COMPUTER VISION

LECTURE 1 4.9.2019

Laura Ruotsalainen, Associate Professor
Department of Computer Science

TODAY'S LECTURE

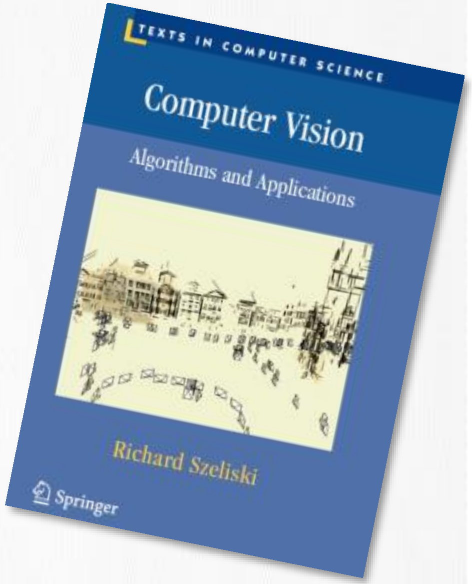
- Course practicalities and contents
- Introduction to Computer Vision
- Basics of Image Understanding

LEARNING OUTCOMES

- During this course you'll learn to
  - recognize and explain the important terms related to computer vision
  - retrieve the mathematical backgrounds for computer vision
  - process images to be in the best form for your algorithms
  - know and use the theoretical basis and most important algorithms for computer vision
  - understand the theory and challenges related to two very important applications where computer vision is used

MATERIALS USED 1

- Textbooks:
  - Rick Szeliski, *Computer Vision: Algorithms and Applications* online at: http://szeliski.org/Book/
  - Hartley, Zissermann (2003). Multiple View Geometry in Computer Vision, 2nd edition (electronic version in Helka)
- Additional reading:
  - Forsyth, Ponce (2013). Computer Vision, a Modern Approach, 2nd edition
- Publications with references given in slides



MATERIALS USED 2

- Some slides are reused from the following sources and clearly marked when and from whom
  - Carnegie Mellon University’s CV course http://www.cs.cmu.edu/~16385/
  - Cornell’s CV course http://www.cs.cornell.edu/courses/cs5670/2019sp/
  - MIT’s Advances in Computer Vision course http://6.869.csail.mit.edu/fa18/materials.html
- MIT’s course page listes also best **CV conferences** and **image data sets** available for use

COURSE CONTENT AND SCHEDULE

- Lectures 4.9. - 18.10. on Wednesdays at 12-14, D123 and Fridays 12-14, C123
- Exercises 11.9. - 16.10. on Wednesdays at 16-18, C122
- Exam 23.10. at 9.00 (27.11.)

Weeks 1-2

Low-level vision

Image formation
Geometry
Image processing
Feature detection

Szeliski 1,2,3,4

Weeks 3-5

Algorithms

Model estimation
3D to 2D
Stereopsis
Optical flow

Szeliski 6, 7, 8, (9),11

Weeks 6-7

Important applications

Structure from motion
Simultaneous Localization and Mapping (SLAM)

Convolutional Neural Networks for segmentation and recognition

Szeliski 14+ given material

TASKS AND GRADING

- Participation at the lectures, **except at the last one**, is not mandatory and will not be awarded. However, it’s highly recommended as questions and discussions are encouraged.
- Exercises: 6 times, 2-4 tasks each,
- Exercise 6 will include a group work where a paper is read and presented to the rest of the class at the last lecture. Please discuss asap with the teacher in case you have some challenges attending the last lecture
- Exam: Done with paper and pen
- Points given from exam = 48p, exercises = 48p, total 96p
- Grade 1 = 47p, 5 = 87 p, middle grades with 10p scale

COURSE REQUIREMENTS

- Prerequisites—*these are essential*!
  - Data structures
  - Linear algebra
  - Vector calculus
- These you'll also need
  - A good working knowledge of Matlab (C++ or Python programming)
  - Deep learning, especially Convolutional Neural Networks
- Course does ***not*** assume prior imaging experience

  computer vision, image processing, graphics, etc.

EXERCISES

- 6 times, 2-4 tasks each
- Some mathematical tasks but mainly implementing computer vision methods
- Exercises are planned to be done with Matlab, but they may be done using C++ or Python and OpenCV. However, it might mean that there will be a bit more to be implemented and detailed advise is not given about the code.
- 50 % of tasks have to be done, max 48 points
- Exercise report has to be **emailed** to the teacher on the ***exercise day latest at 15.00***
- Answers to the tasks will be discussed and given at the exercise class. Answers will be published at the Wiki page also, but personal advise will not be given

