	
February 17	Template matching, image pyramids and optical flow	Szeliski Ch.4.1, 8.1 and 8.4	
February 24	Grouping and segmentation	Szeliski Ch.5.2 and 5.3	Homework 2, due 3/2
March 2	Camera geometry and Structure-from-Motion	Szeliski Ch.2.1 and 7	
March 9	3D reconstruction	Szeliski Ch. 11	
March 16	Object recognition (I)	Szeliski Ch. 14 and notes	Homework 3, due 3/30
March 30	Object recognition (II)	Szeliski Ch. 14 and notes	
April 6	Object recognition (III)	Szeliski Ch. 14 and notes	
April 13	Deep learning for computer vision	Notes	Homework 4, due 4/20
April 20	Context	Szeliski Ch. 14.5 and notes	
April 27	Action and activity recognition	Notes	
May	**Final Exam**		

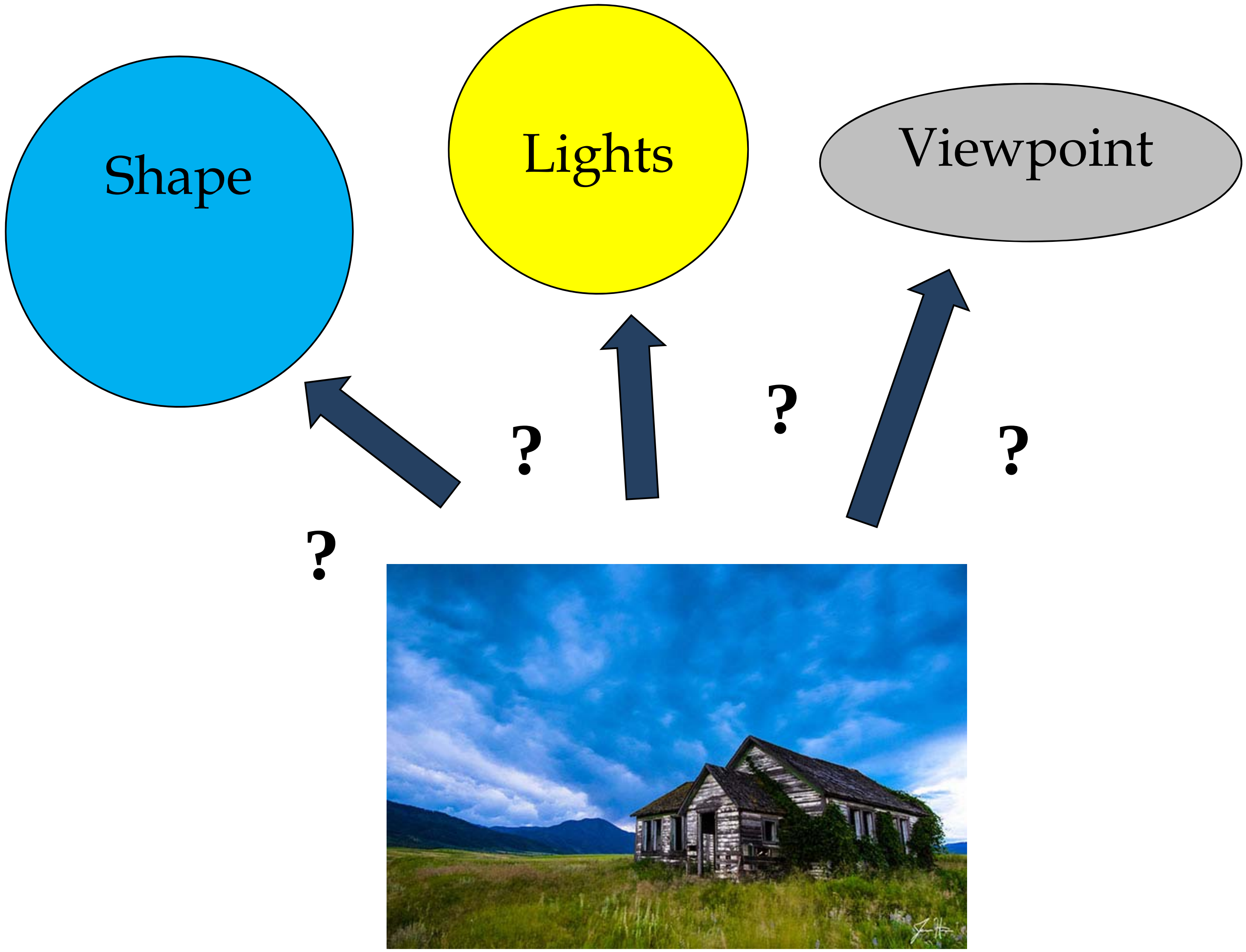
What is Computer Vision

- …
- How is it different from image processing?

Graphics vs. Vision



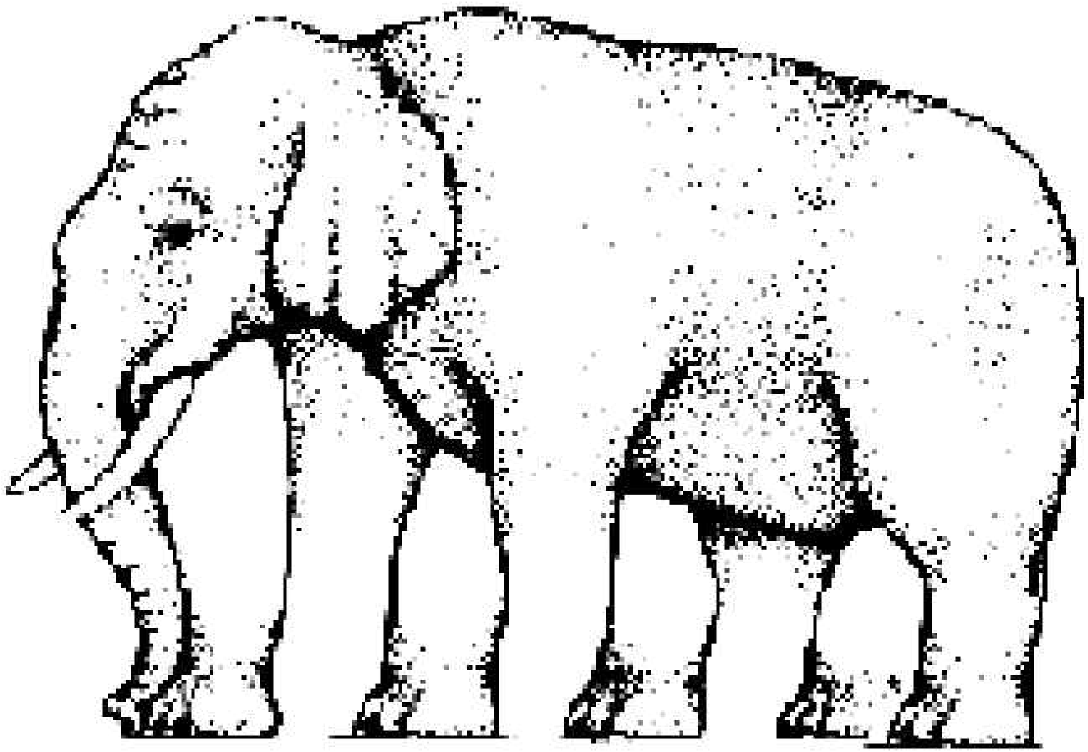
Graphics vs. Vision



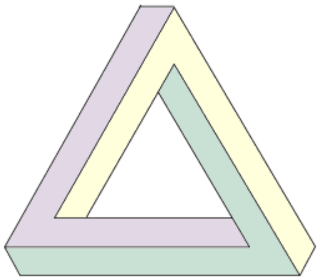
Vision is Hard

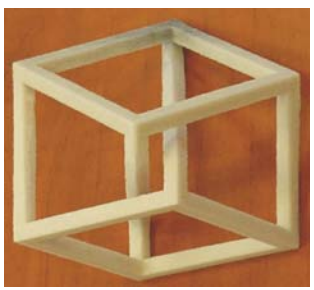
Vision is Hard

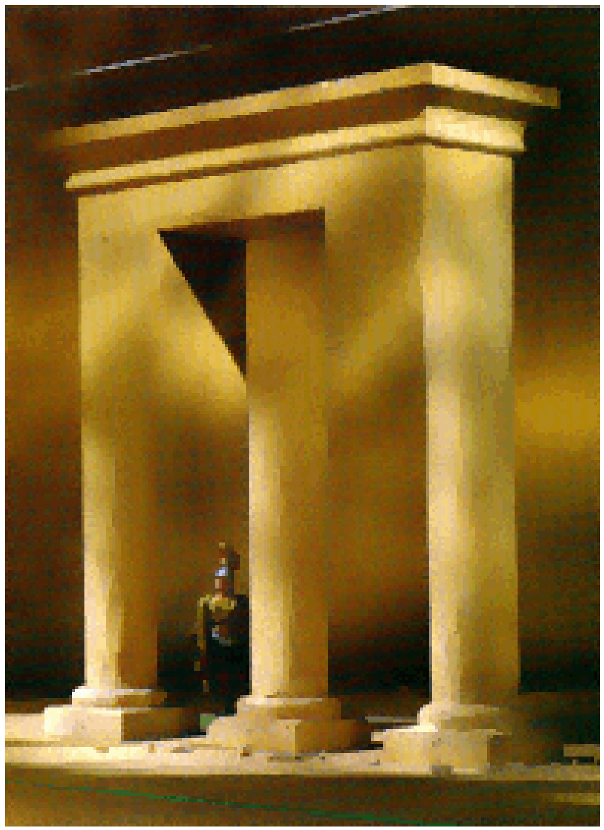
Vision is Hard

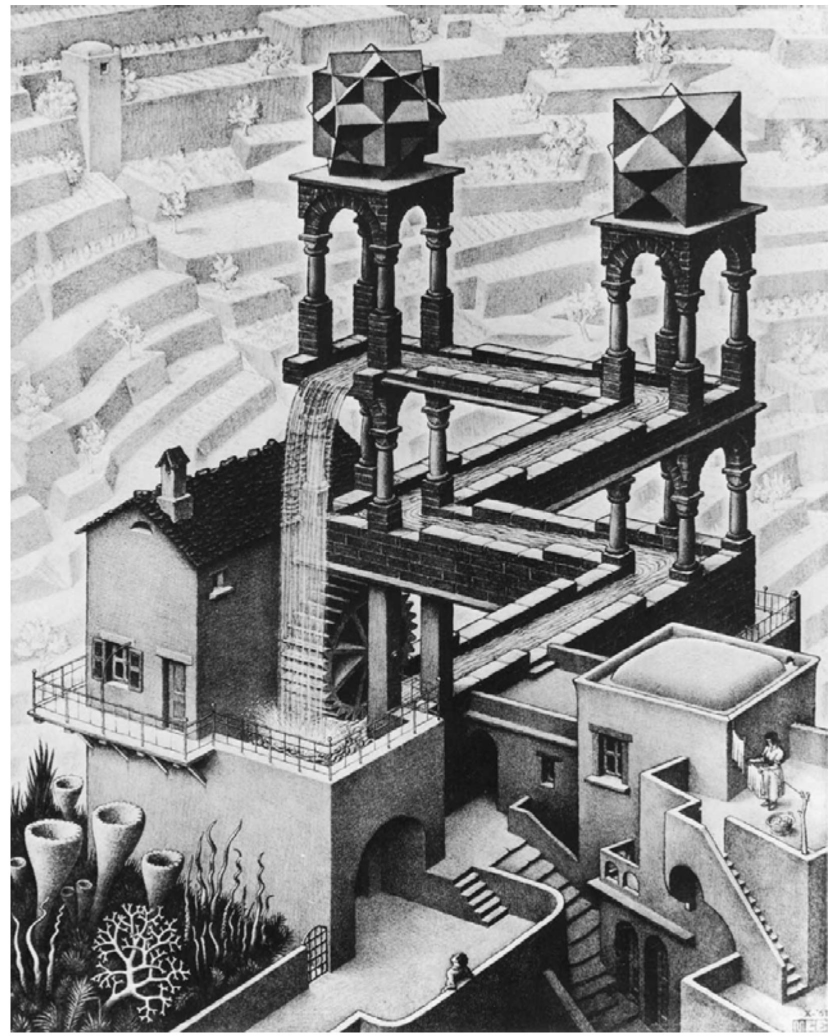


Vision is Hard

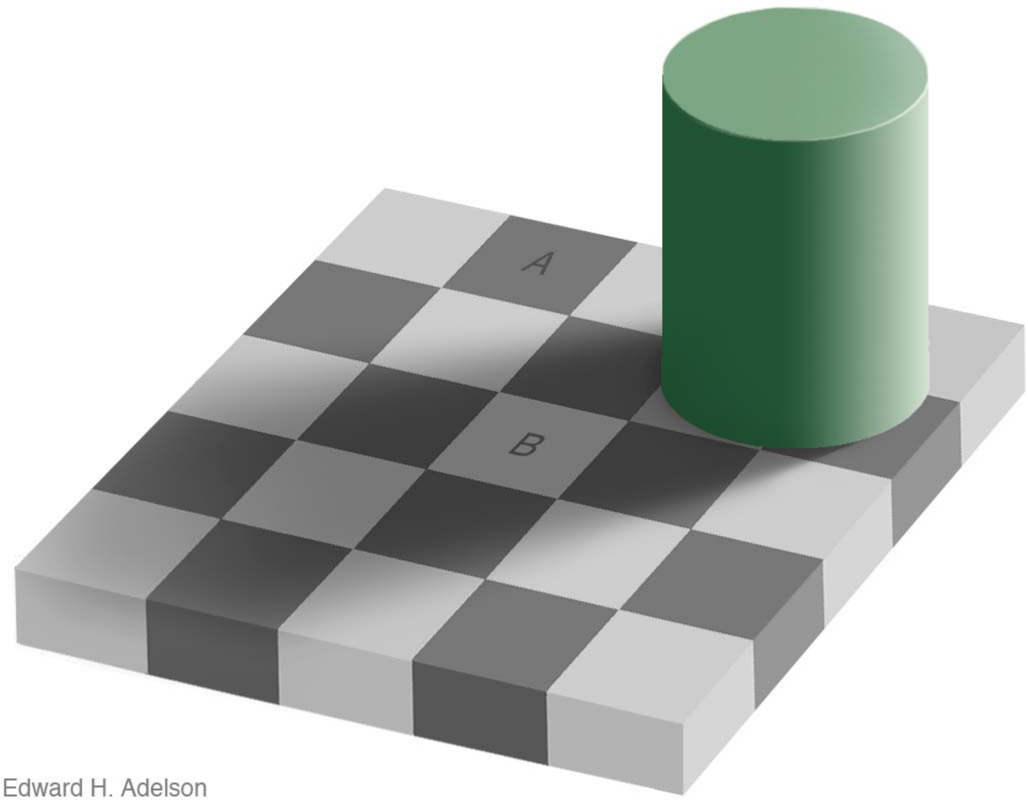






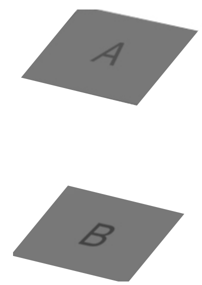


Vision is Hard

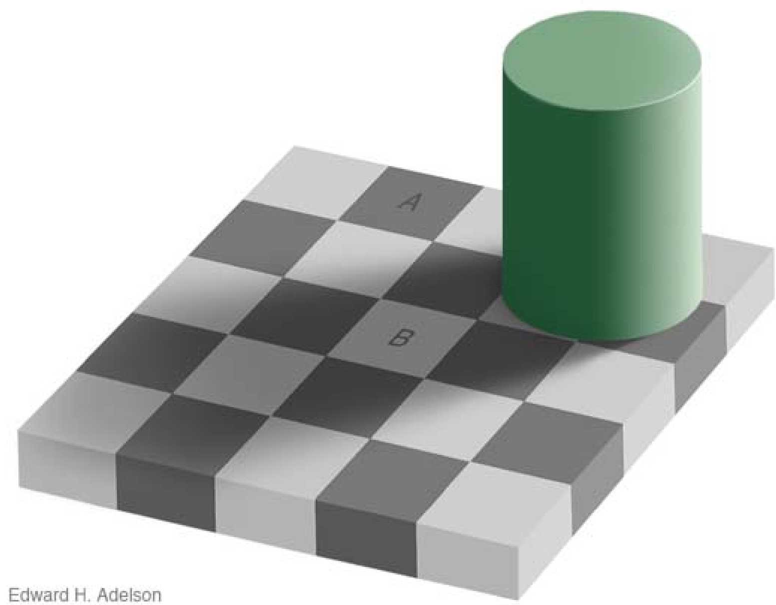


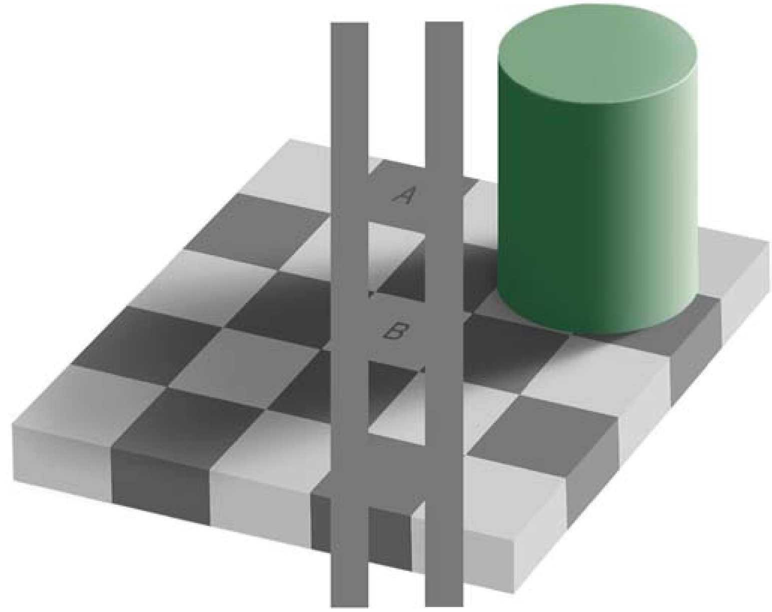
Edward H. Adelson

Vision is Hard



Vision is Hard

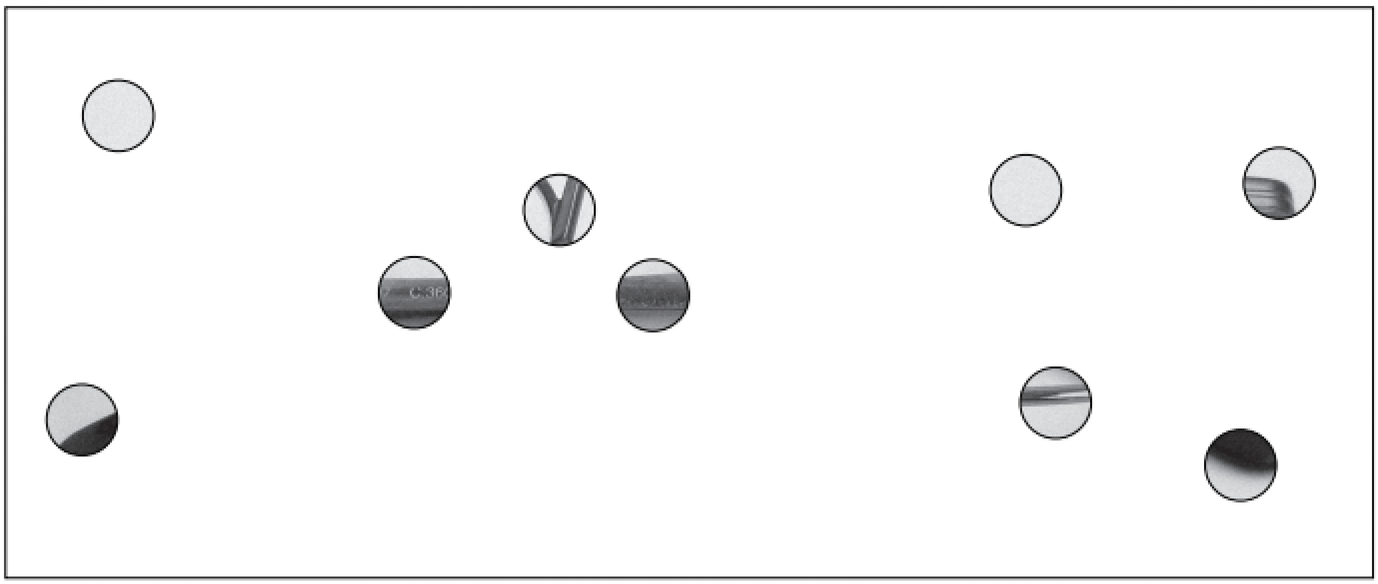




Vision is Hard

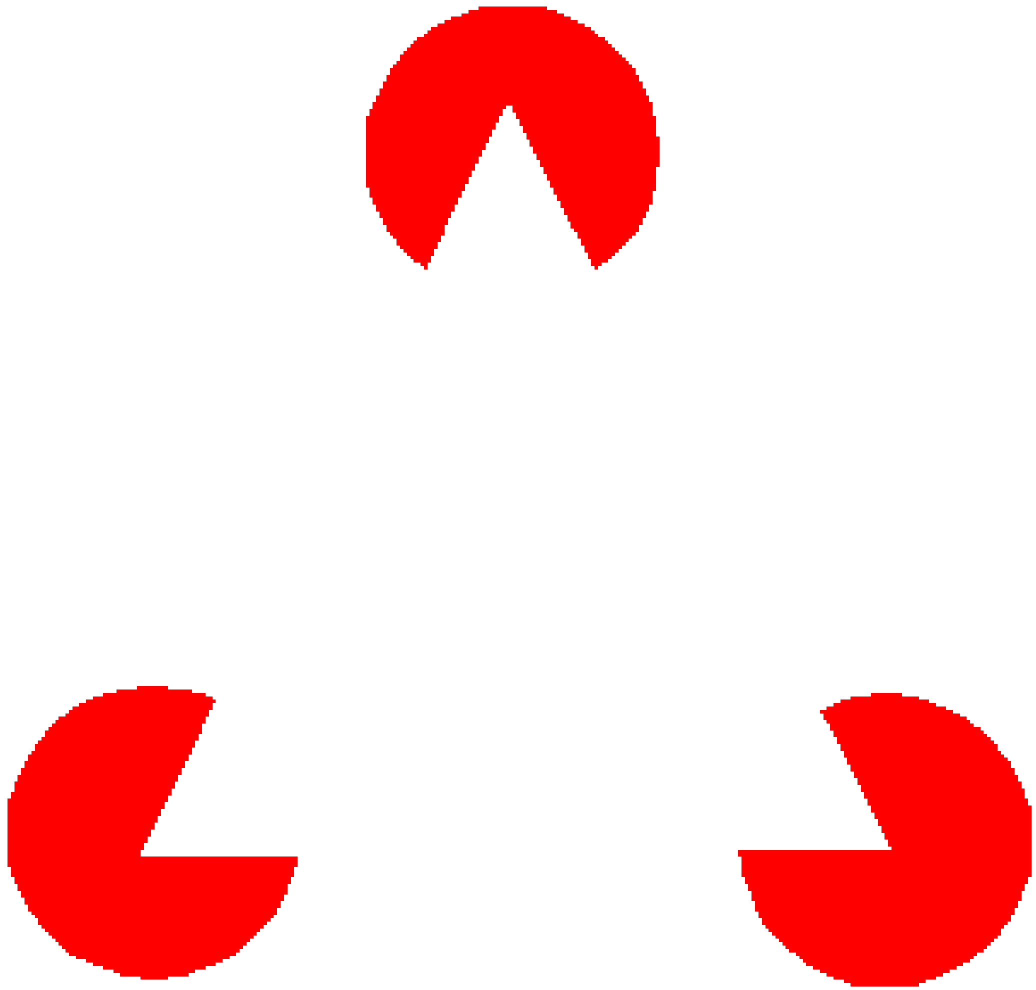
- A 2D picture may be produced by many different 3D scenes