TEACHER

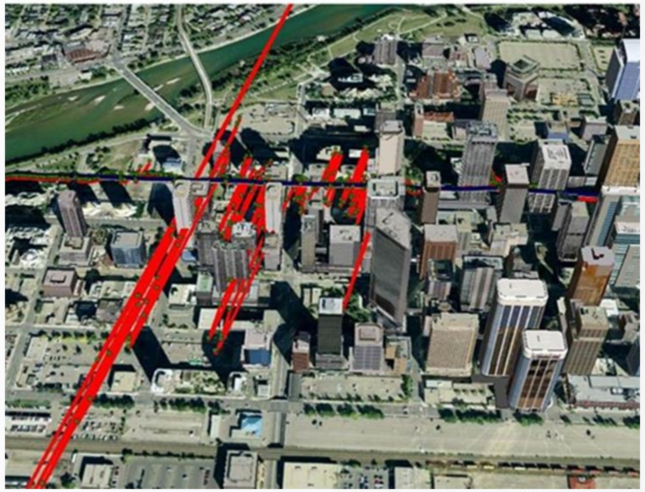
- Laura Ruotsalainen (Exactum B227, @helsinki.fi)
- Associate Professor in Spatiotemporal Data Analysis for Sustainabililty Science, since August 2018 https://www.helsinki.fi/en/researchgroups/spatiotemporal-data-analysis
    - Finnish Geospatial Research Institute, Department of Navigation and Positioning, 2010 – 2019, leader of the Sensors and Indoor Navigation research group
- PhD ”Vision-aided Pedestrian Navigation for Challenging GNSS Environments” 2013

SPATIOTEMPORAL DATA ANALYSIS AT CS

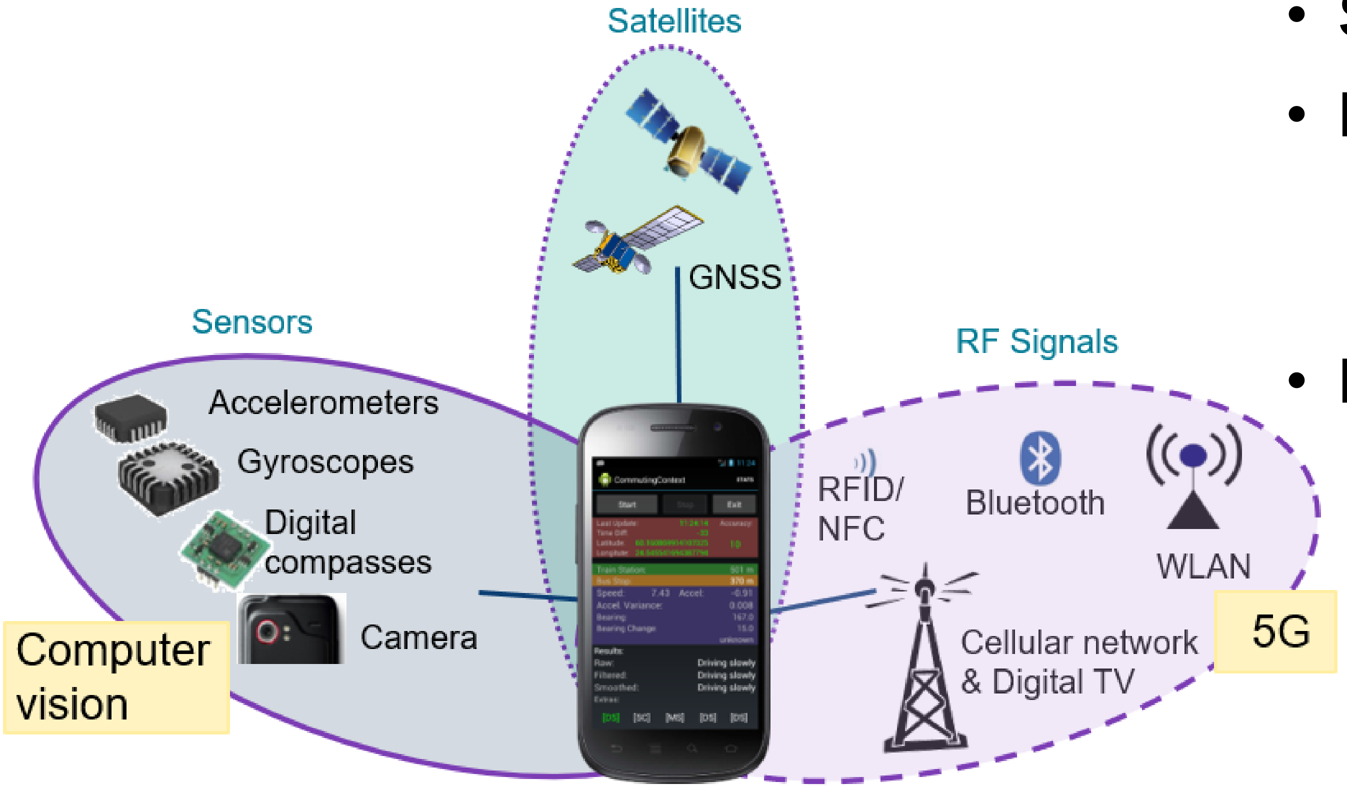
- Navigation in challenging environments (urban, indoors, Arctic)
- Machine learning and estimation
- Autonomous systems (Road transport, UAVs), pedestrians
- Computer vision methods for navigation and situational awareness
- Mitigation of intentional interference of satellite positioning

HELSINKI CENTRE FOR DATA SCIENCE

HELSINKI INSTITUTE OF SUSTAINABILITY SCIENCE



NAVIGATION IN CHALLENGING SITUATIONS



- **Statistical Error Modelling**
- **Recursive Bayesian Estimation**
  - Measurement fusion
  - Cooperative positioning
- **Machine Learning**
  - Recognizing environment and motion
  - Route prediction
  - Improving computer vision / radio signal processing

INTRODUCTION TO COMPUTER VISION

COMPUTER VISION

- Computer vision: algorithms retrieving understanding from the image
  - Computer vision is a subfield of AI
- Machine vision: use of computer vision in industrial or practical processes
- Photogrammetry: generating 2D or 3D model of the scene using images taken from different poses (or multiple cameras)

COMPUTER VISION

- Goal of the computer vision is to give computers human level perception
- Research started at the 70s, when this was anticipated to be a task for a student during one summer’s training
- 2018 investments for computer vision: North America 120 M$, China 3.9B$
- Field has been active for decades, but last 4 years have been hectic => Deep learning and especially Convolutional Neural Networks
  - Huge need for research and development both at academia and companies => great work opportunities

IMPROVING PHOTOS (COMPUTATIONAL PHOTOGRAPHY)

Super-resolution (source: 2d3)

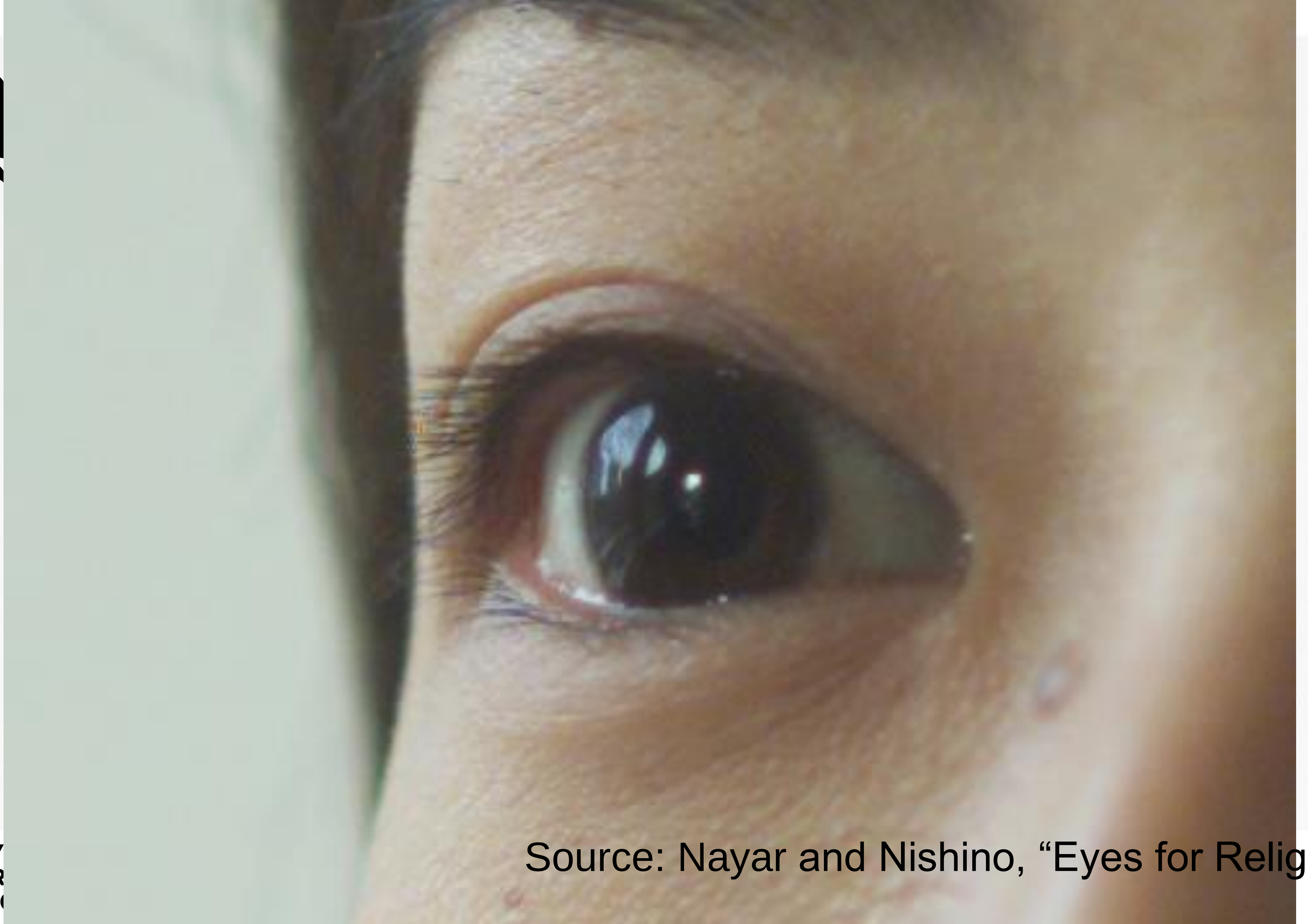
Low-light photography
(credit: Hasinoff et al., SIGGRAPH ASIA 2016)