Vision is Hard



Vision is Hard

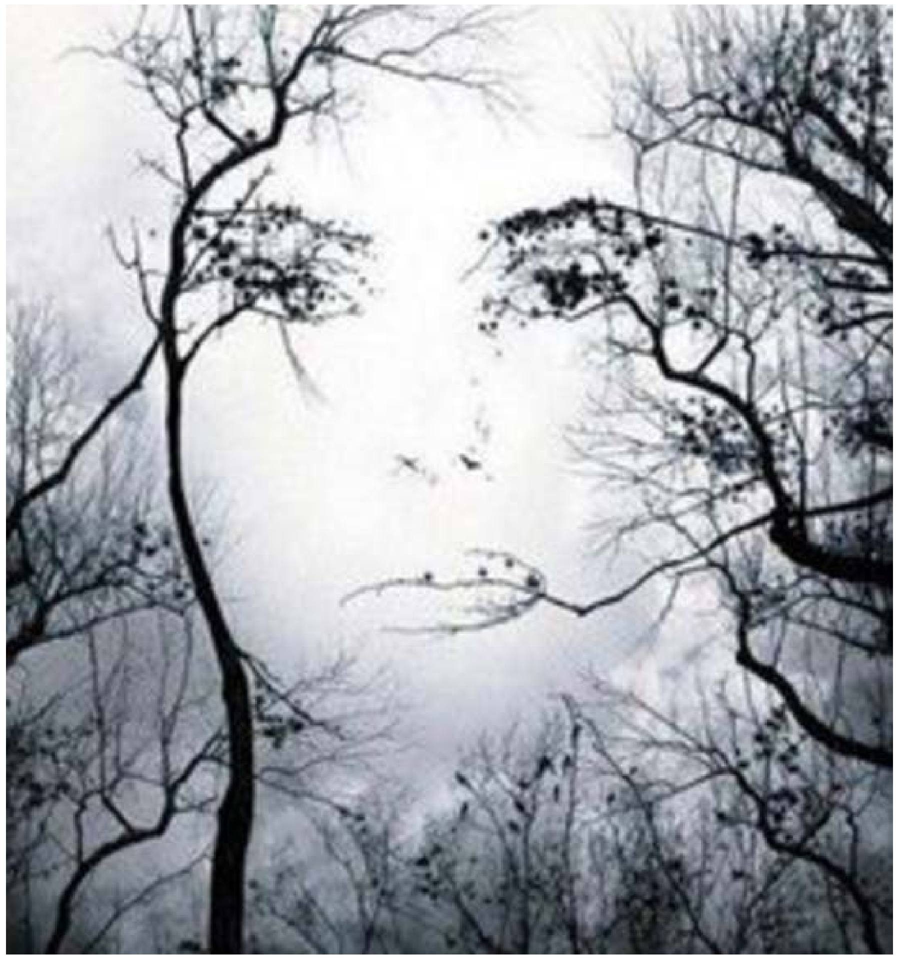
Vision is Inferential

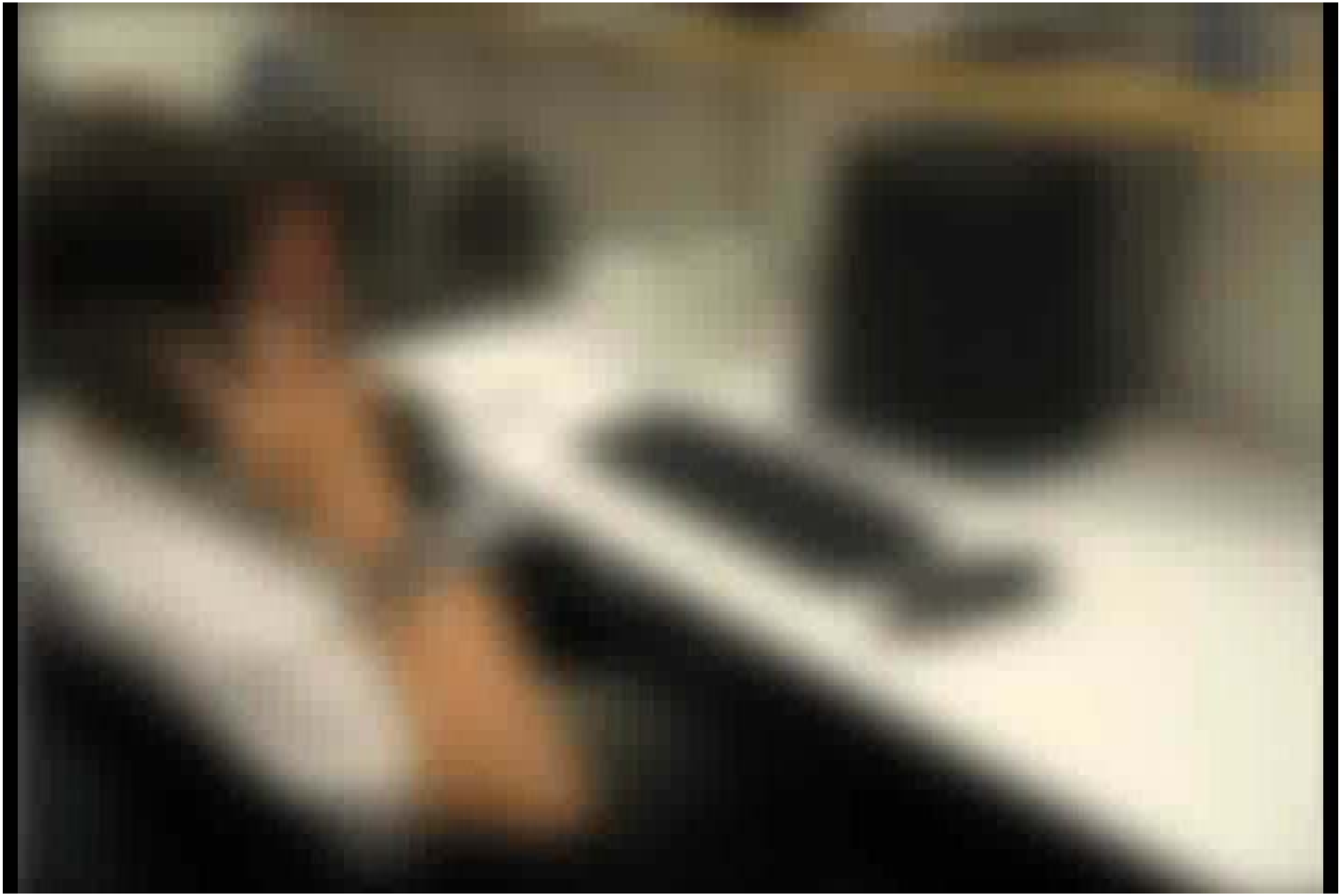


Vision is Fascinating

Vision is Fascinating

Vision is Fascinating





Why is Vision Hard?

- Loss of information due to projection from 3D to 2D
  - Infinite scenes could have generated a given image
- Image colors depend on surface properties, illumination, camera response function and interactions such as shadows
  - HVS very good at ignoring distractors
- Noise
  - sensor noise and nonlinearities, quantization
- Lots of data
- Conflicts among local and global cues
  - Illusions

The Horizon

- Not all hard to explain phenomena are unusual…

Vanishing Points

How Vision is Used Now

- Examples of state-of-the-art

Some of the following slides by Steve Seitz and Derek Hoiem

Earth Viewers (3D modeling)

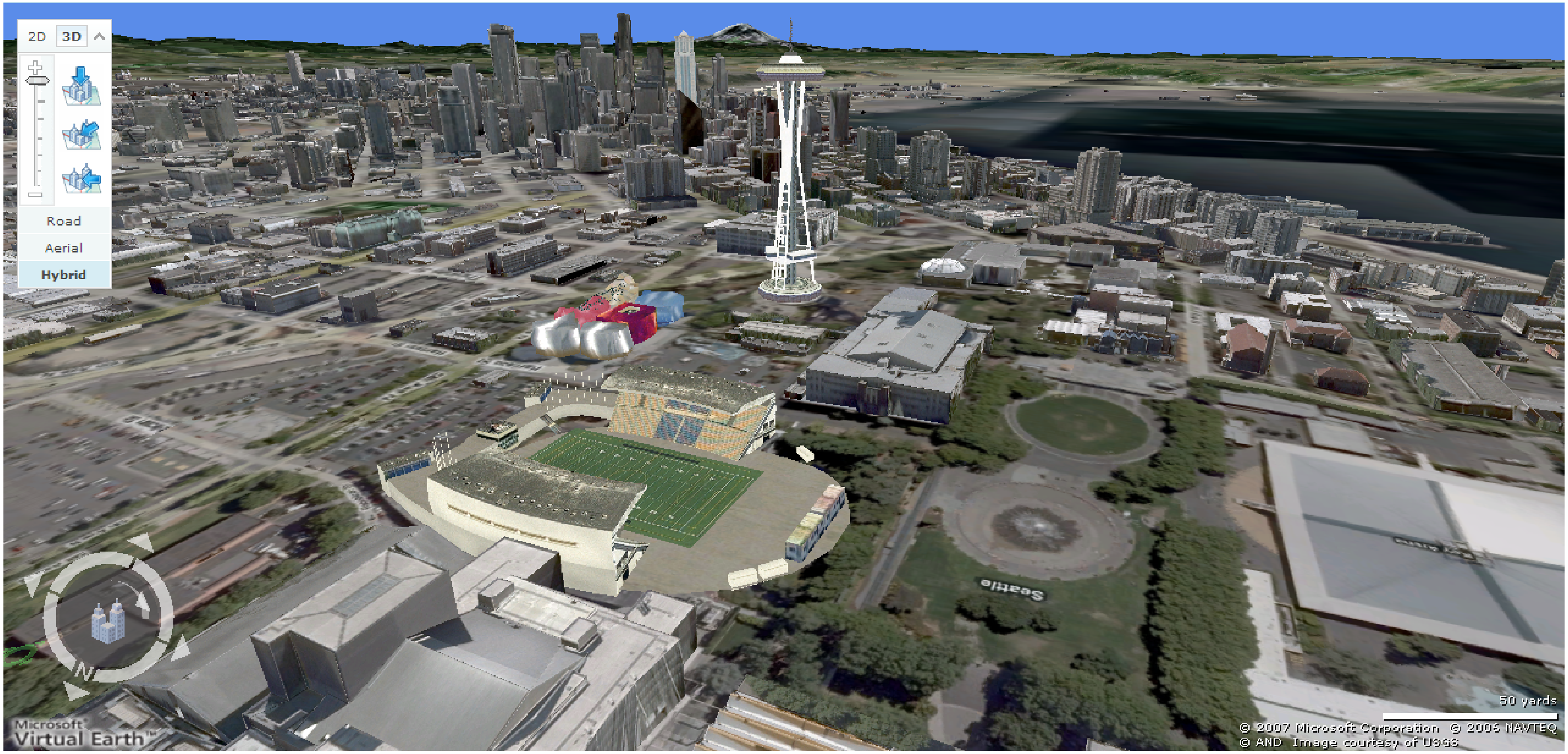
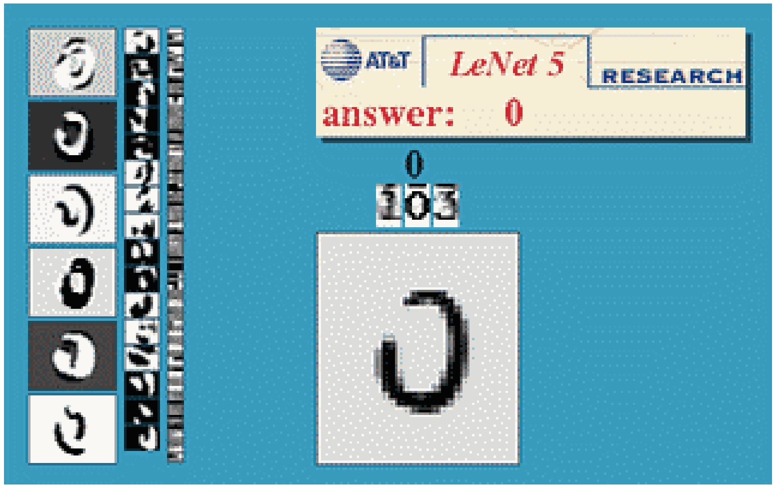


Image from Microsoft's Virtual Earth
(see also: Google Earth)

Optical Character Recognition (OCR)



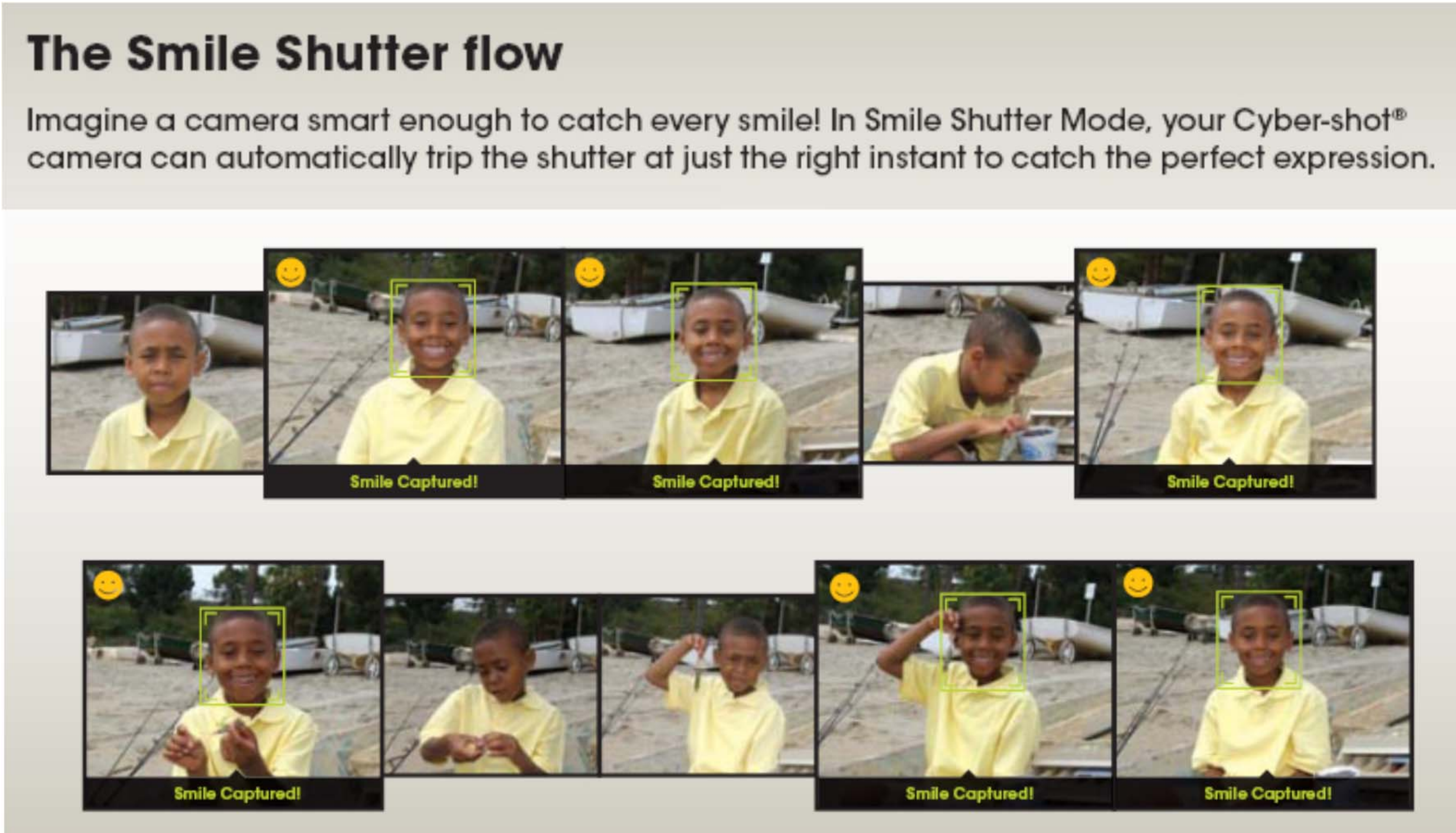
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face Detection

- Most digital cameras detect faces (and more)

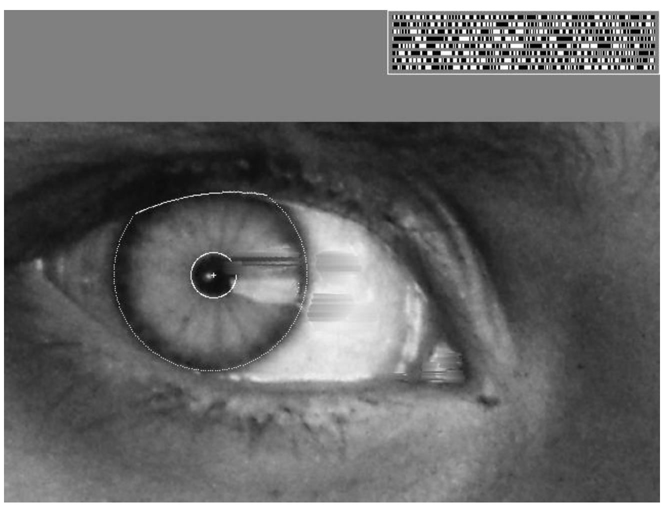
Smile detection

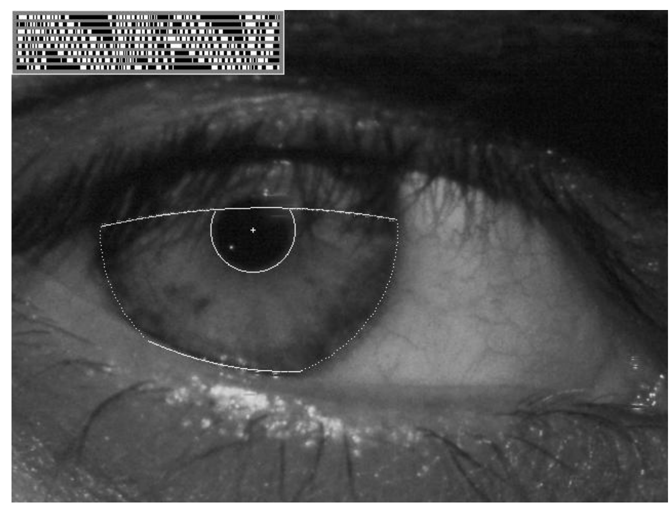


[Sony Cyber-shot® T70 Digital Still Camera](#)

Vision-based Biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*"





Login without a Password…

Fingerprint scanners on
many new laptops,
other devices

Face recognition systems now
beginning to appear more widely
http://www.sensiblevision.com/

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html

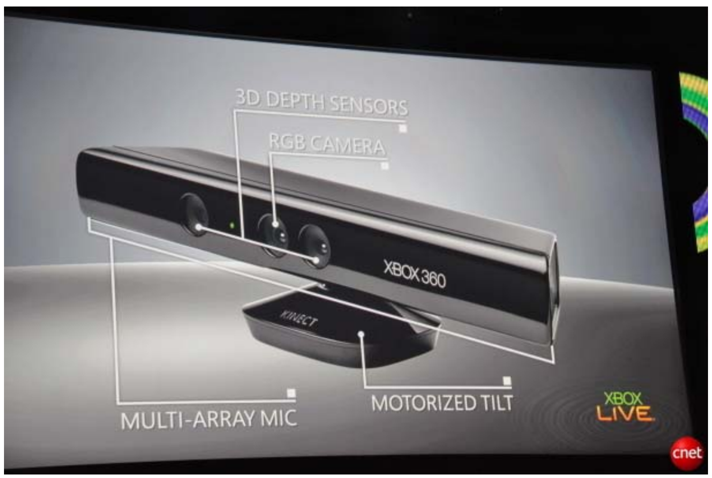
Smart cars

- [Mobileye](http://mobileye.com): Vision systems currently in many cars

http://mobileye.com/technology/applications/vehicle-detection/forward-colision-warning/
http://mobileye.com/technology/applications/pedestrian-detection/pedestrian-collision-warning/

"Subaru thinks cameras are better than radar cruise"
http://www.roadandtrack.com/new-cars/news/a6852/subaru-camera-controlled-cruise/

Interactive Games: Kinect



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

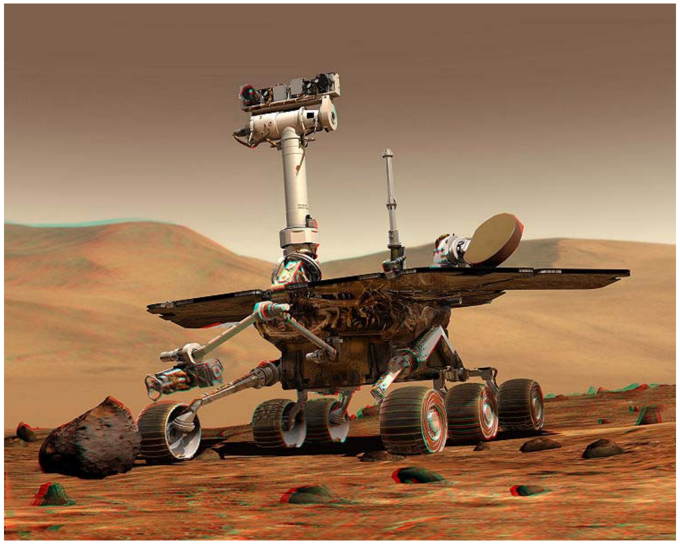
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “Computer Vision on Mars” by Matthies et al.