Inpainting / image completion
(image credit: Hays and Efros)

FORENSICS

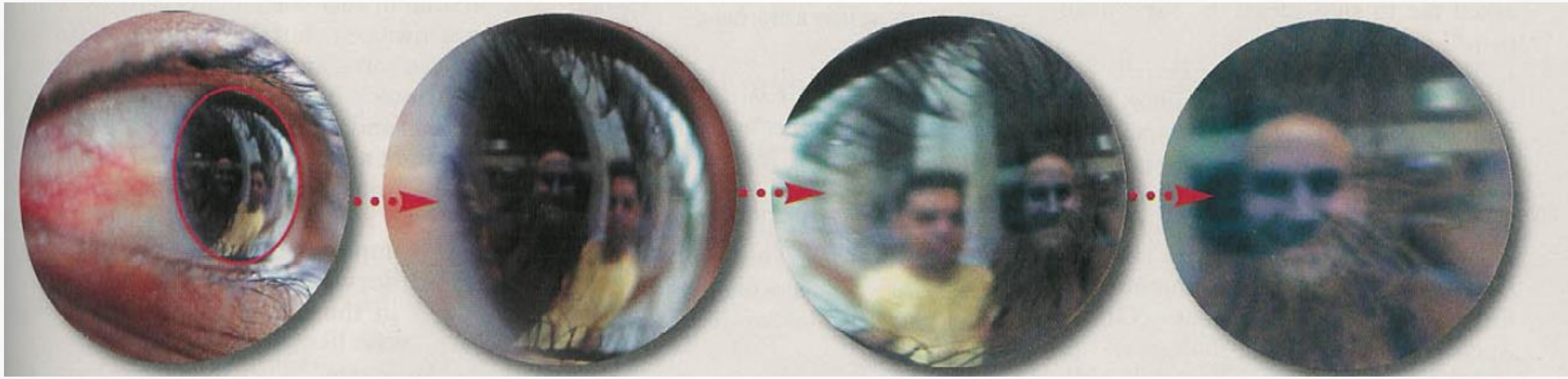
Forensics

Source: Nayar and Nishino, “Eyes for Relighting”



Source: Nayar and Nishino, “Eyes for Relighting”

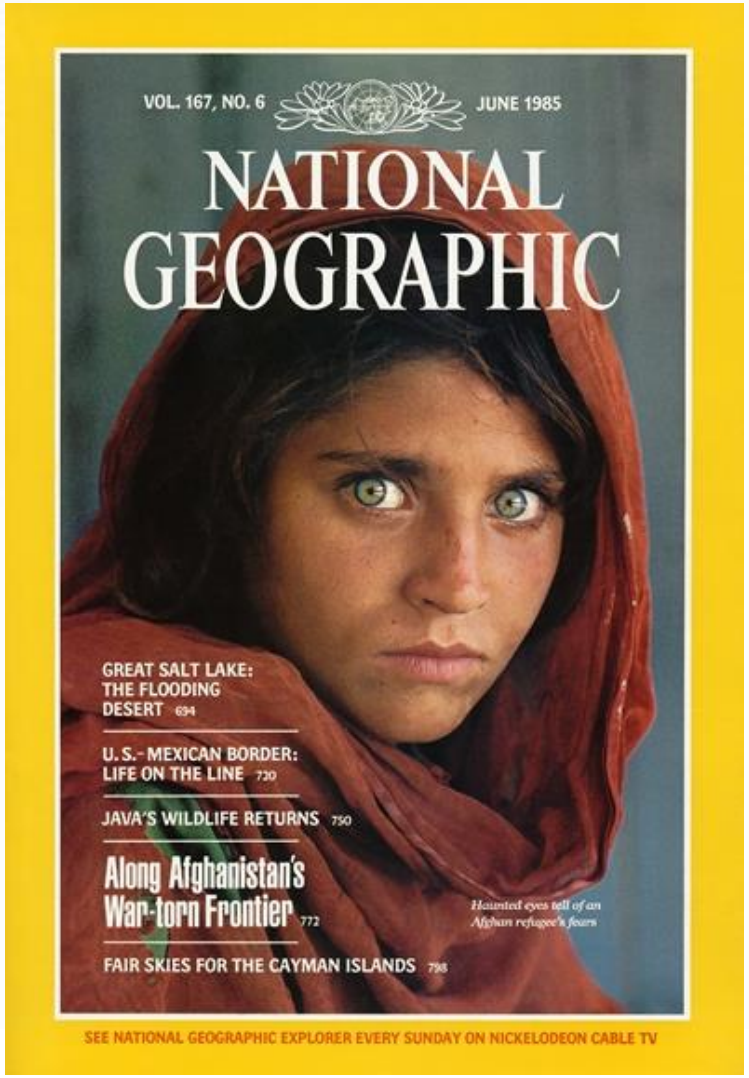
FORENSICS



Source: Nayar and Nishino, “Eyes for Relighting”

FACE RECOGNITION

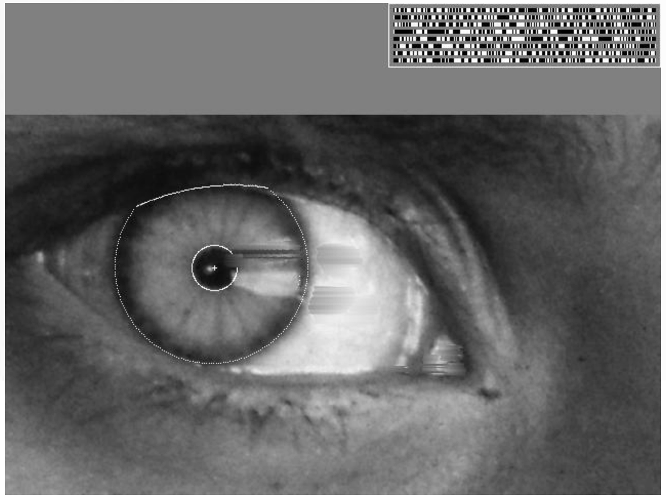
Who is she?

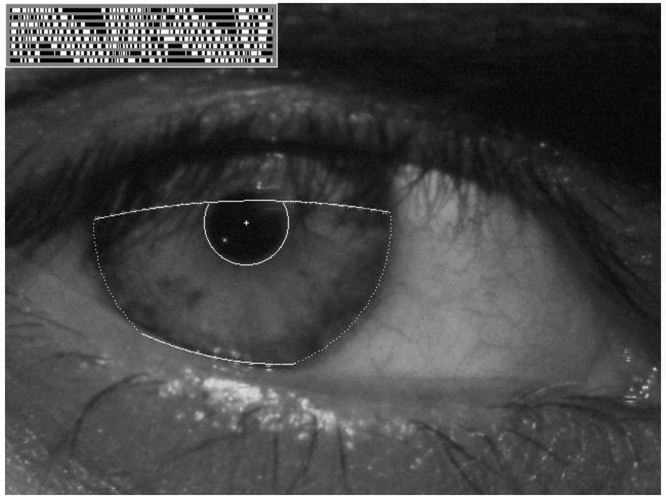


Source: S. Seitz

VISION-BASED BIOMETRICS

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



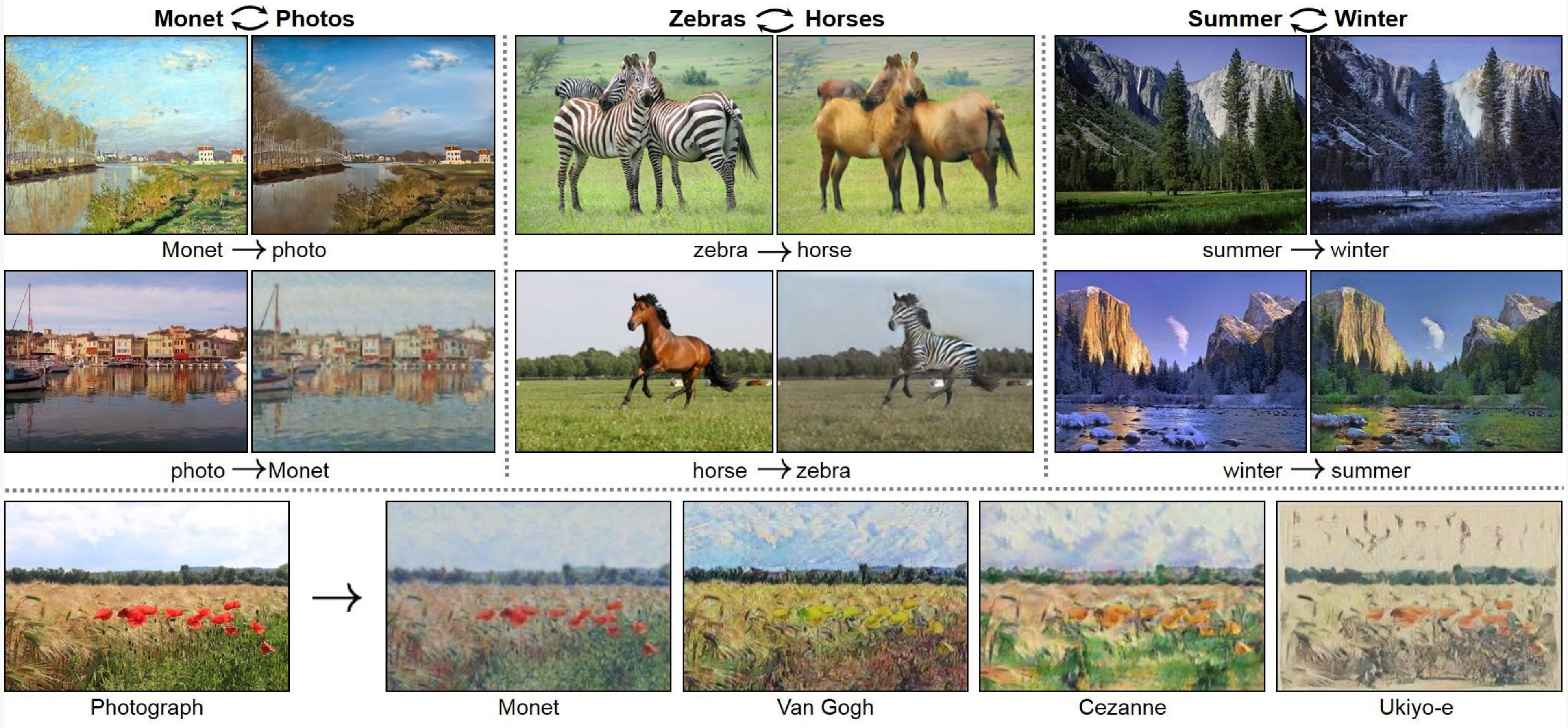


Source: S. Seitz

MORPHING

Morphing https://www.youtube.com/watch?v=wZurRt0TidI