Industrial robots

Vision-guided robots position nut runners on wheels

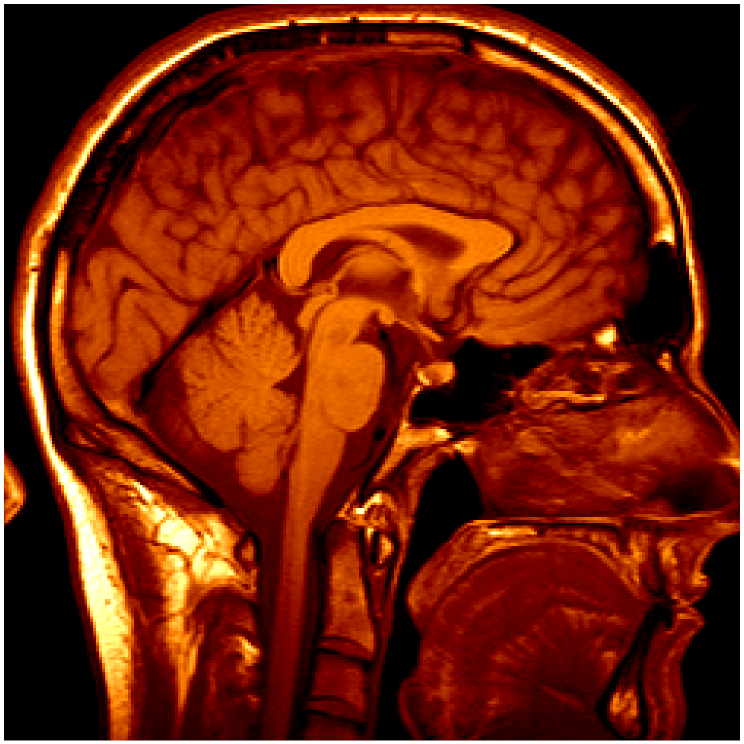
Mobile Robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

http://www.robocup.org/

Medical Imaging

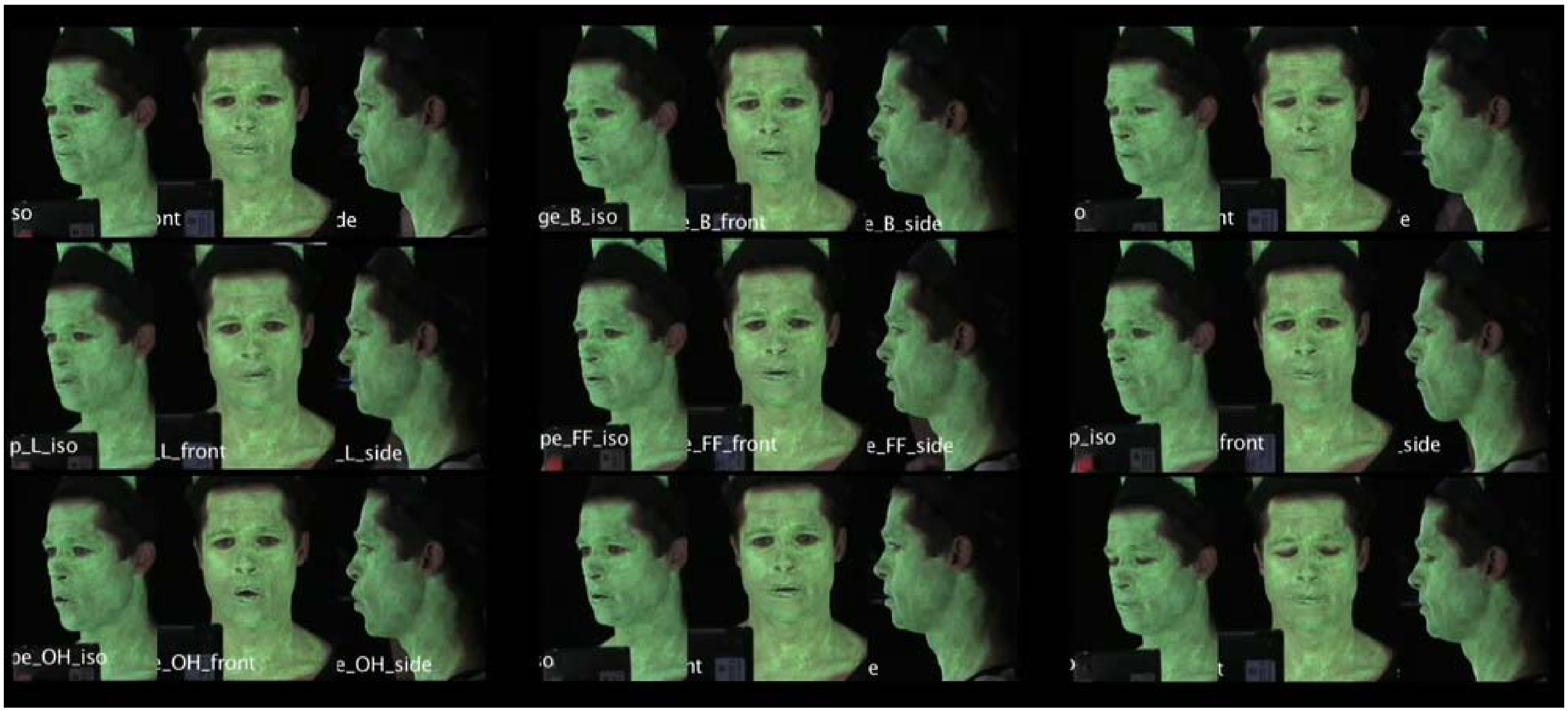


3D imaging
MRI, CT

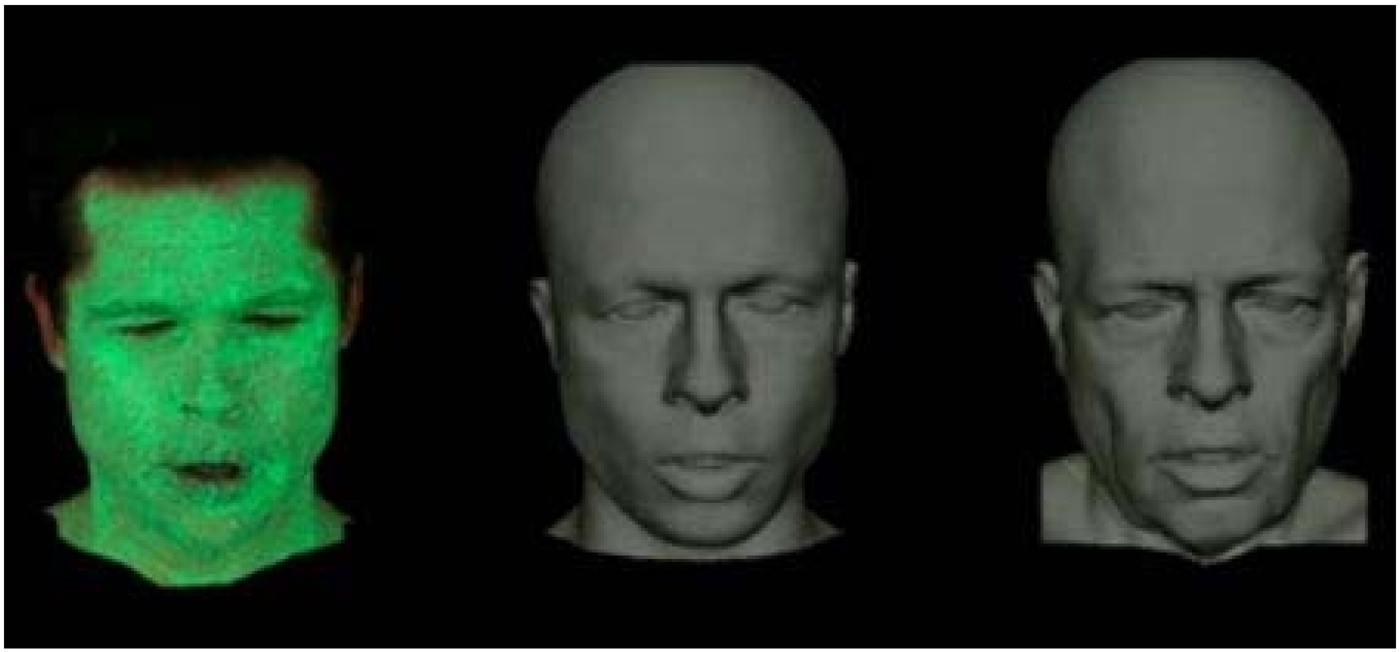
Image guided surgery
[Grimson et al., MIT]

The Curious Case of Benjamin Button (2008)

The Curious Case of Benjamin Button (2008)



The Curious Case of Benjamin Button (2008)



The Curious Case of Benjamin Button (2008)

The Curious Case of Benjamin Button (2008)

The Curious Case of Benjamin Button (2008)

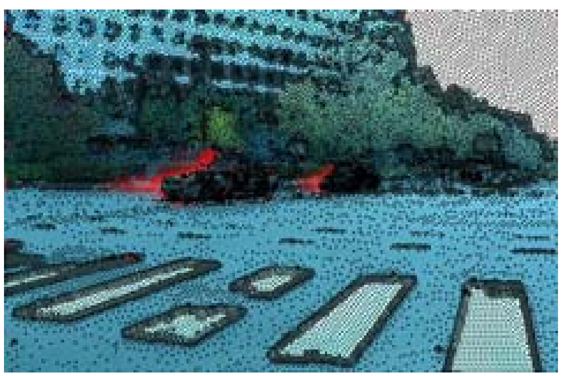
The Curious Case of Benjamin Button (2008)

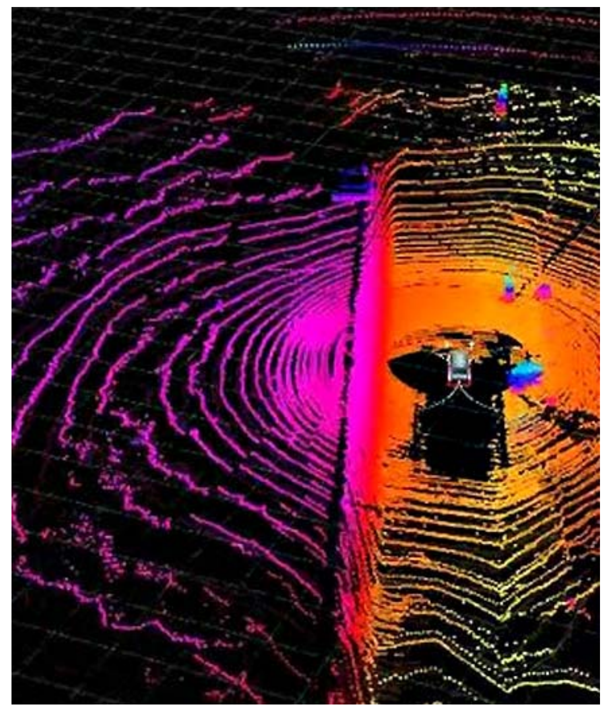
The DARPA Grand Challenge

- 2004: goal drive 150 miles in the Mojave desert autonomously
  - Longest any participant traveled: 7.4 miles
- 2005: 5 teams completed race
  - Won by Stanley from Stanford (VW Touareg)

The DARPA Urban Challenge

- 2007: navigate “urban” environment to complete missions
- 50 human-driven cars also on road
- Maps and locations of stop signs given to teams
- Improved laser sensors





The DARPA Urban Challenge

Images courtesy of DARPA

The DARPA Urban Challenge

The DARPA Urban Challenge

The DARPA Urban Challenge

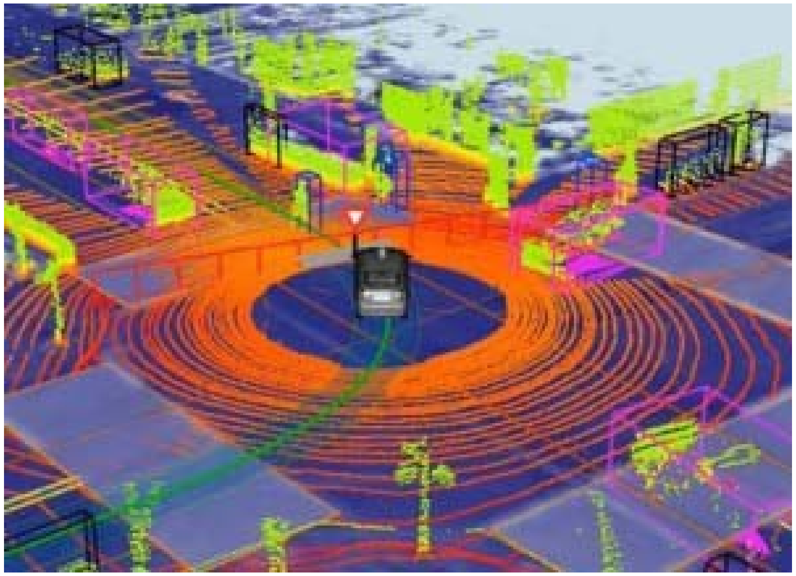
Google Car

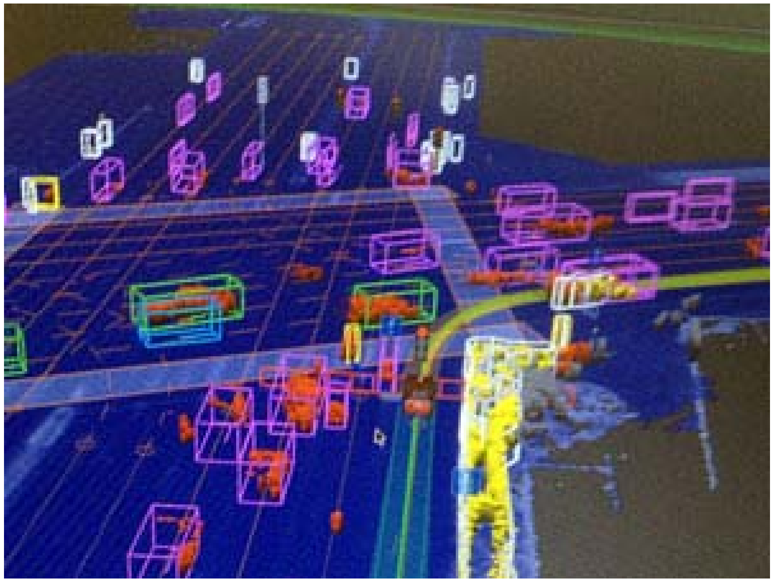
Google Car

- No steering wheel, no brake pedal

Google Car

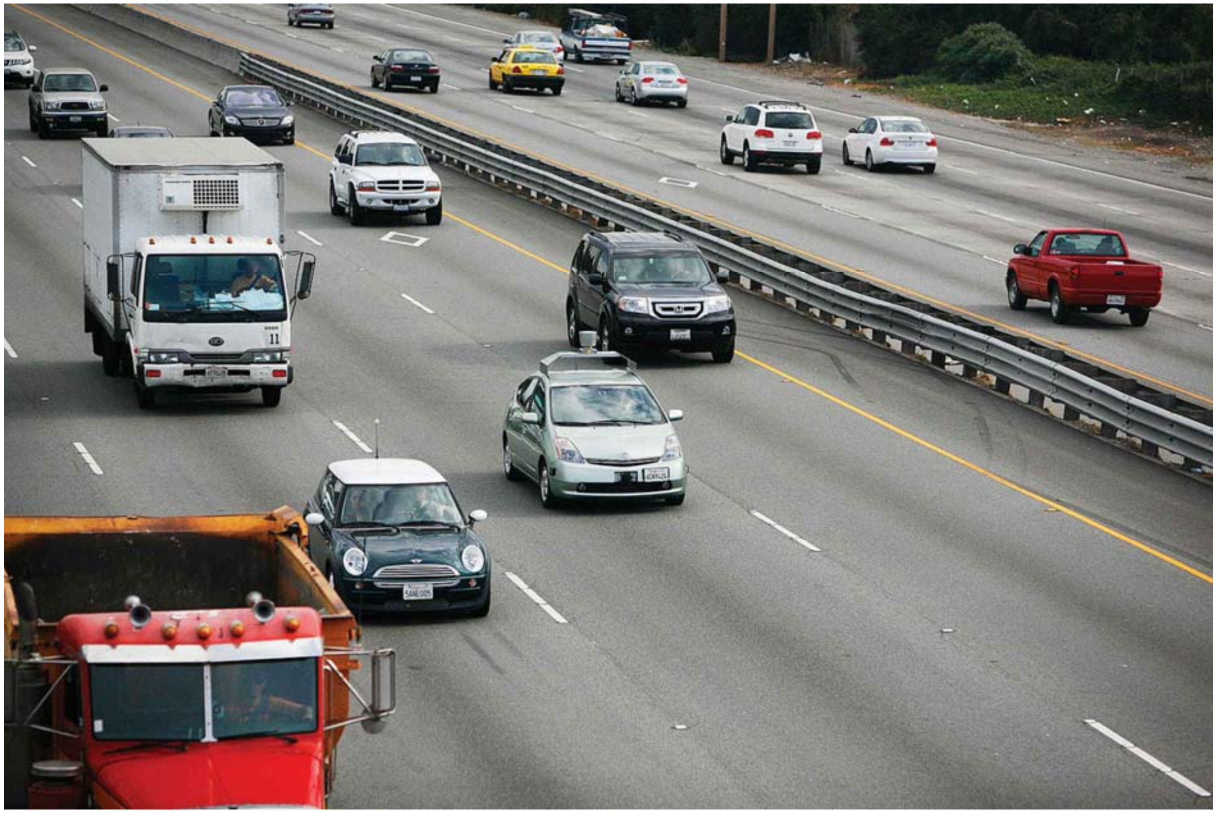
Google Car





Google Car

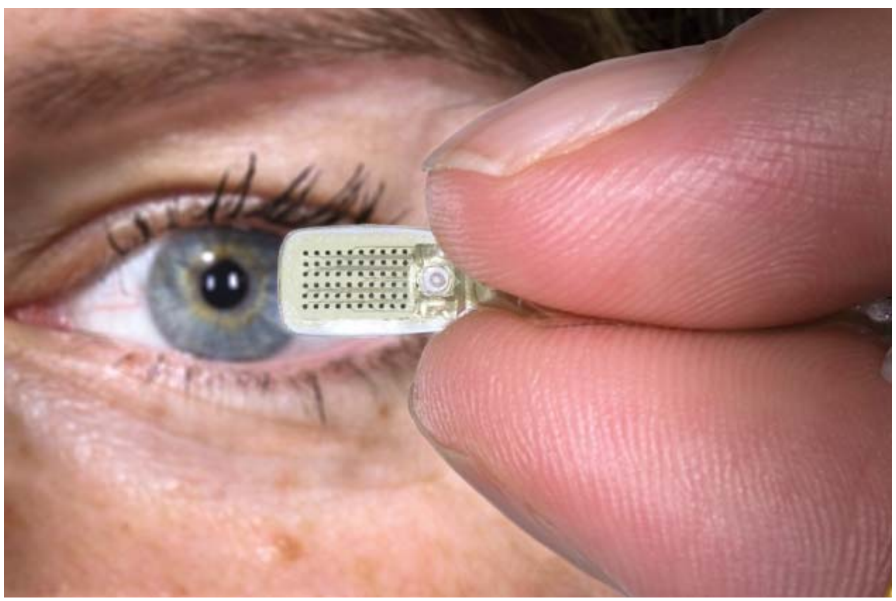
- Passed 1,000,000 accident-free miles



- What are the remaining obstacles?