IMAGE SYNTHESIS



Zhu, et al., *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*, ICCV 2017

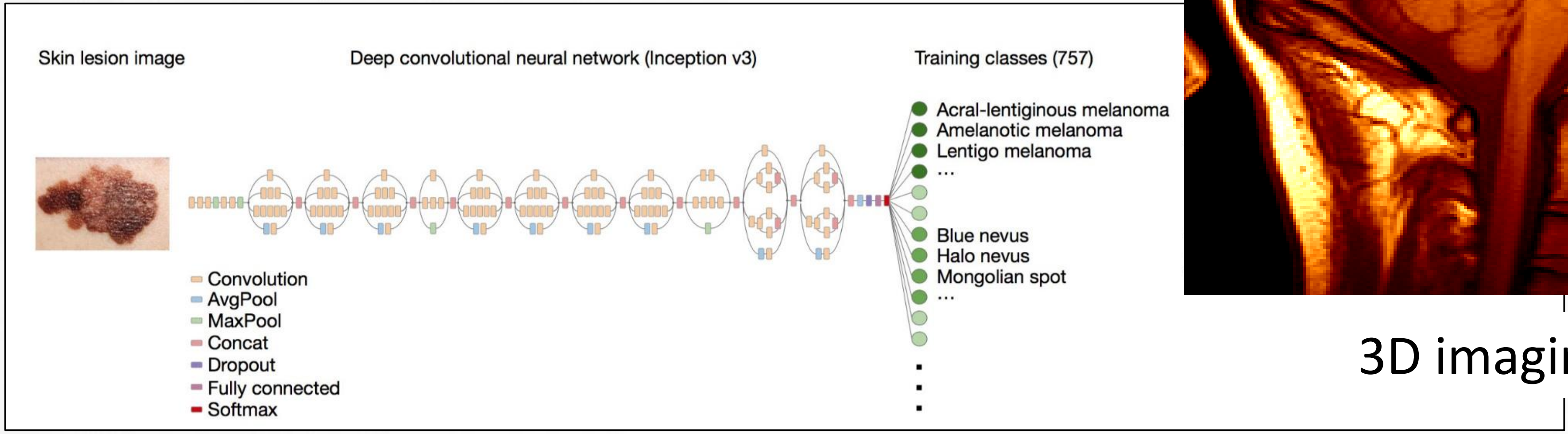
SPORTS

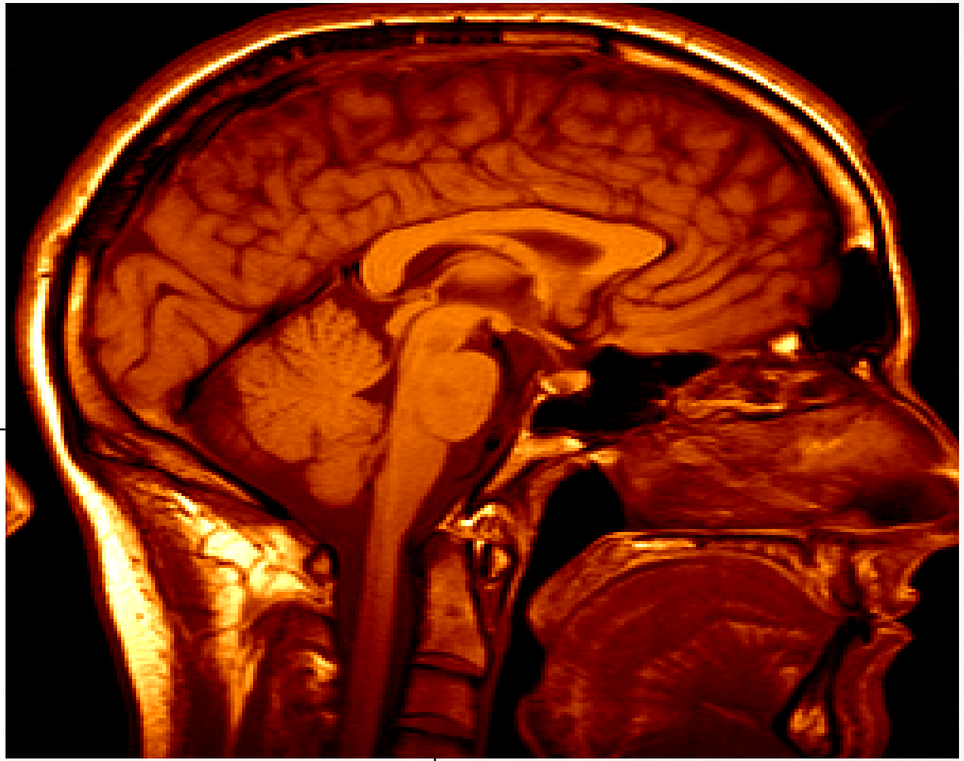
Sportvision first down line
Nice explanation on www.howstuffworks.com

Highlights of the men's 4x200m relay final on Day 5.

Source: S. Seitz

MEDICAL IMAGING





3D imaging (MRI, CT)

Skin cancer classification with deep learning

https://cs.stanford.edu/people/esteva/nature/

OPTICAL CHARACTER RECOGNITION (OCR)

- If you have a scanner, it probably came with OCR software

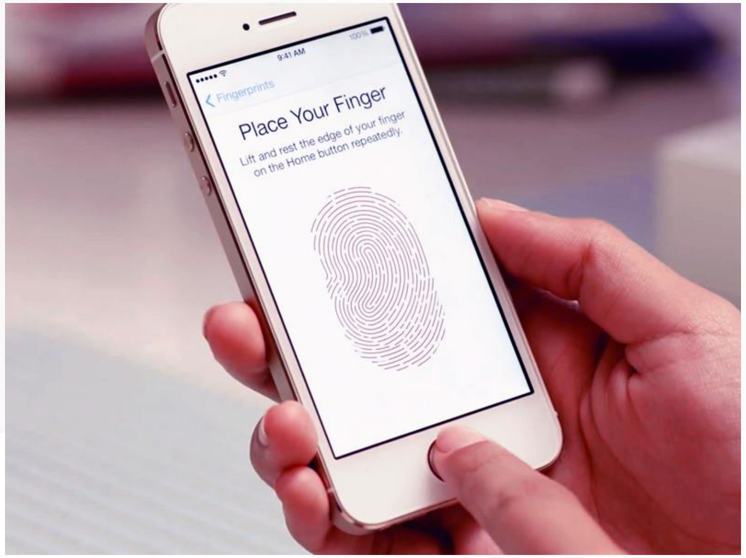
Digit recognition, AT&T labs (1990's)
http://yann.lecun.com/exdb/lenet/