The Real Bionic Woman

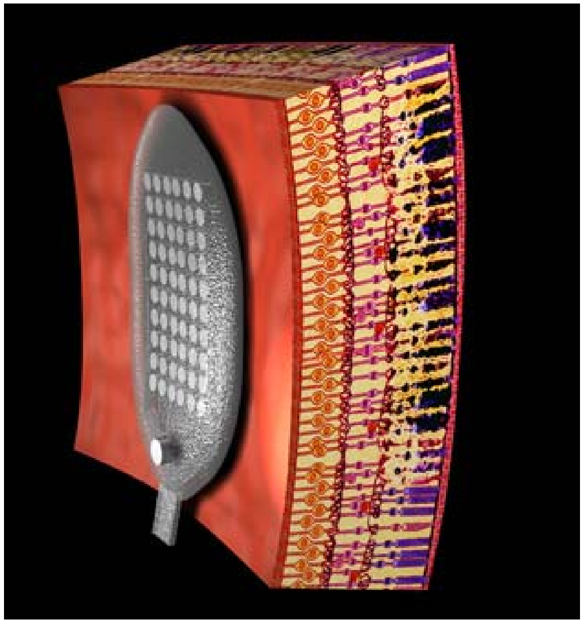
- Lisa Kulik
  - – Blind since 2000



The Real Bionic Woman

The Real Bionic Woman

- Electrodes implanted in Lisa's eye
- Connected to camera mounted on glasses
- 6x10 image
- Brain has to be trained
- Lisa can see moon, fireworks



OrCam

- Smart camera mounted on glass frames
- Speaks through earpiece
- Recognizes text and objects

OrCam


I'm Jon Snow all the way, ready to save the world from White Walkers. On Halloween, I went as Khaleesi and Dax [Shepard], my husband, went as Khal Drogo. We have two dogs and a baby, and they were all dragons.
Just always makes me think of Veronica. She was normally in a pair of Frye motorcycle boots and a Gap corduroy jacket, which was her uniform for a lot of the series.

Tools Needed for Course

- Math
  - Linear Algebra
  - Signal Processing (to be covered)
  - Some calculus
  - Some geometry
  - Some probability

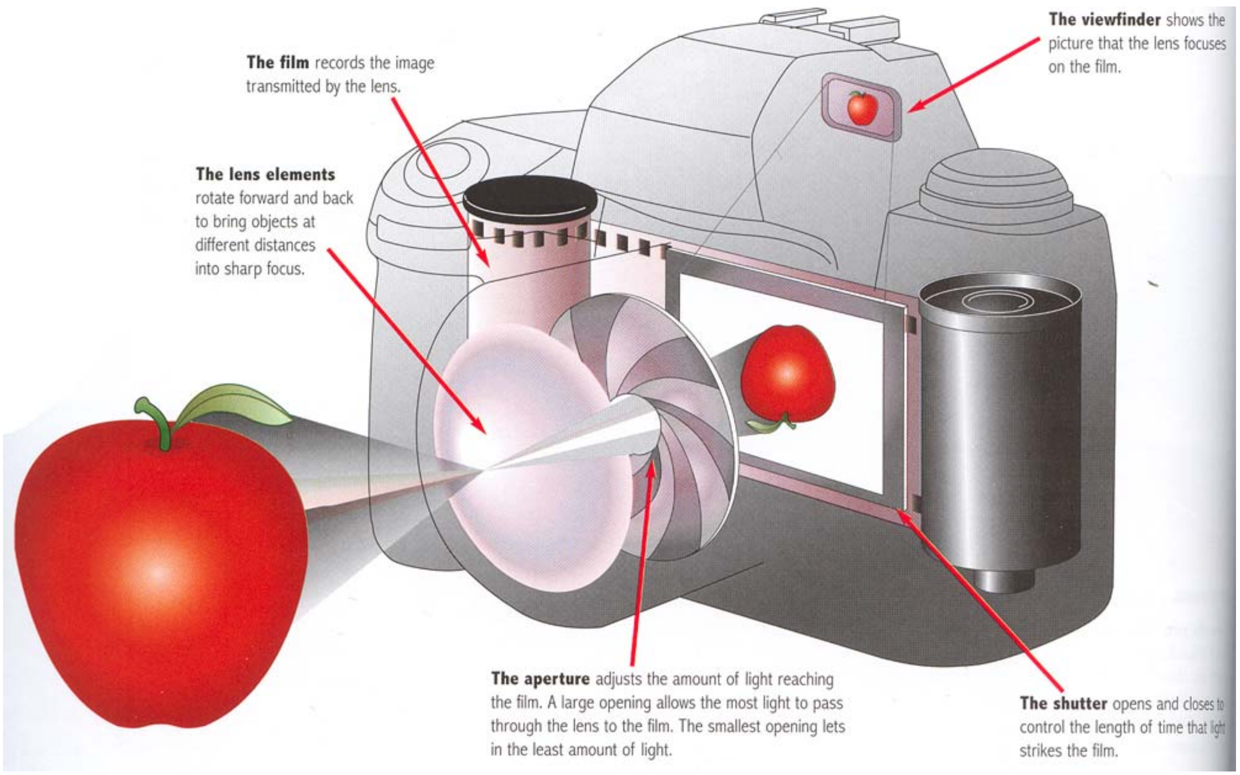
- Computer Science
  - Algorithms
  - Programming
    - Suggested matlab or python
    - Other languages acceptable if you can manipulate images effortlessly

Camera Body

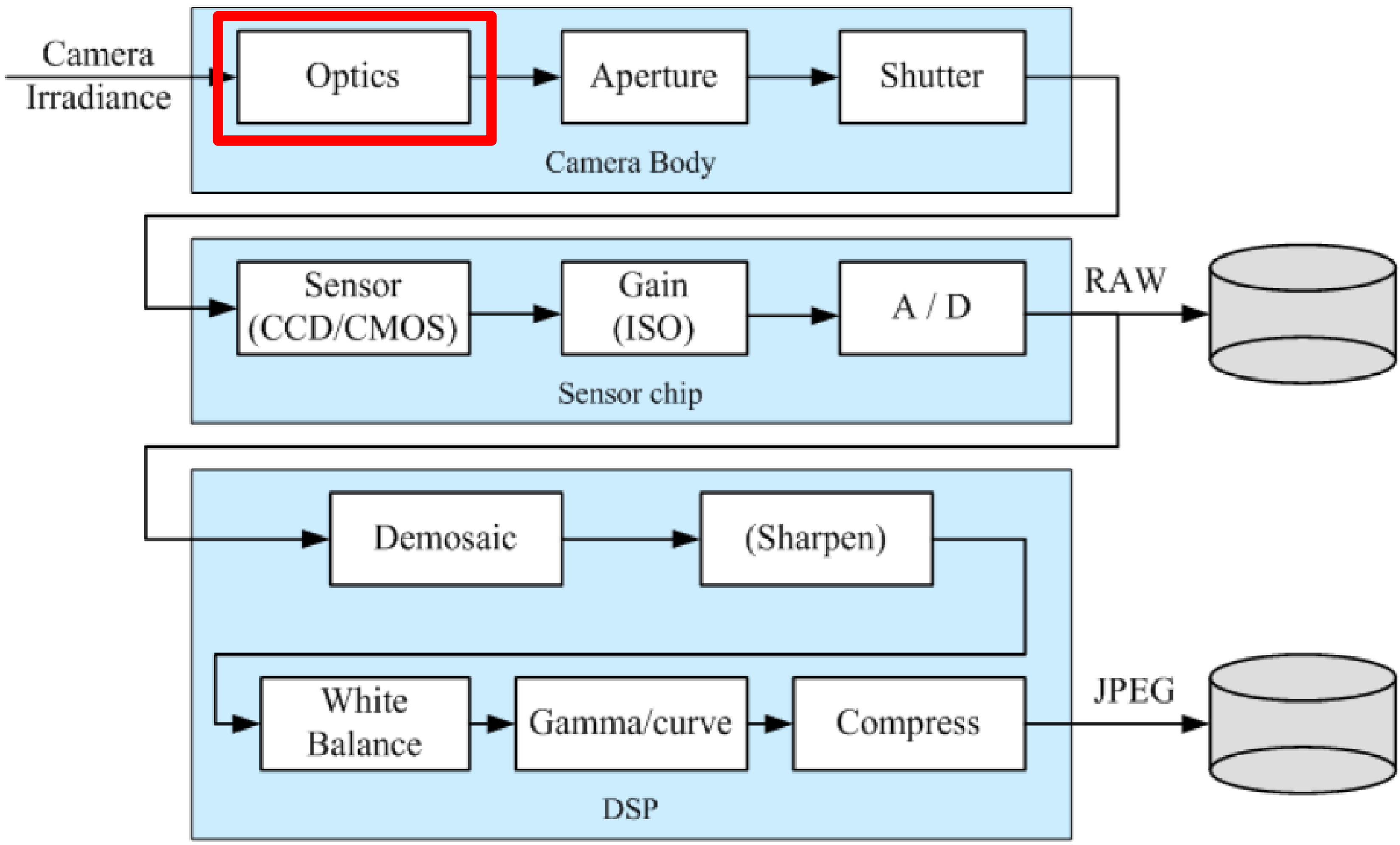
Slides by G. Doretto

Camera Body

- Lens and viewpoint determine perspective
- Aperture and shutter speed determine exposure
- Aperture and other effects determine depth of field
- Film or sensor record image



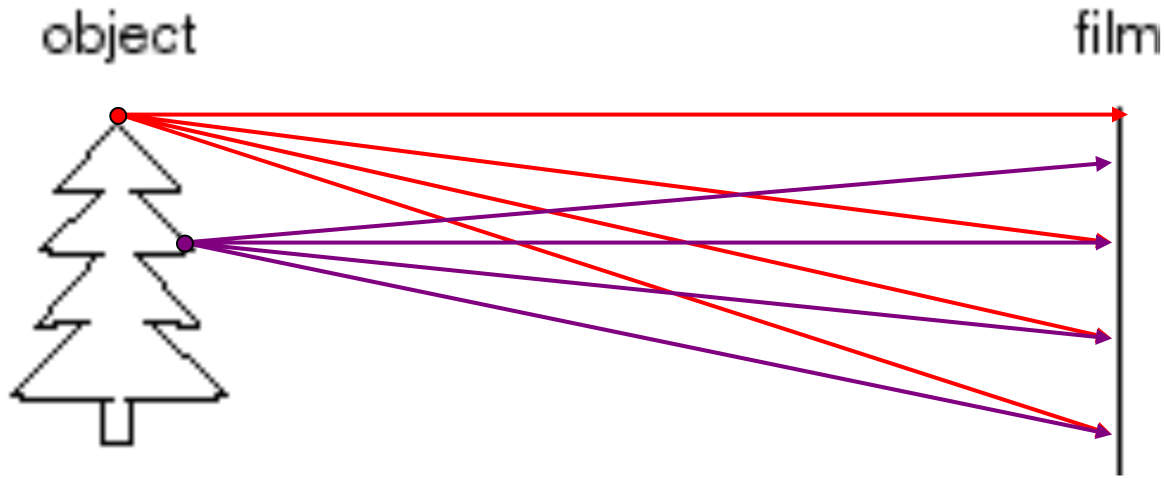
Camera Body: Optics



Camera Body: Optics

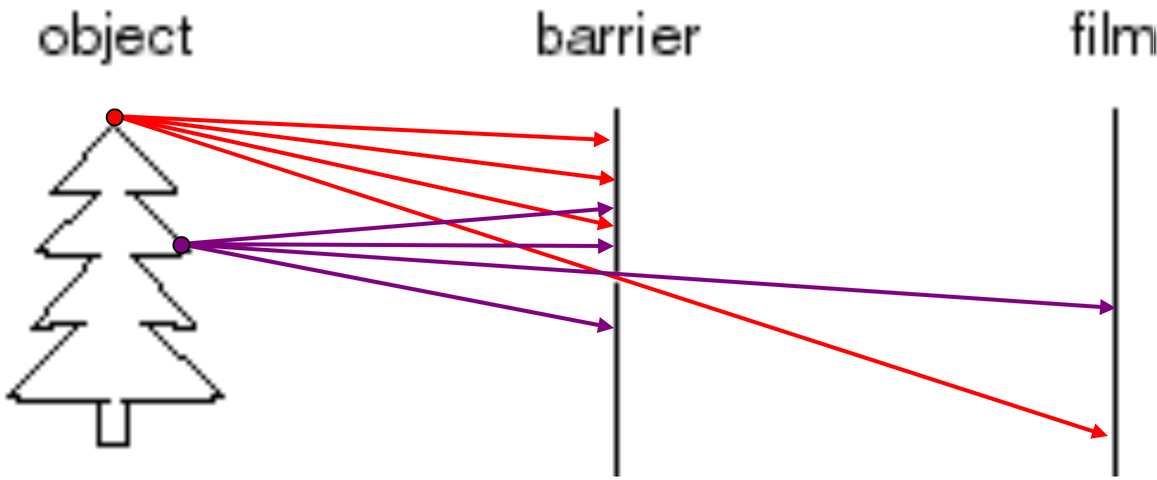
- Pinhole camera model
- Focusing light: Thin lens model
- Field of view (zoom)

How do we see the world?



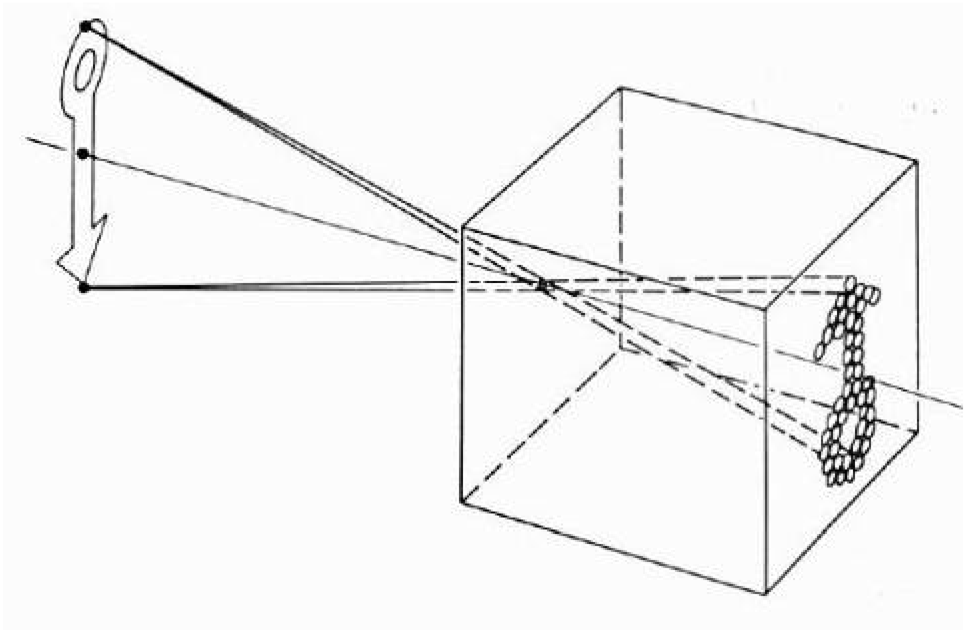
- Let’s design a camera
  - Idea 1: put a piece of film in front of an object
  - Do we get a reasonable image?

Pinhole camera



- Add a barrier to block off most of the rays
  - This reduces blurring
  - The opening known as the **aperture**
  - How does this transform the image?

Pinhole camera model



- Pinhole model:
  - Captures **pencil of rays** – all rays through a single point
  - The point is called **Center of Projection (COP)**
  - The image is formed on the **Image Plane**
  - **Effective focal length *f*** is distance from COP to Image Plane

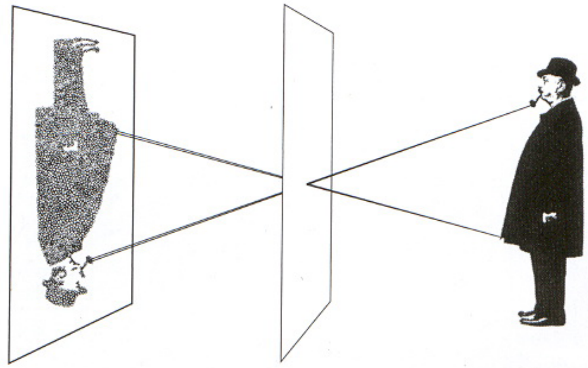
Home-made pinhole camera

Why so blurry?

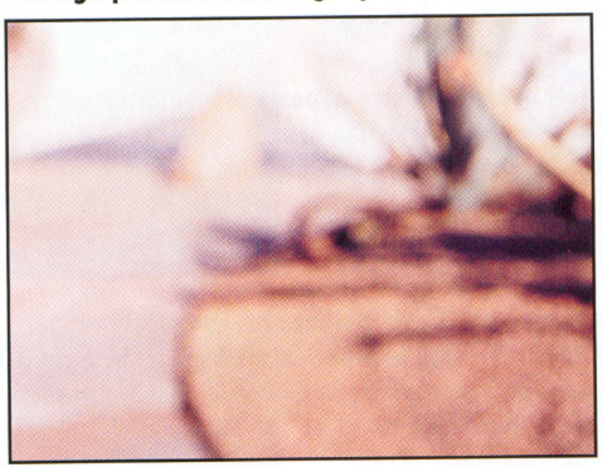
http://www.debevec.org/Pinhole/

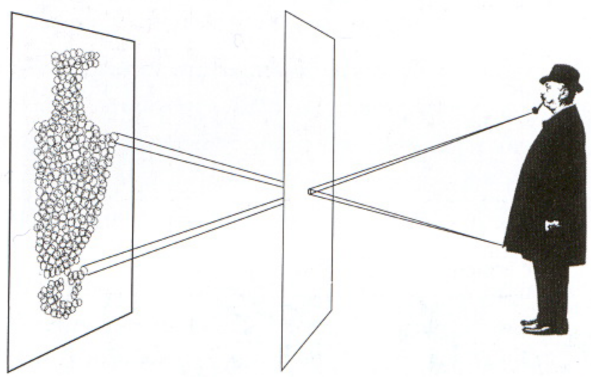
Pinhole size?

Photograph made with small pinhole



Photograph made with larger pinhole

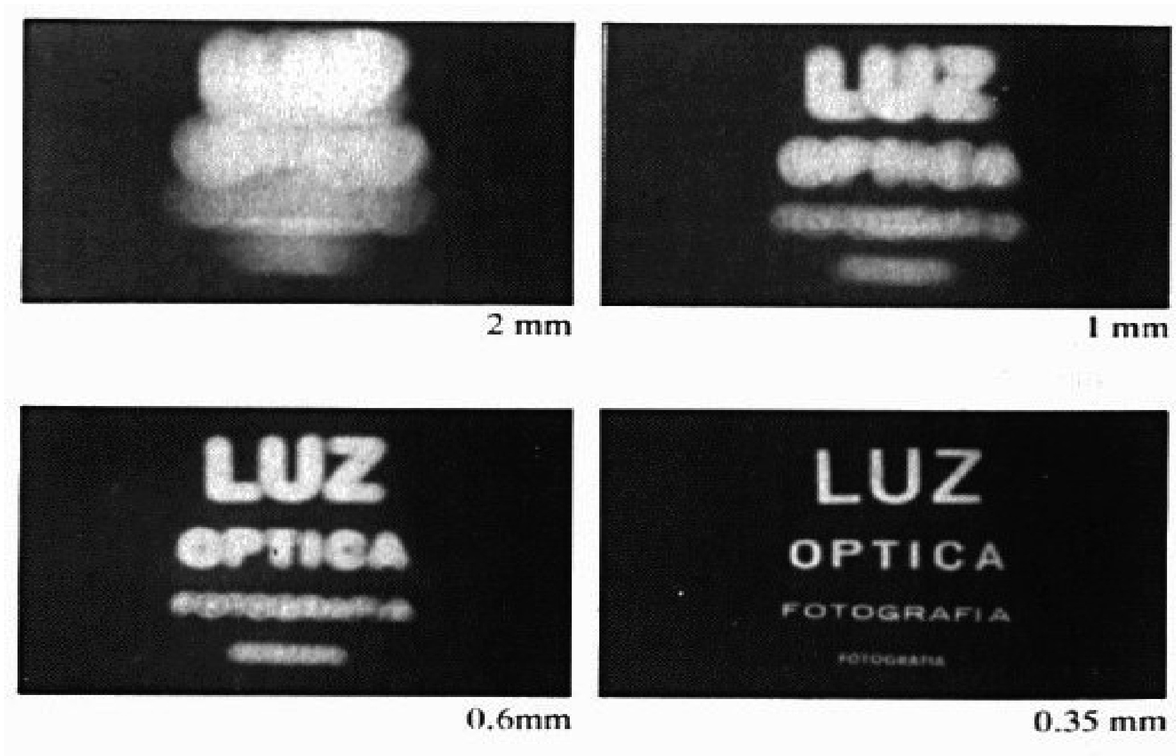




From Photography, London et al.

Shrinking the aperture

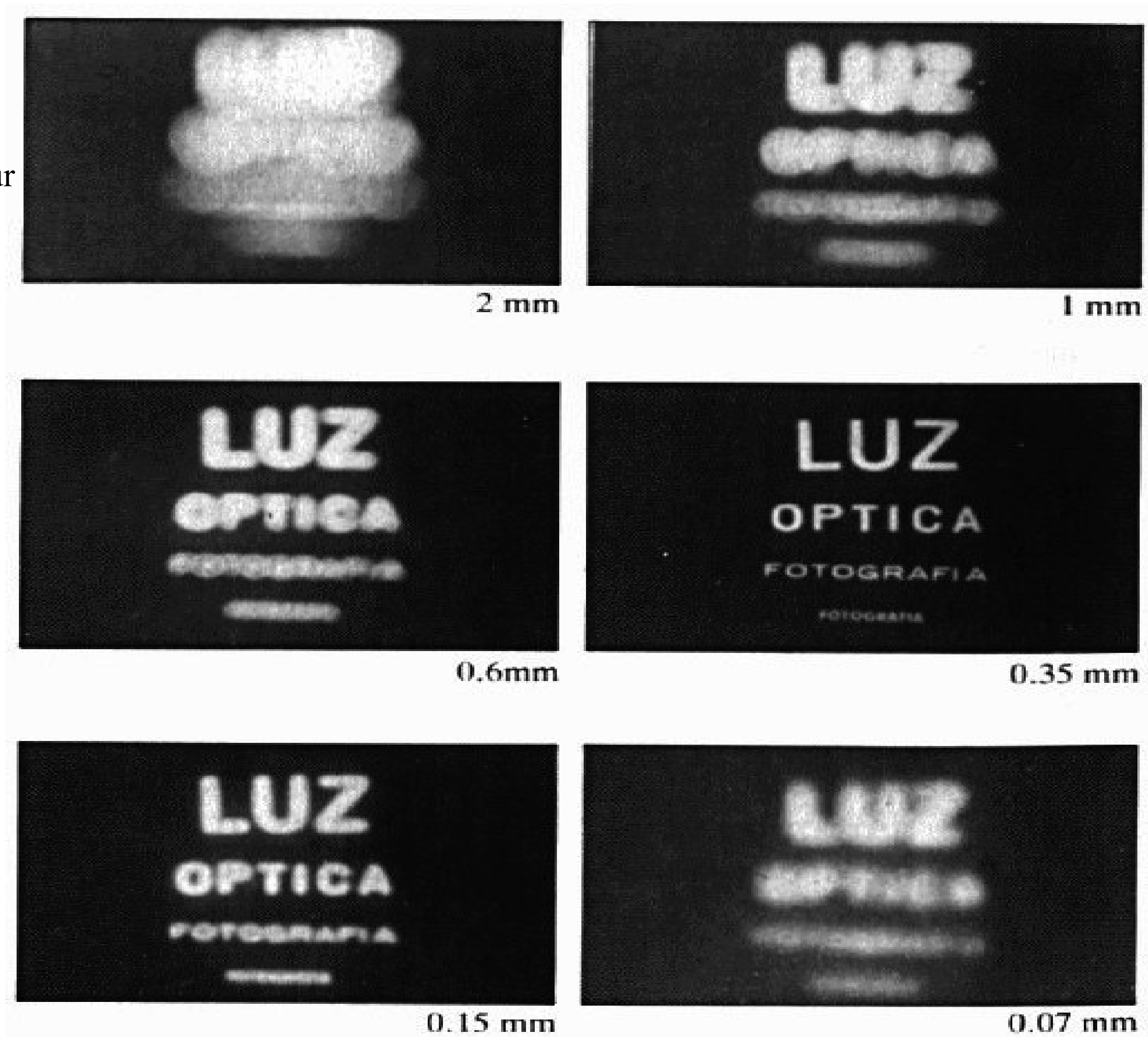
Large pinhole
→ Geometric blur



- Why not make the aperture as small as possible?
  - Less light gets through
  - Diffraction effects…

Shrinking the aperture

Large pinhole
➔ Geometric blur



Optimal pinhole
➔ Very little light

Small pinhole
➔ Diffraction blur

Camera Body: Optics

- Pinhole camera model
- Focusing light: Thin lens model
- Field of view (zoom)

Problem with pinhole?

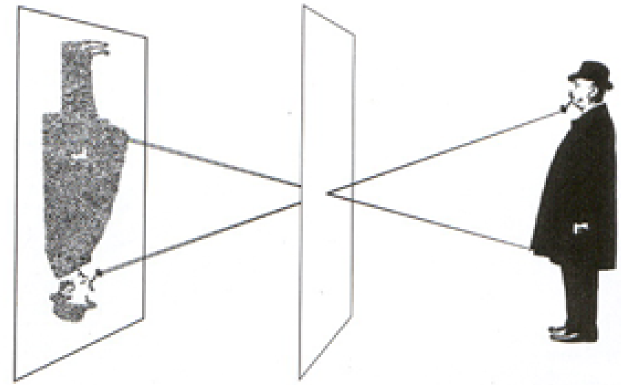
- Not enough light!
- Diffraction limits sharpness

Solution: refraction!

From Photography, London et al.

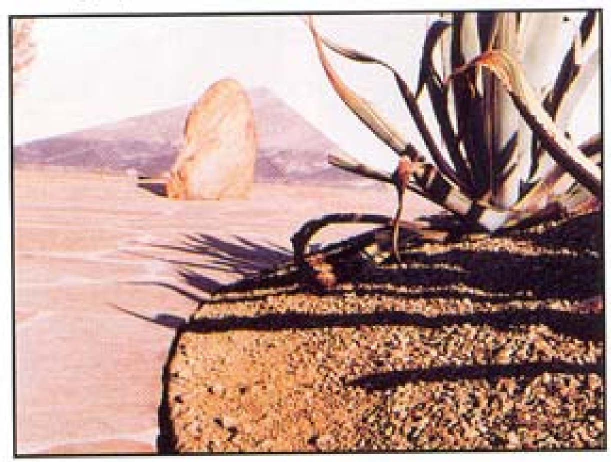
- Gather more light!
- But need to be focused

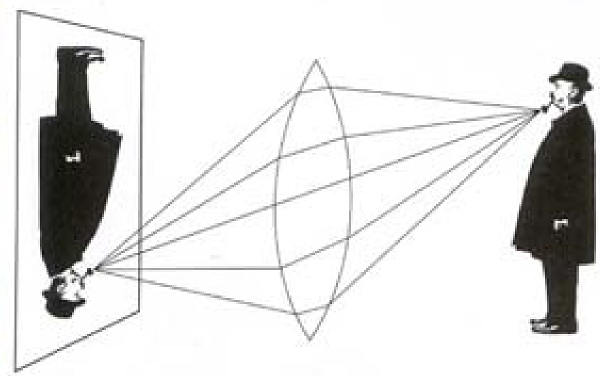
Photograph made with small pinhole



To make this picture, the lens of a camera was replaced with a thin metal disk pierced by a tiny pinhole, *equivalent in size to an aperture of f/182. Only a few rays of light from each point on the subject got through the tiny opening, producing a soft but acceptably clear photograph. Because of the small size of the pinhole, the exposure had to be 6 sec long.*

Photograph made with lens



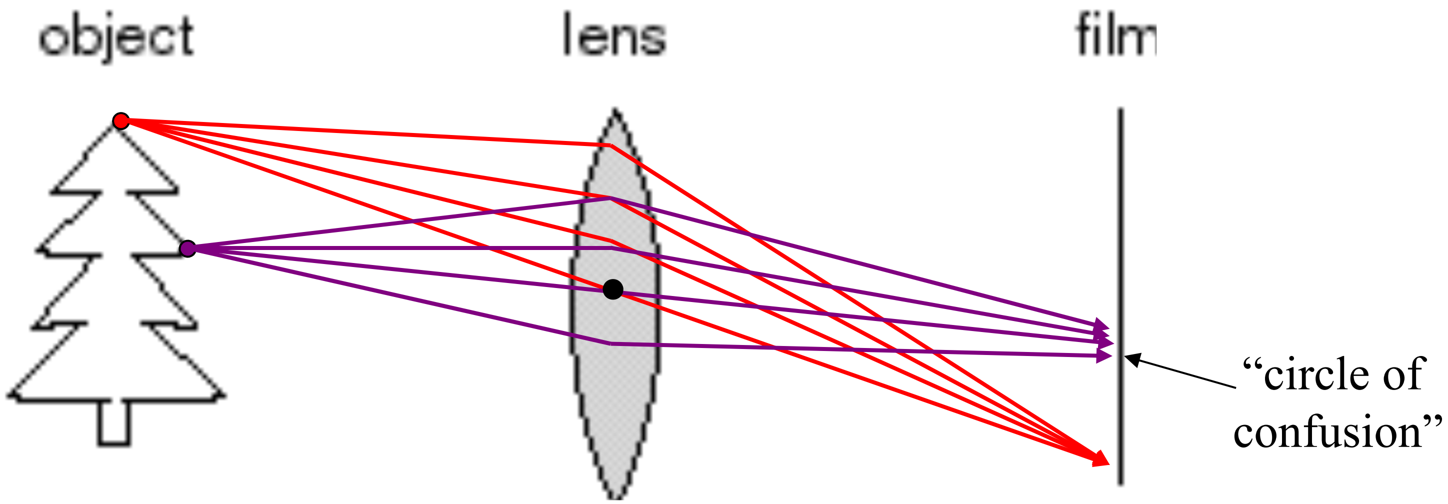


This time, using a simple convex lens with an f/16 aperture, *the scene appeared sharper than the one taken with the smaller pinhole, and the exposure time was much shorter, only 1/100 sec.*

The lens opening was much bigger than the pinhole, letting in far more light, but it focused the rays from each point on the subject precisely so that they were sharp on the film.

From Photography, London et al.

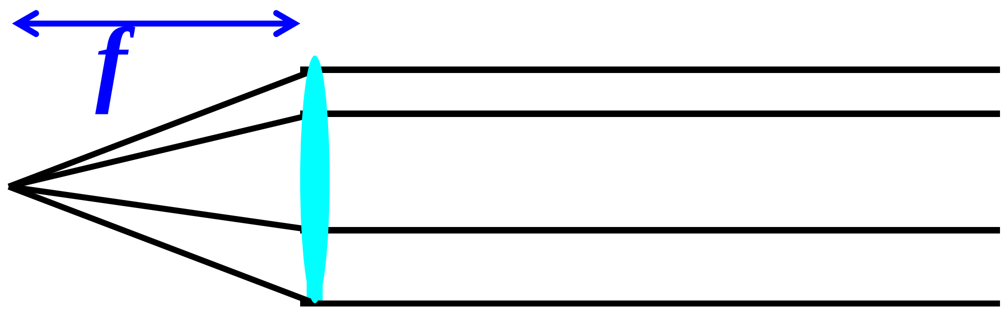
Focus and Defocus



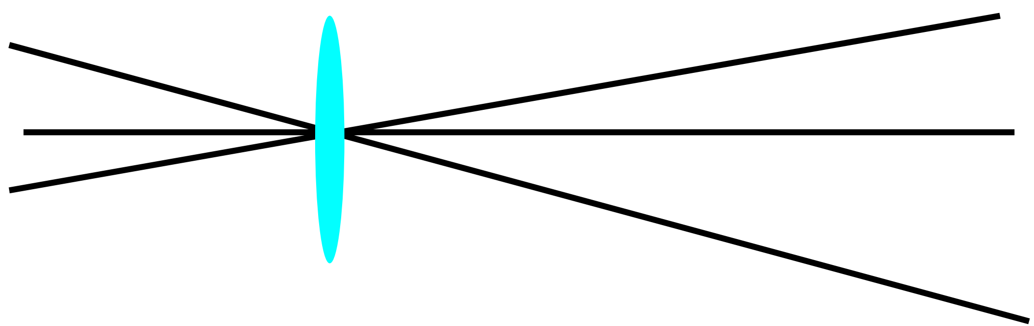
- A lens focuses light onto the film
  - There is a specific distance at which objects are "in focus"
    - other points project to a "circle of confusion" in the image
  - Changing the shape of the lens changes this distance

Thin lens optics

- Simplification of geometrical optics for well-behaved lenses
- All parallel rays converge to one point on a plane located at the focal length *f*

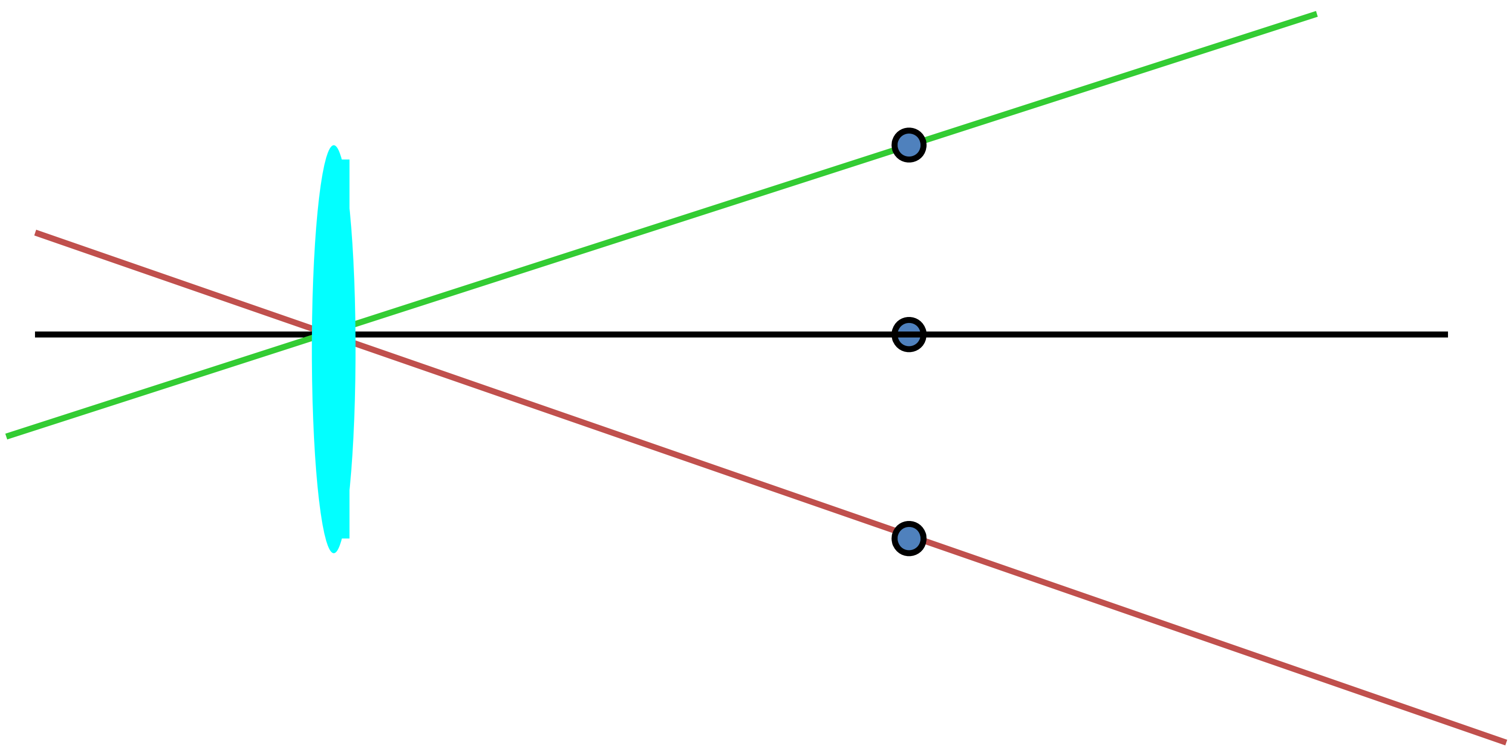


- All rays going through the center are not deviated
  - Hence same perspective as pinhole



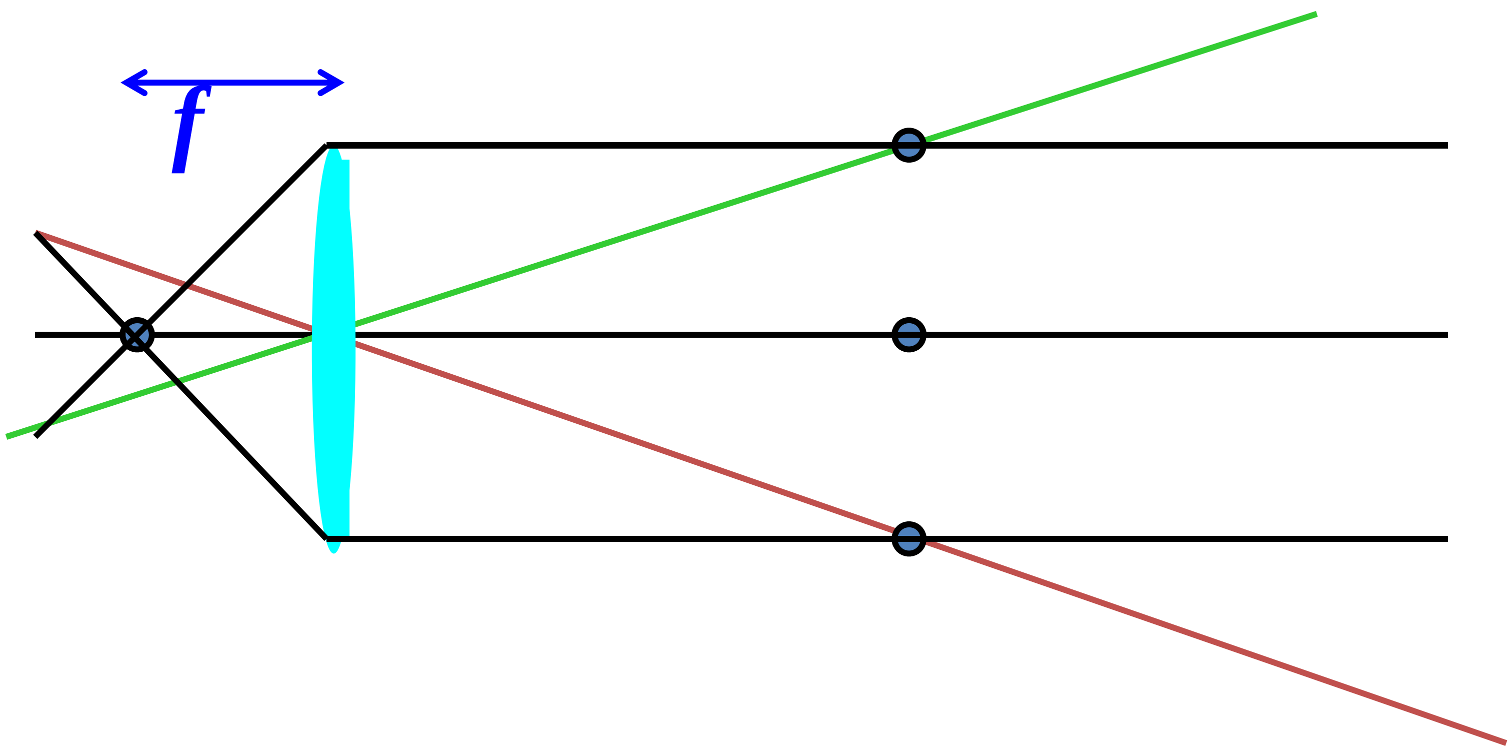
How to trace rays

- Start by rays through the center



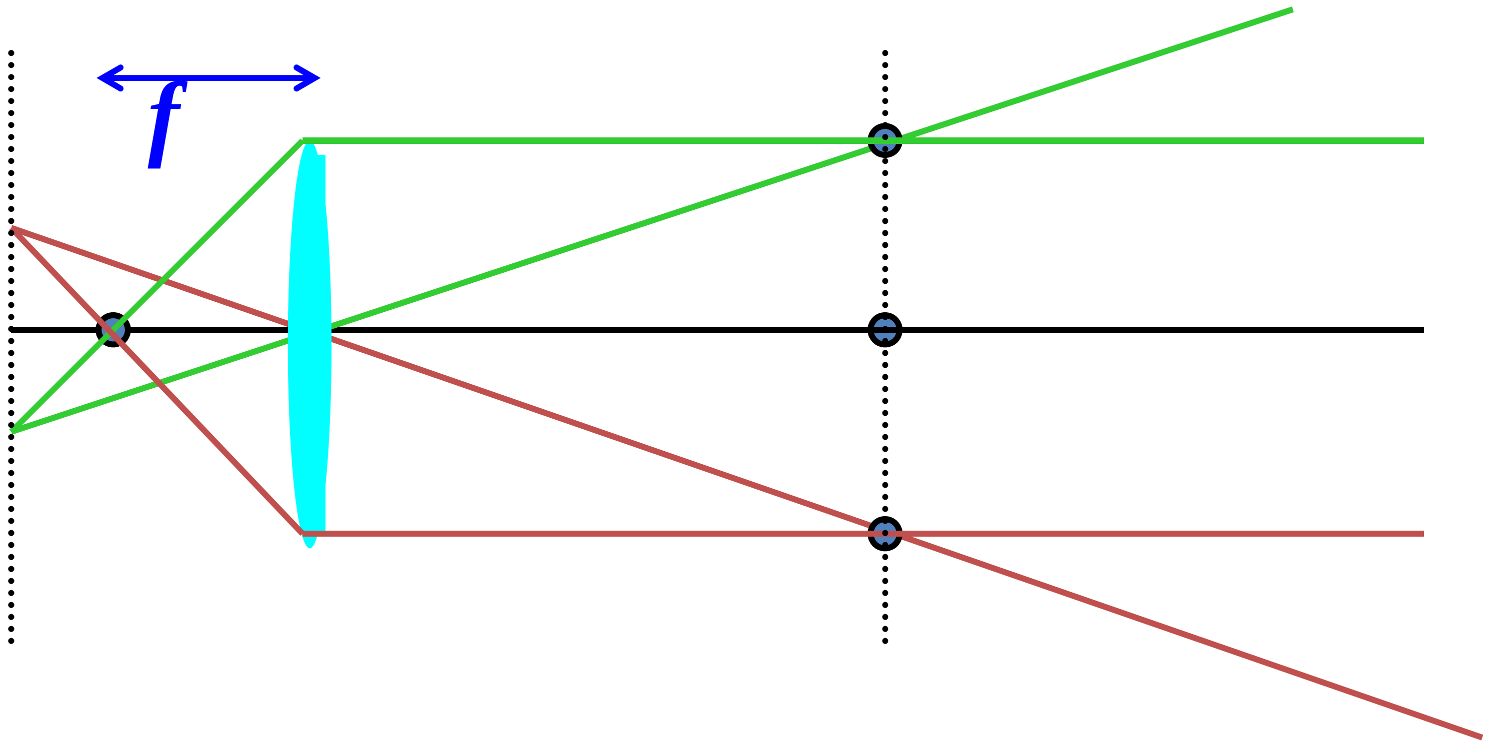
How to trace rays

- Start by rays through the center
- Choose focal length, trace parallels

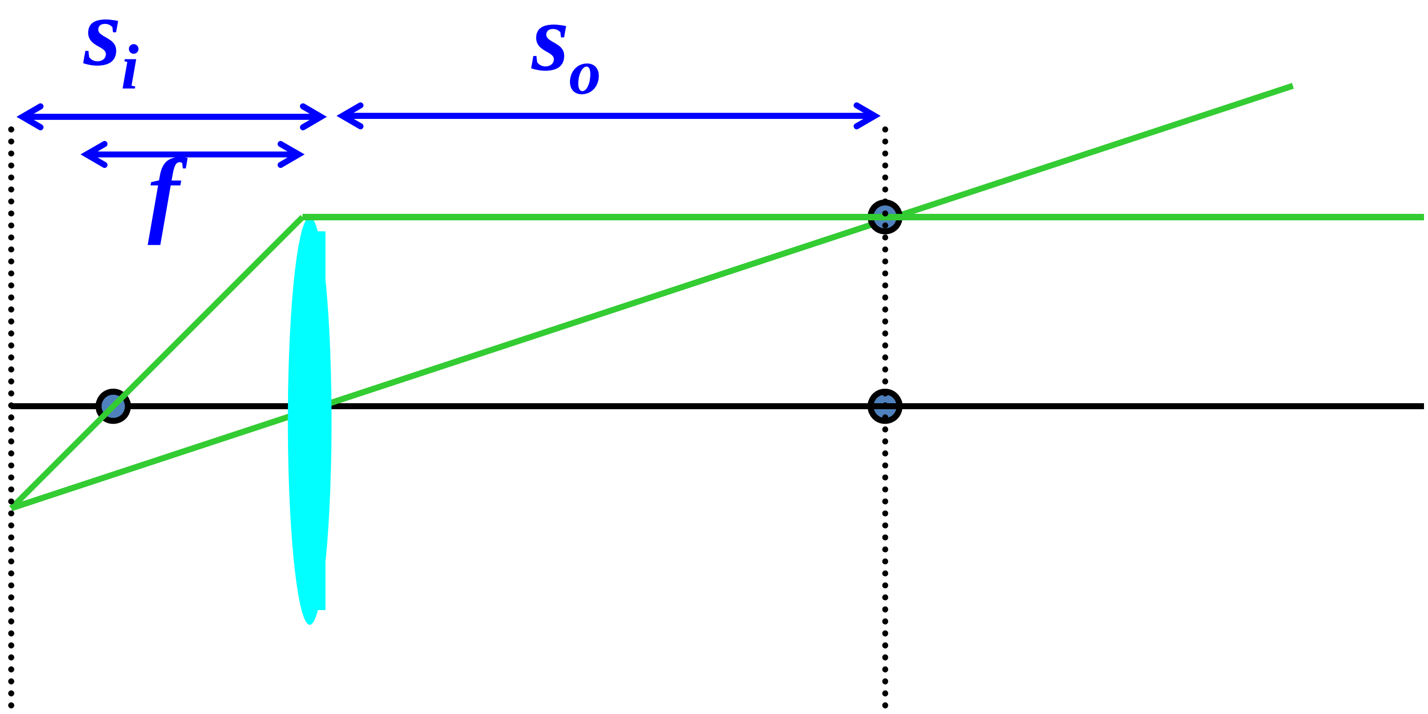


How to trace rays

- Start by rays through the center
- Choose focal length, trace parallels
- You get the focus plane for a given scene plane
  - All rays coming from points on a plane parallel to the lens are focused on another plane parallel to the lens

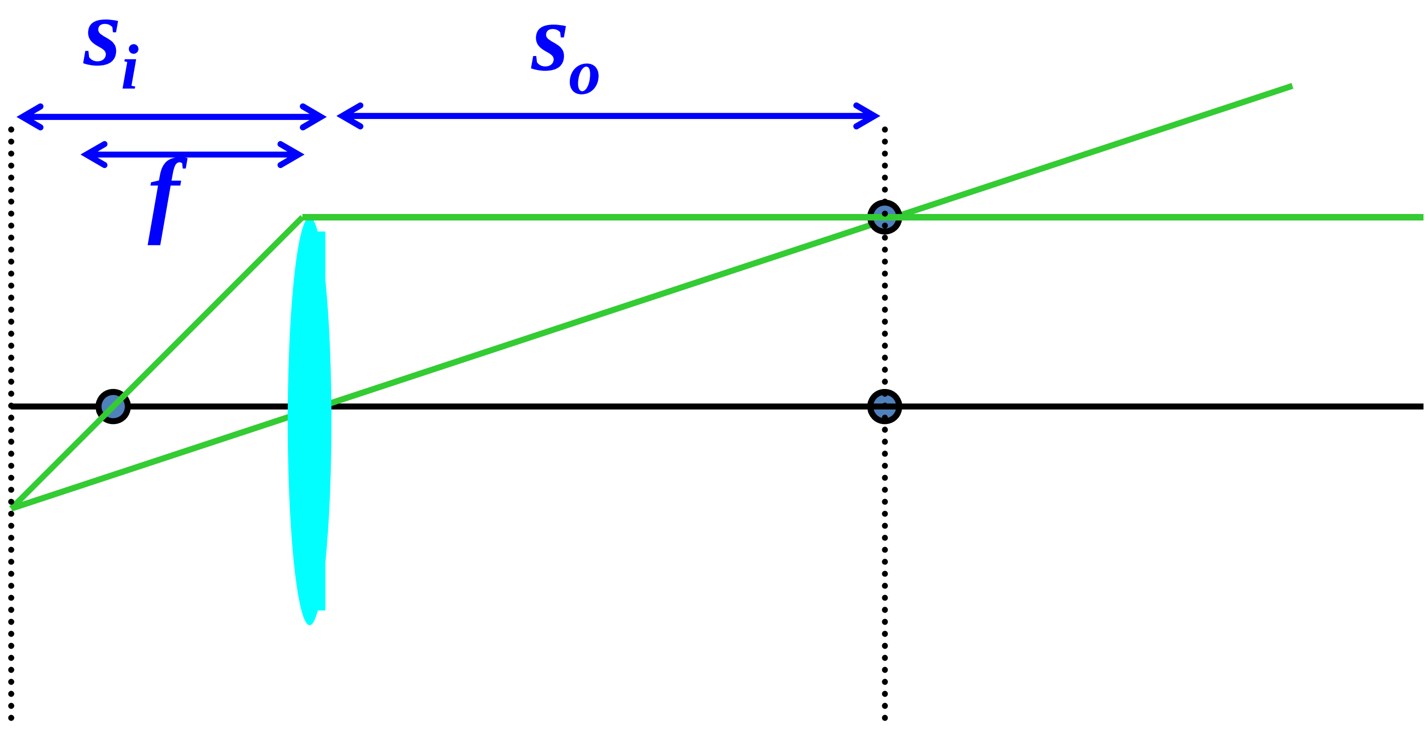


Thin lens formula



Thin lens formula

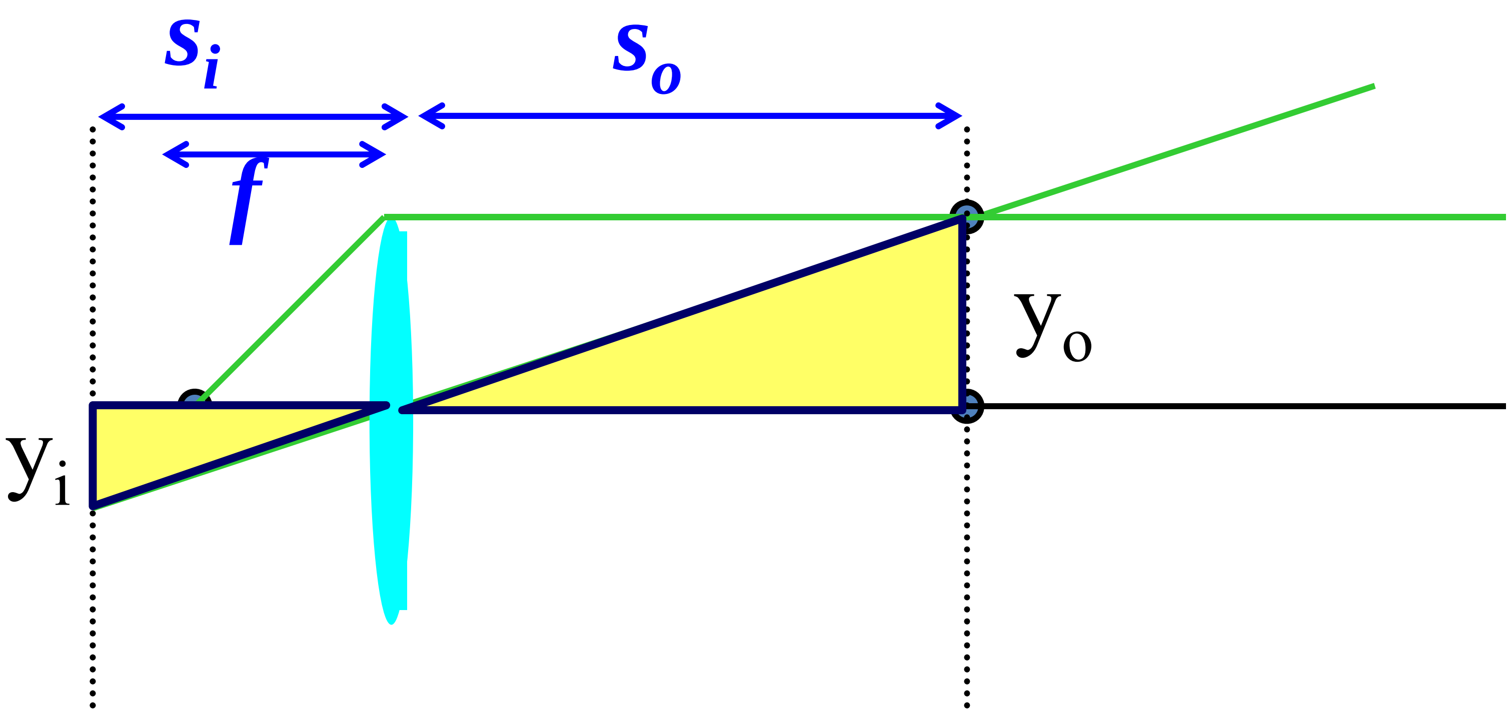
Similar triangles everywhere!



Thin lens formula

Similar triangles everywhere!

$$y_i/y_o = s_i/s_o$$

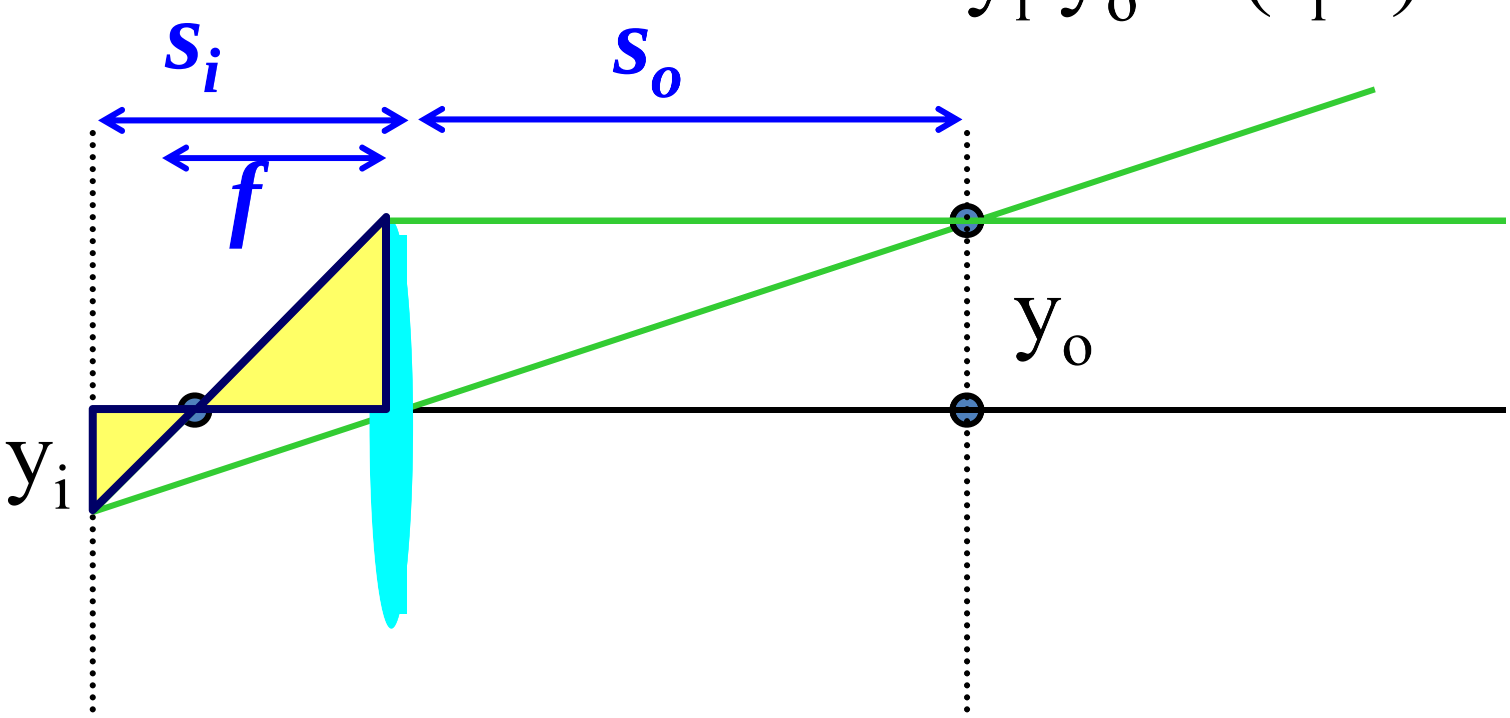


Thin lens formula

Similar triangles everywhere!

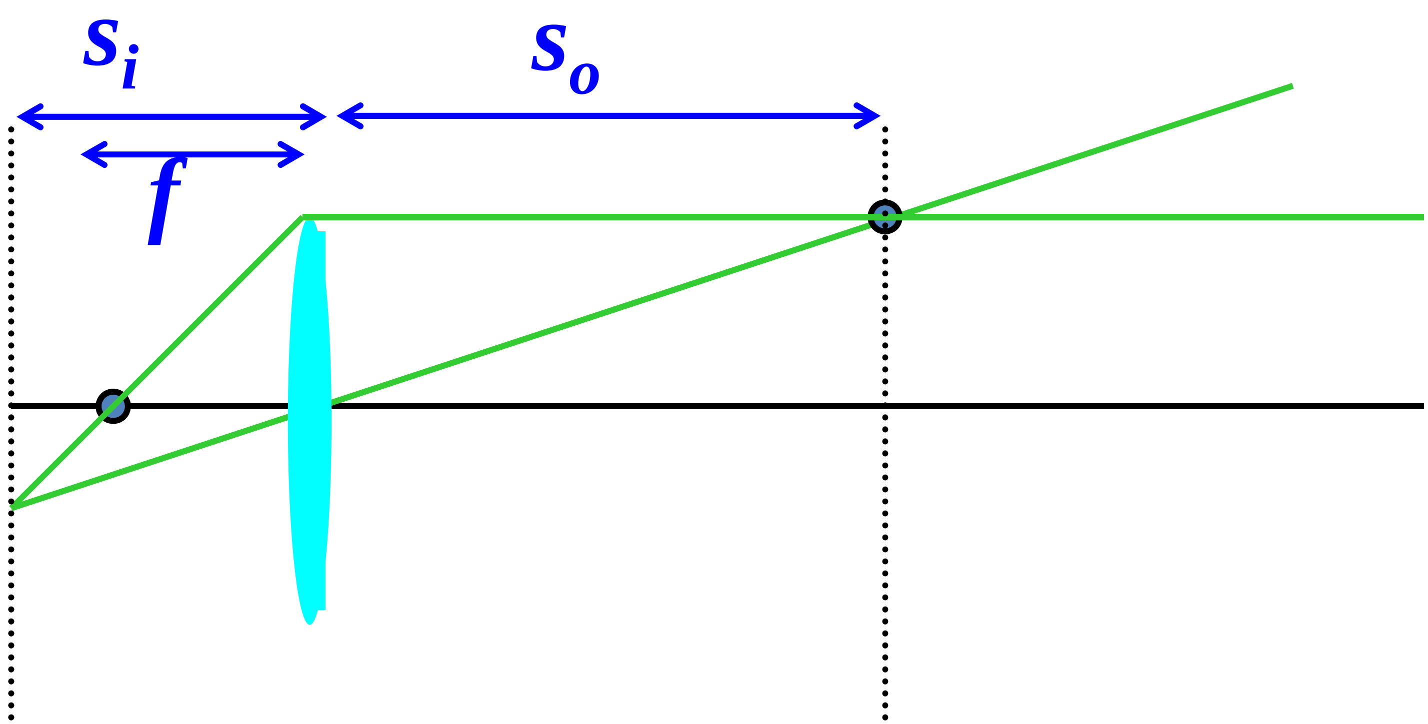
$$y_i/y_o = s_i/s_o$$

$$y_i/y_o = (s_i\text{-}f)/f$$



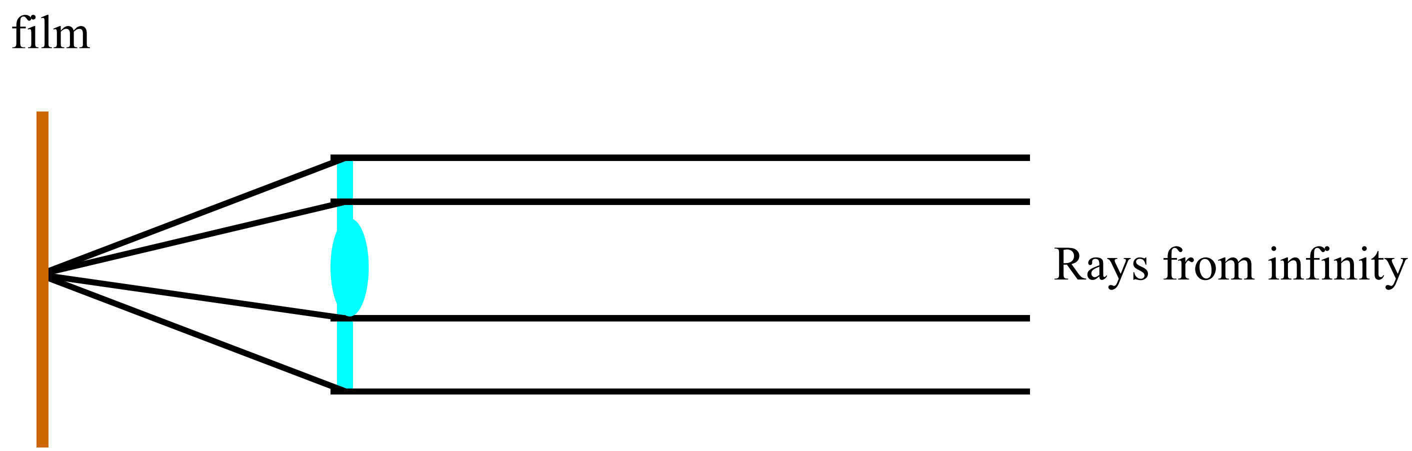
Thin lens formula

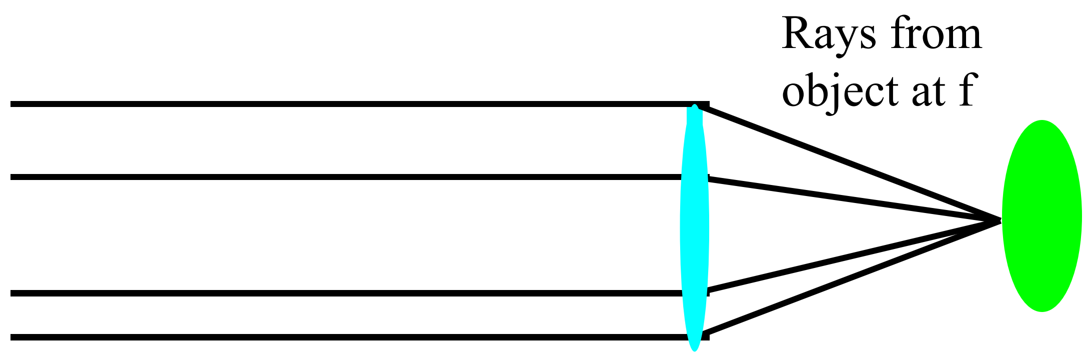
$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f}$$



Minimum focusing distance

- By symmetry, an object at the focal length requires the film to be at infinity.

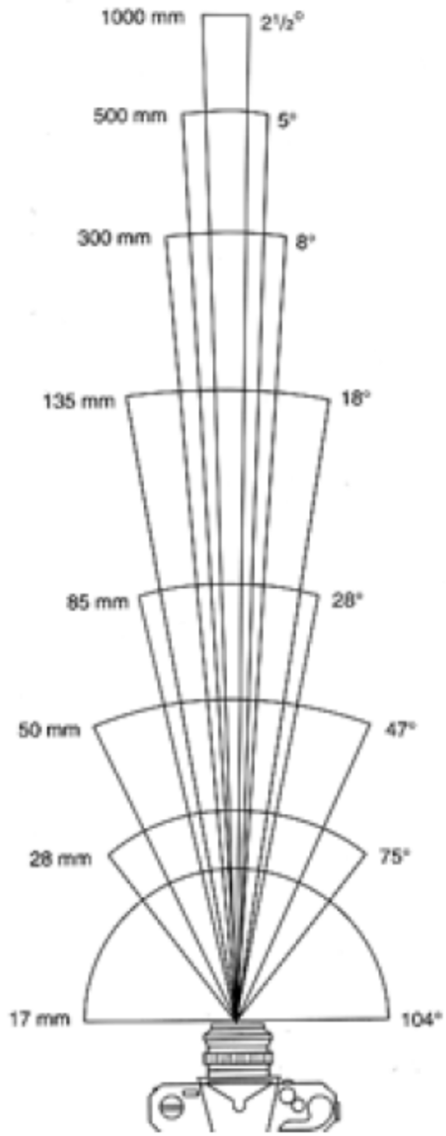




Camera Body: Optics

- Pinhole camera model
- Focusing light: Thin lens model
- Field of view (zoom)

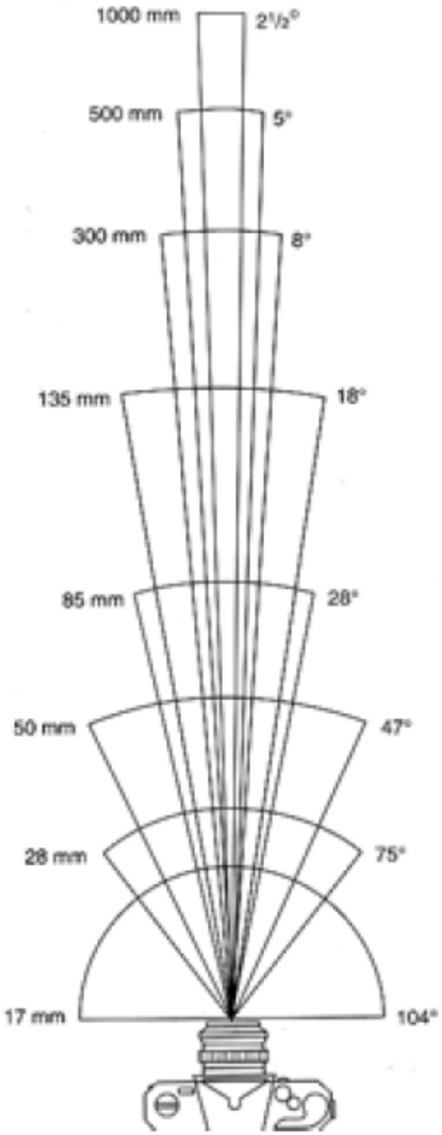
Field of View (Zoom)

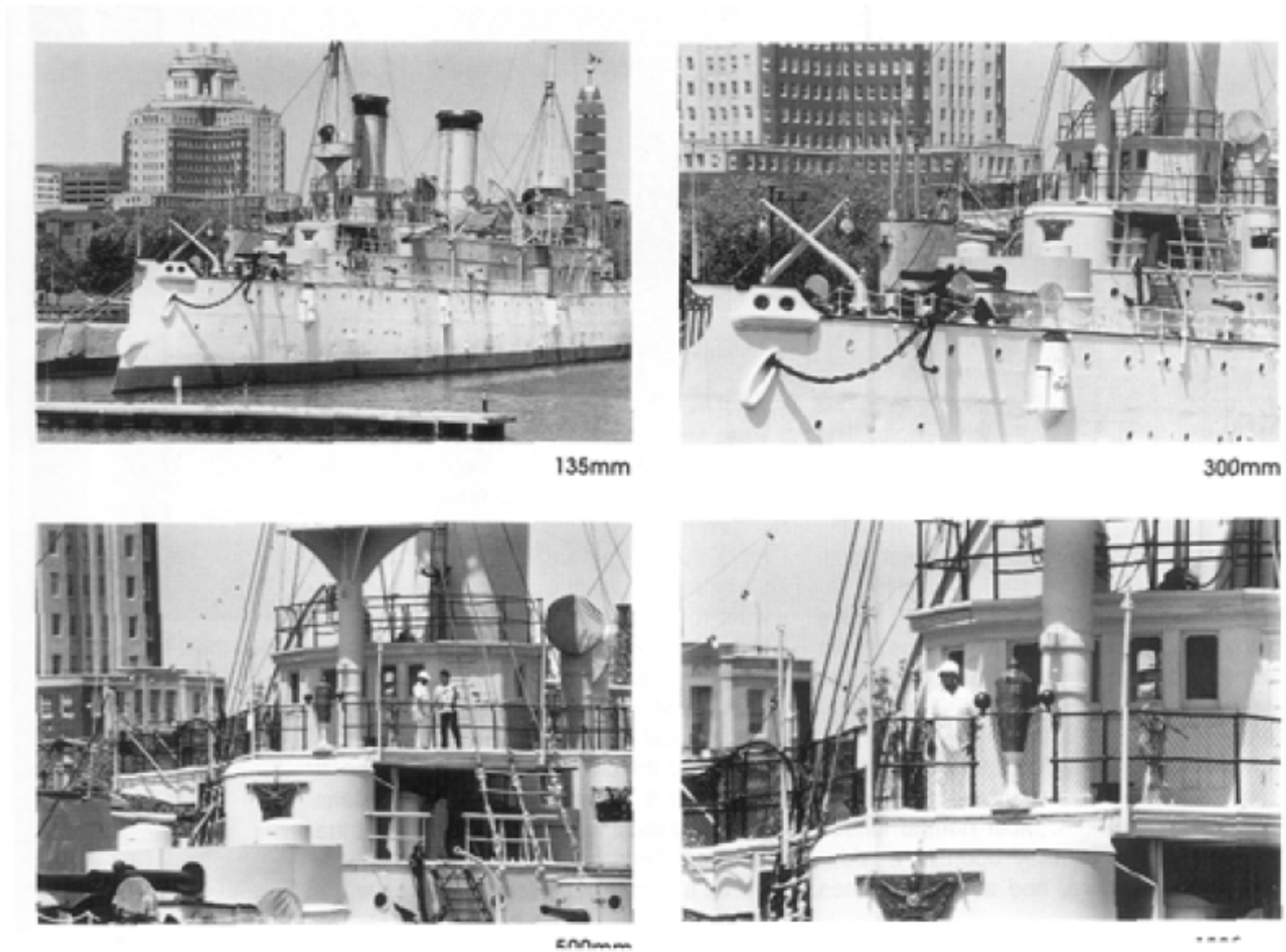


FOV measured diagonally on a 35mm full-frame camera (24 × 36mm)

From London and Upton

Field of View (Zoom)

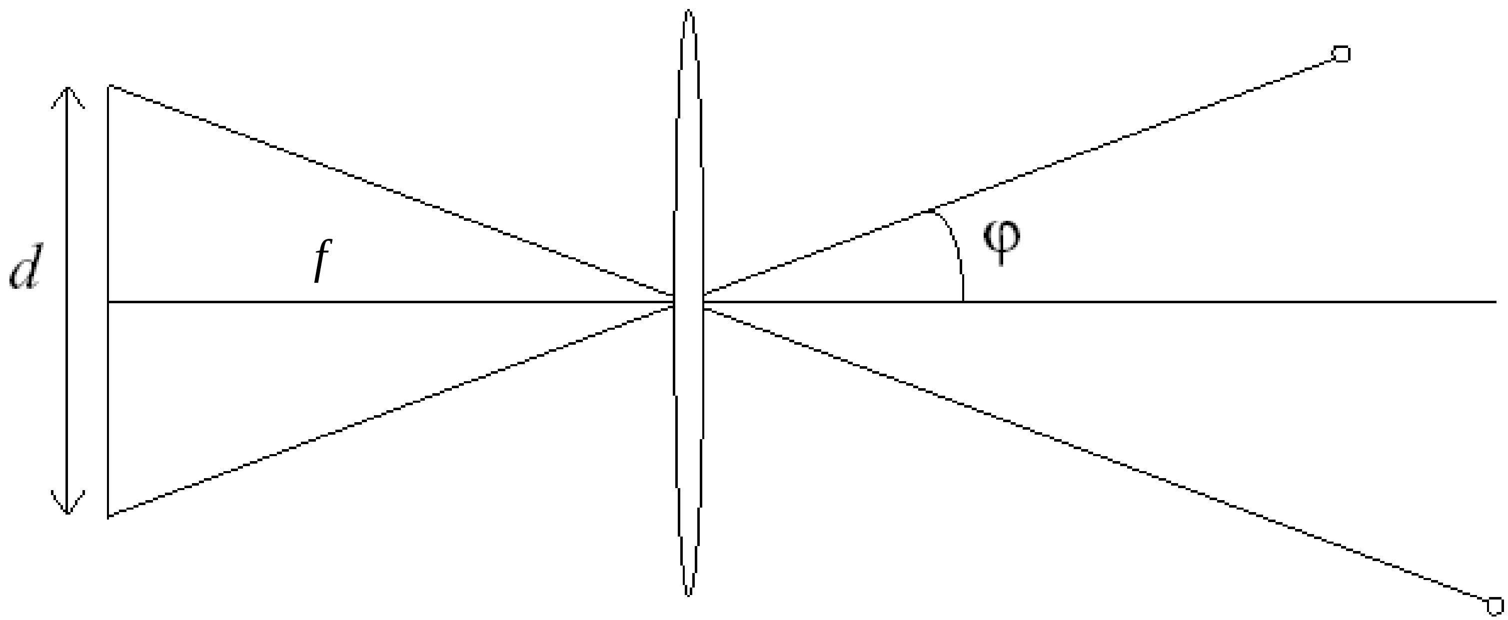




From London and Upton

FOV measured diagonally on a 35mm full-frame camera (24 × 36mm)

FOV depends on Focal Length



Size of field of view governed by size of the camera retina:

$$\varphi = \tan^{-1}\left(\frac{d}{2f}\right)$$

FOV = 2 x ϕ (in degrees) Smaller FOV = larger Focal Length

Field of View vs. Focal Length

Large FOV, small f
Camera close to car

Small FOV, large f
Camera far from the car

Field of View vs. Focal Length

- Portrait: distortion with wide angle
- Why?

Wide angle

Standard

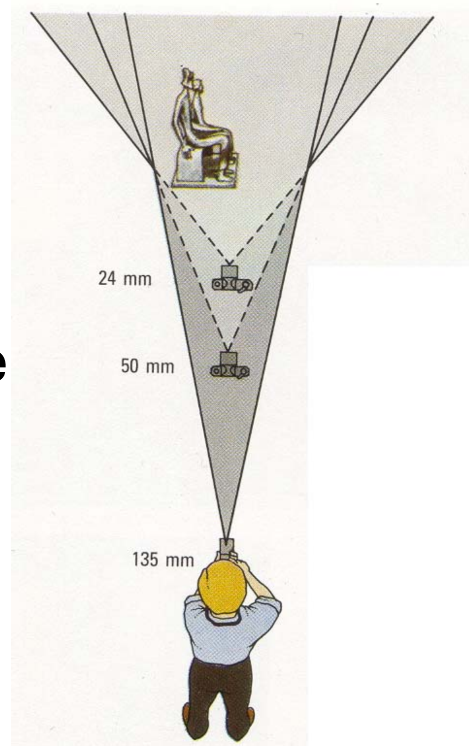
Telephoto

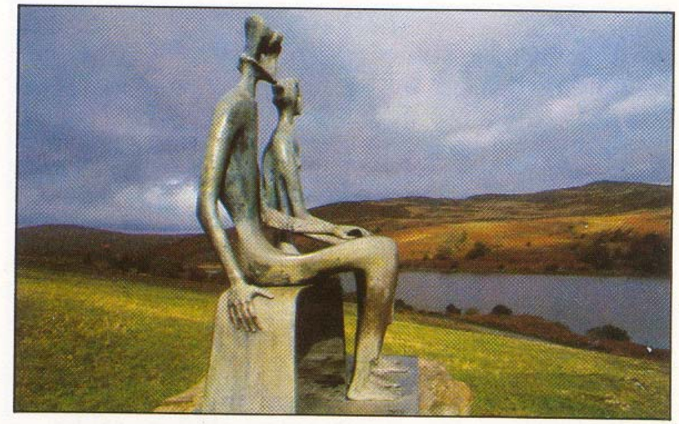
Field of View vs. Focal Length

https://photographylife.com/what-is-distortion

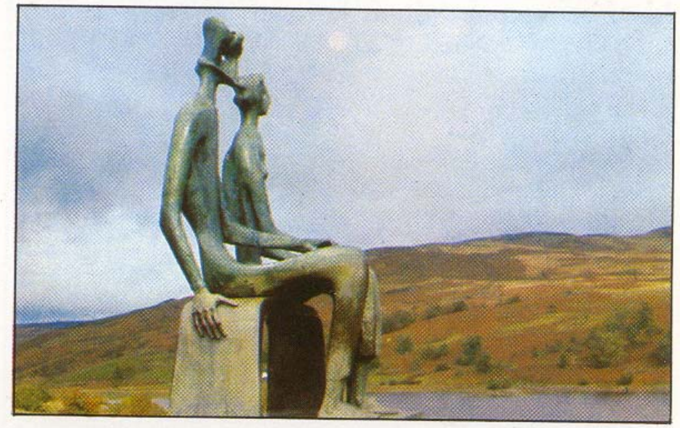
Changing the focal length vs. changing the viewpoint

- Telephoto makes it easier to select background (a small change in viewpoint is a big change in background)
  - changing the focal length lets us move back from a subject, while maintaining its size on the image
  - but moving back changes perspective relationships





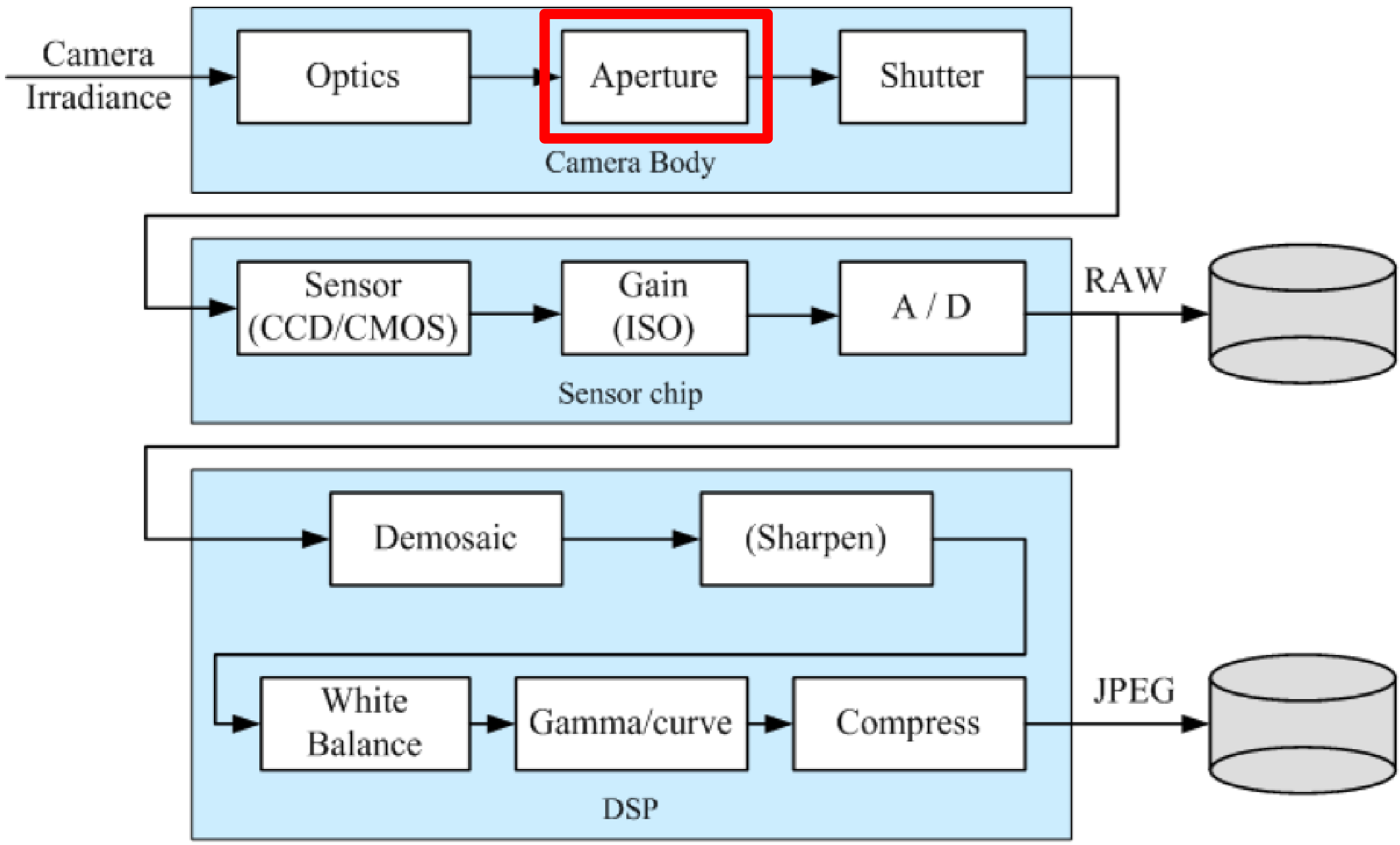
Grand-angulaire 24 mm



Normal 50 mm

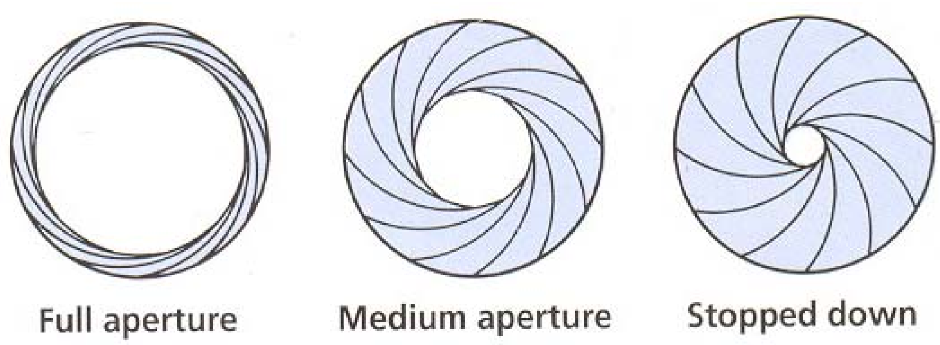
Longue focale 135 mm

Camera Body: Aperture

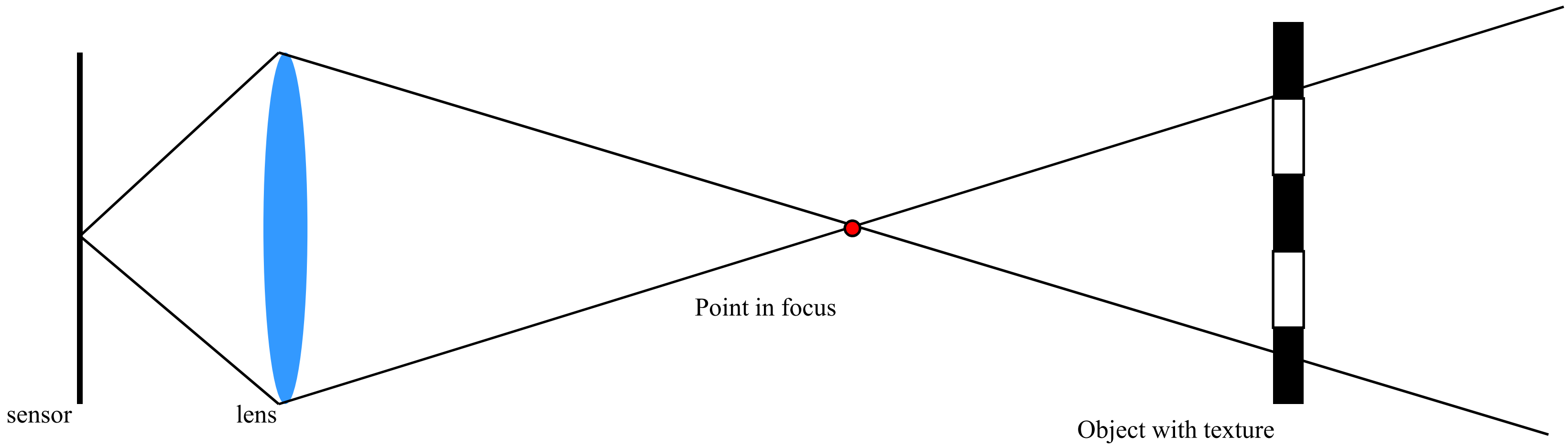


Aperture

- Diameter of the lens opening (controlled by diaphragm)
- Expressed as a fraction of focal length, in ***f-number*** ***N***
  - f/2.0 on a 50mm lens means that the aperture is 25mm
  - f/2.0 on a 100mm lens means that the aperture is 50mm
- Confusing: small f-number = big aperture
- What happens to the area of the aperture when going from f/2.0 to f/4.0?
- Typical f-numbers are (each of them counts as one f/stop) f/2.0, f/2.8, f/4, f/5.6, f/8, f/11, f/16, f/22, f/32
  - See the pattern?

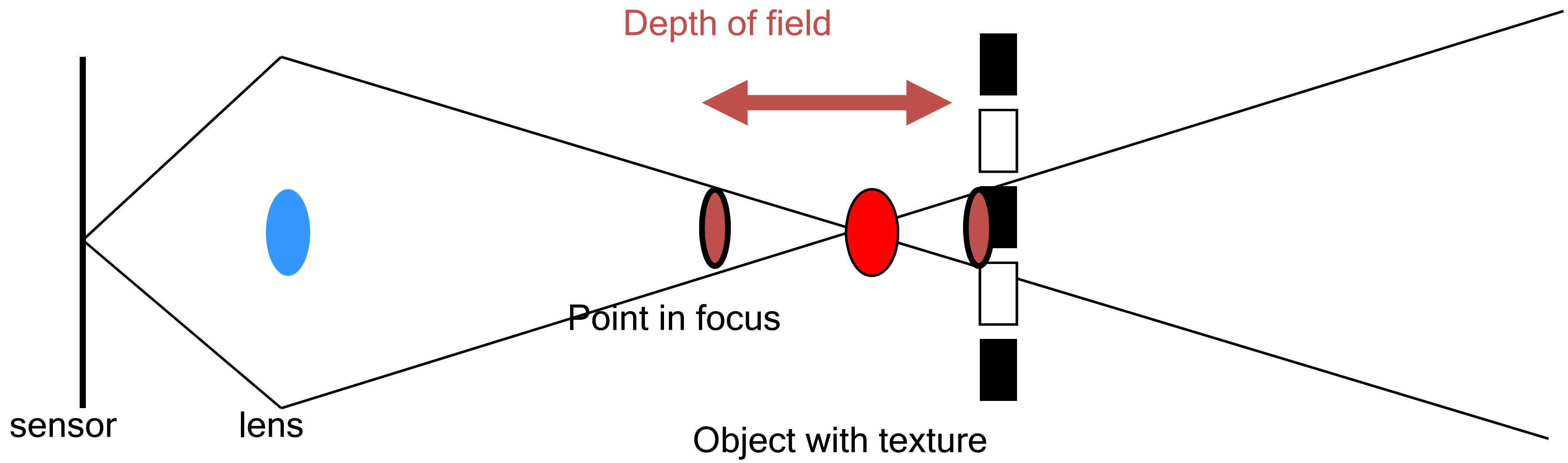


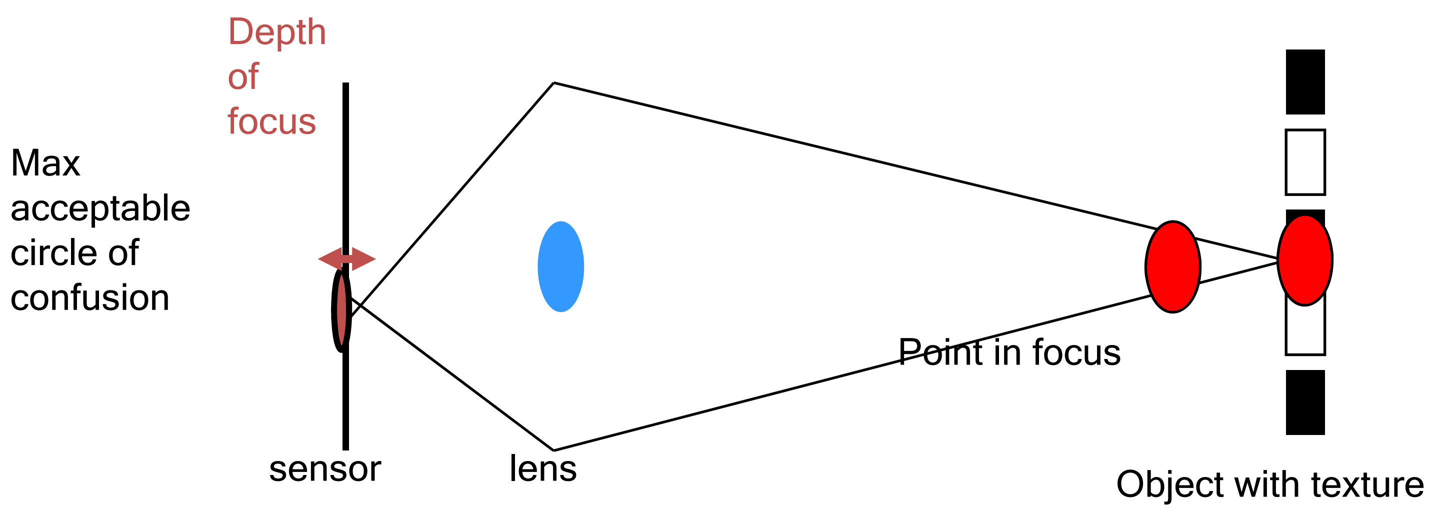
Depth of field



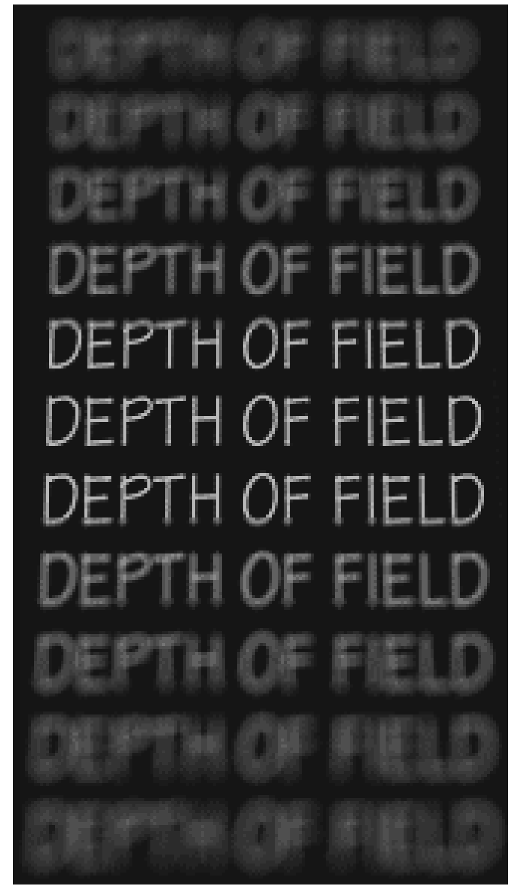
Depth of field

- We allow for some tolerance





Depth of Field

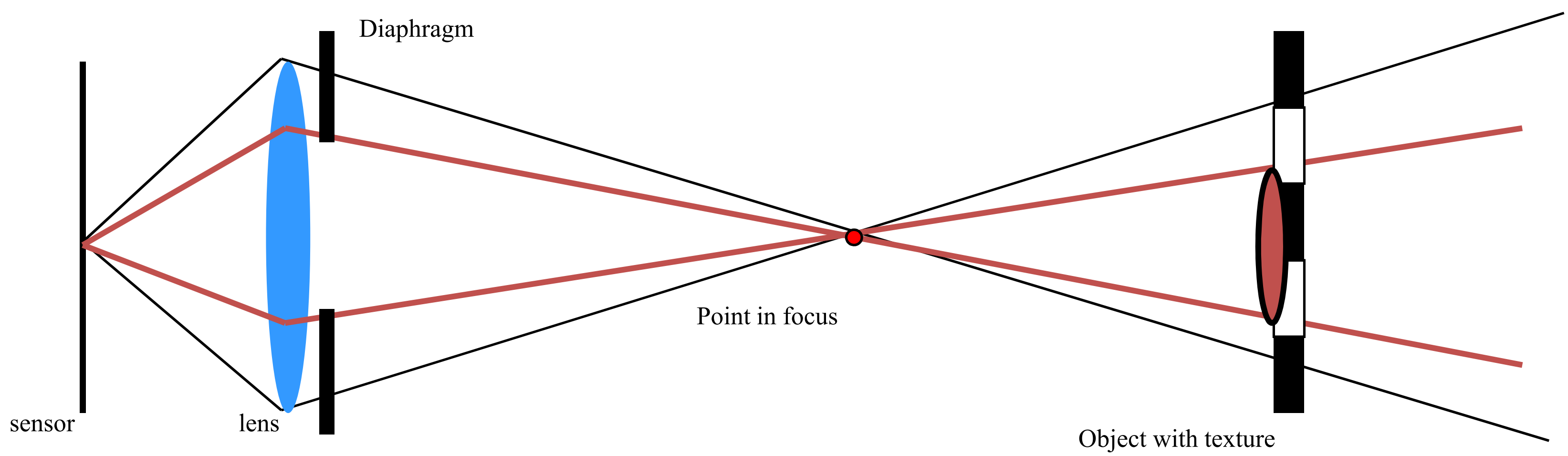


http://www.cambridgeincolour.com/tutorials/depth-of-field.htm

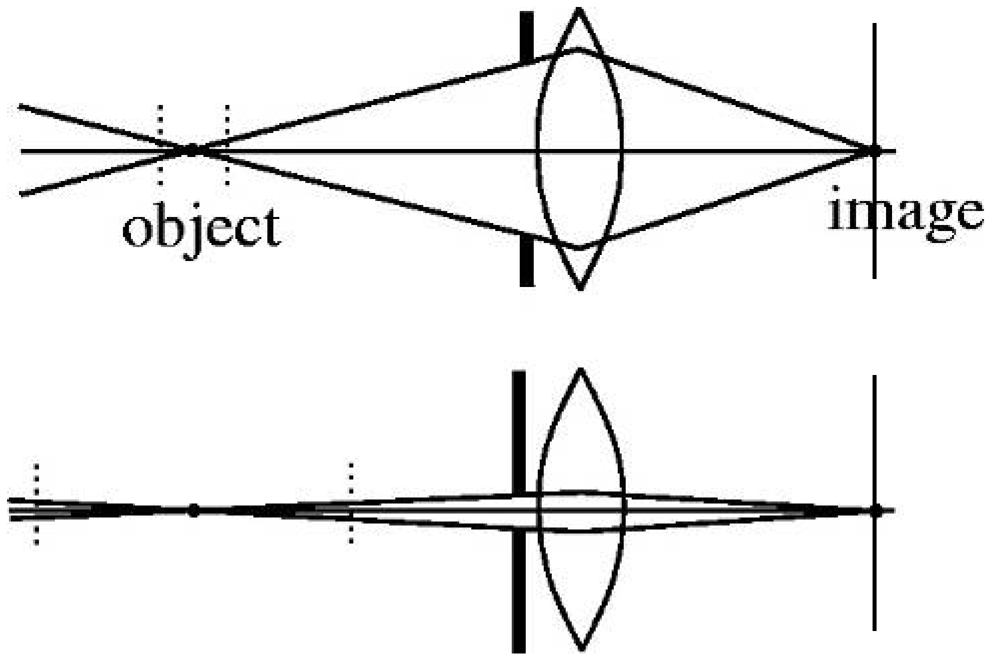
Nice Depth of Field Effect

Aperture controls Depth of Field

- What happens when we close the aperture by two stops?
  - Aperture diameter is divided by two
  - Depth of field is doubled



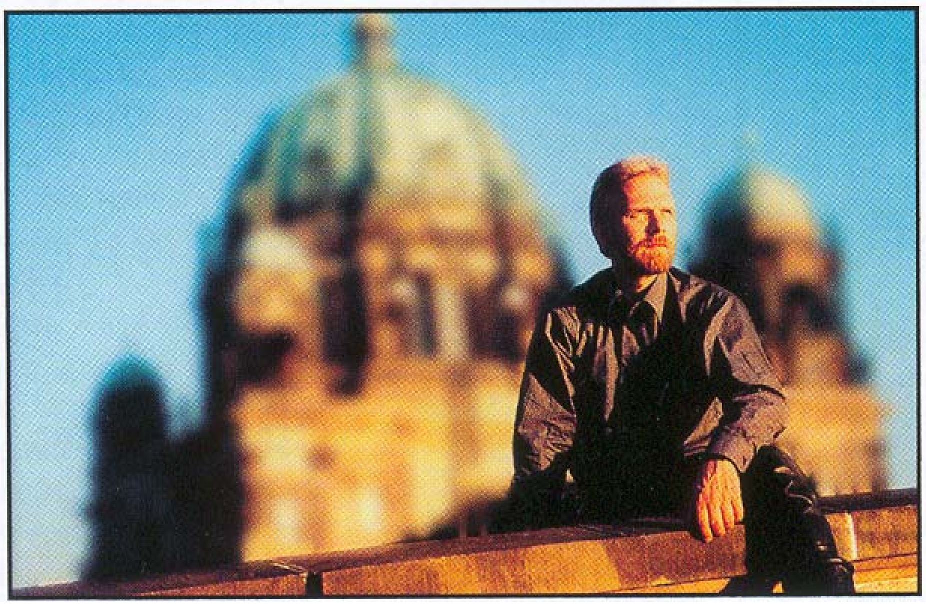
Aperture controls Depth of Field



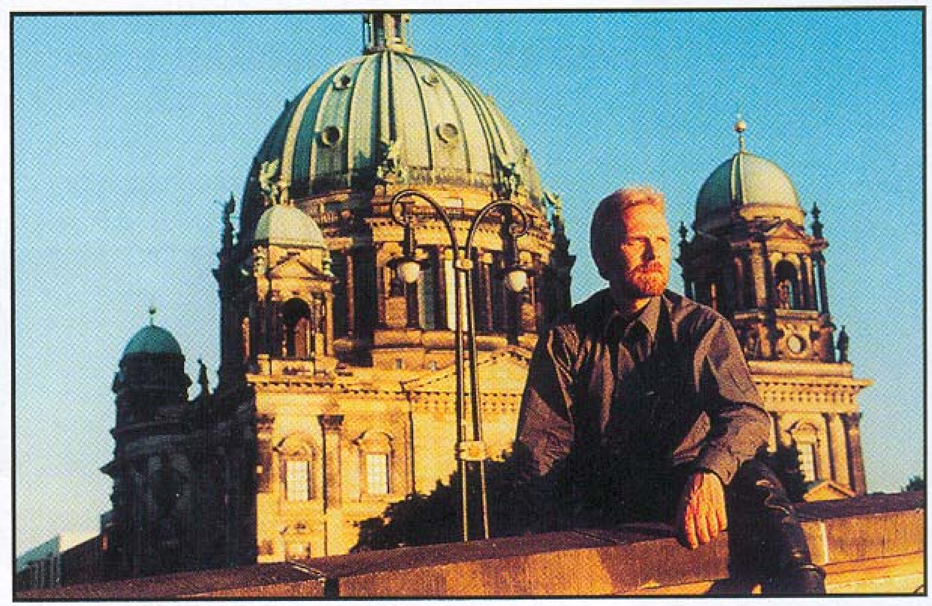
- A smaller aperture increases the range in which the object is approximately in focus
- But small aperture reduces amount of light - need to increase exposure

Aperture controls Depth of Field

Large aperture opening

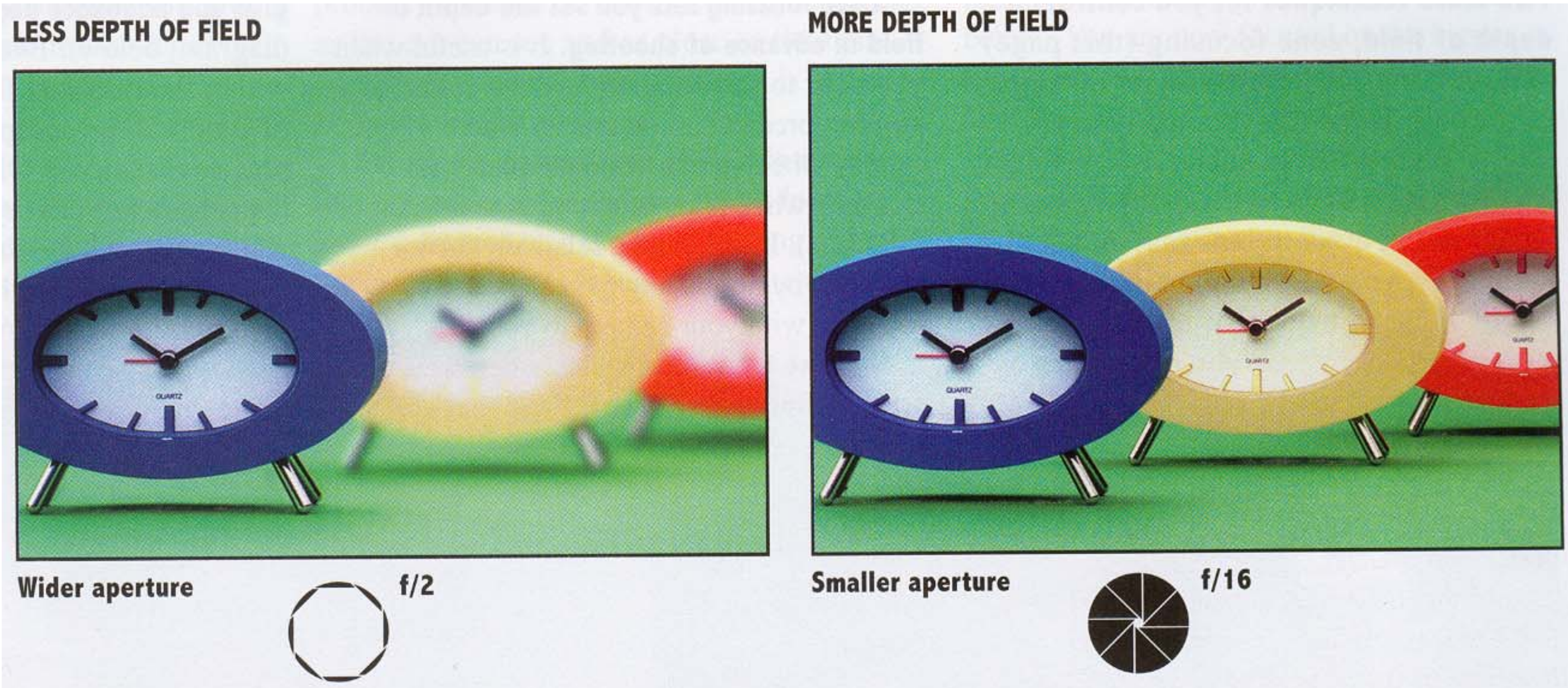


Small aperture opening



From Photography, London et al.

Depth of field



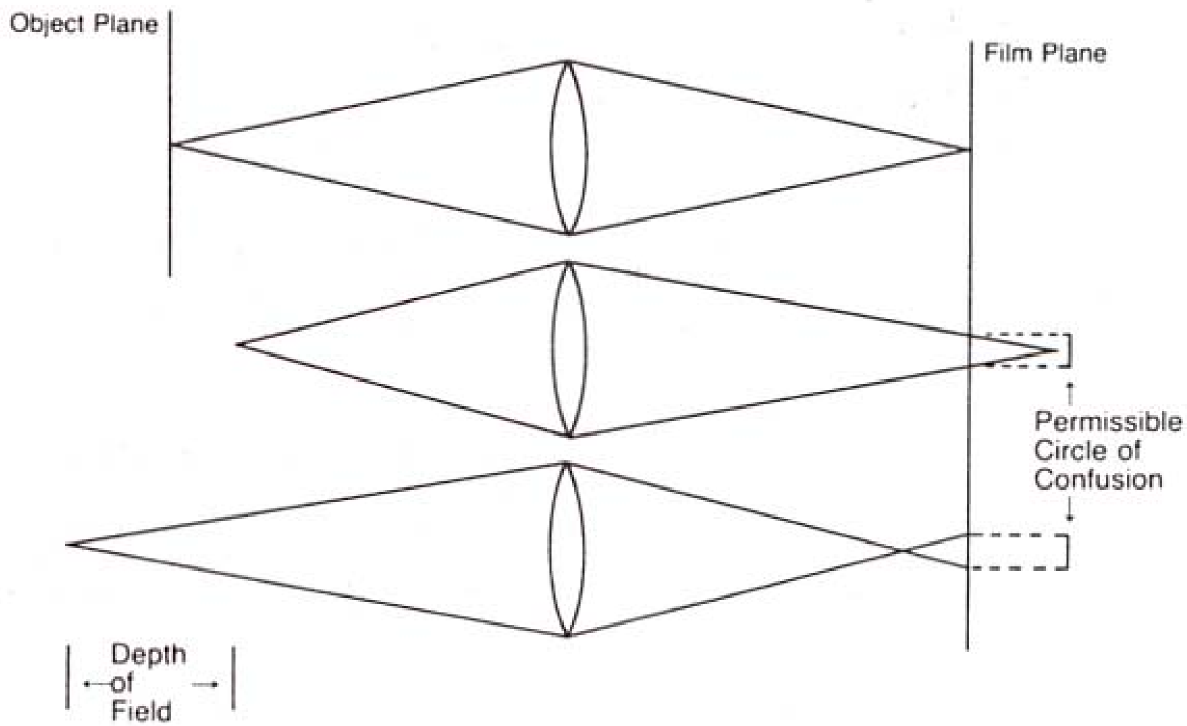
From Photography, London et al.

Depth of field

Large apeture = small DOF

Small apeture = large DOF

Circle of confusion (C)



- C depends on sensing medium, reproduction medium, viewing distance, human vision,...
  - for print from 35mm film, 0.02mm is typical
  - for high-end DSLR, 6µ is typical (1 pixel)
  - larger if downsizing for web, or lens is poor


N = f/6.3
C = 2.5µ
U = 17m (56')
f = 27mm
D_TOT = 12.5m (41')
1 pixel on this video projector
C = 2.5µ x 2816 / 1024 pixels
D_EFF = 34m (113')

Slide Credits

- This set of sides also contains contributions kindly made available by the following authors
  - Derek Hoiem
  - Gianfranco Doretto
  - Steve Seitz
  - John Oliensis
  - Alexei Efros
  - Frédo Durand
  - Steve Marschner
  - Marc Levoy