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

FACE AND FINGERPRINT DETECTION

Nearly all cameras detect faces in real time

Face unlock on Apple iPhone X
See also http://www.sensiblevision.com/

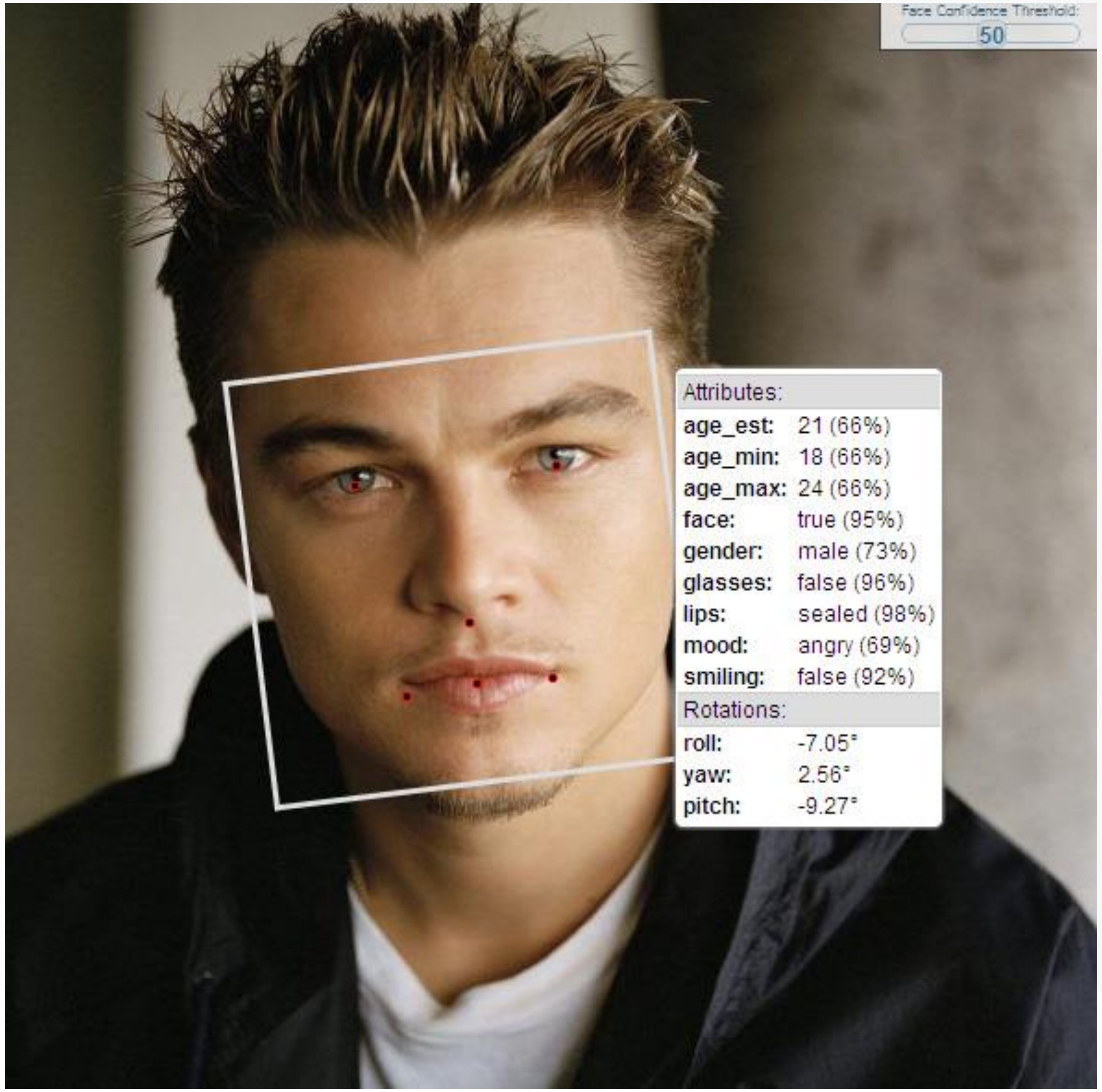


Fingerprint scanners on many new smartphones and other devices

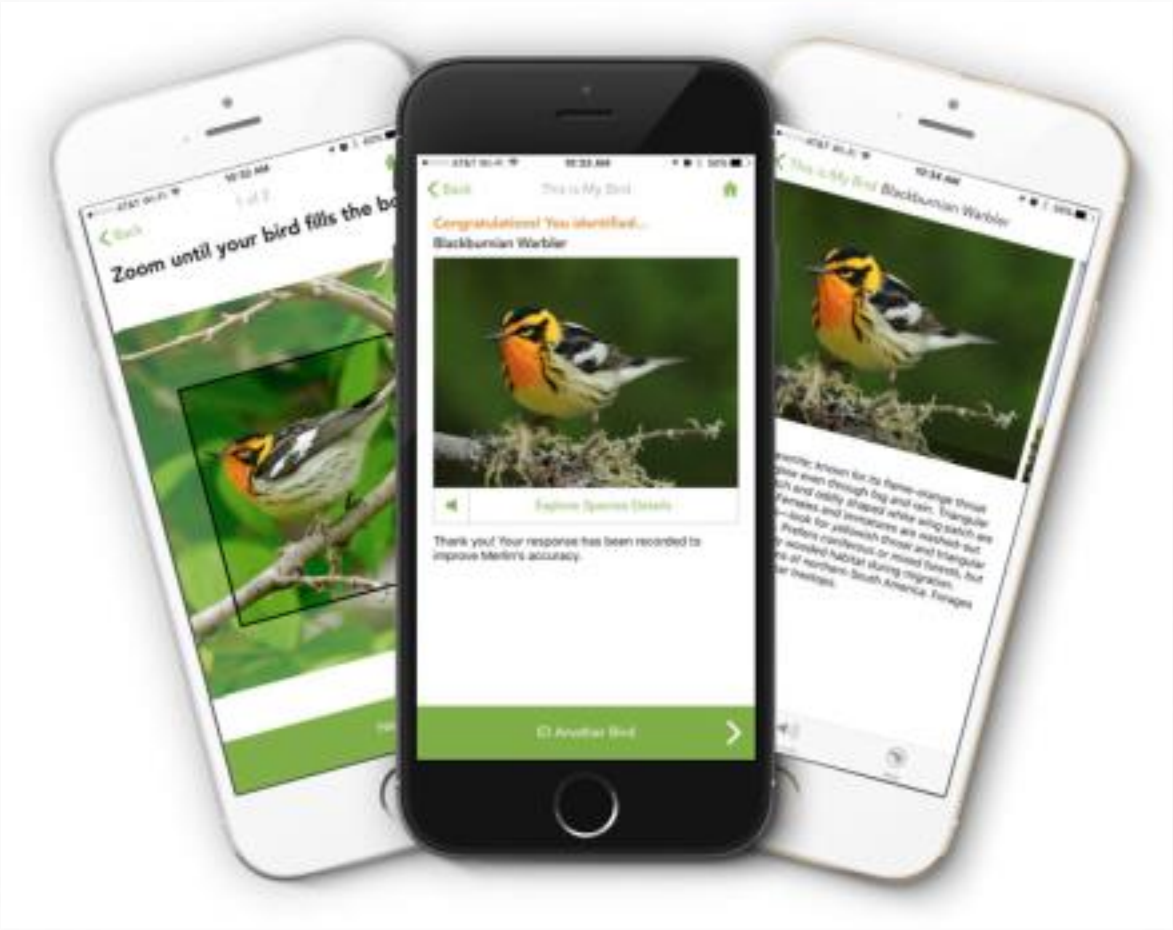
FACE RECOGNITION

FACE++

What about privacy, China?



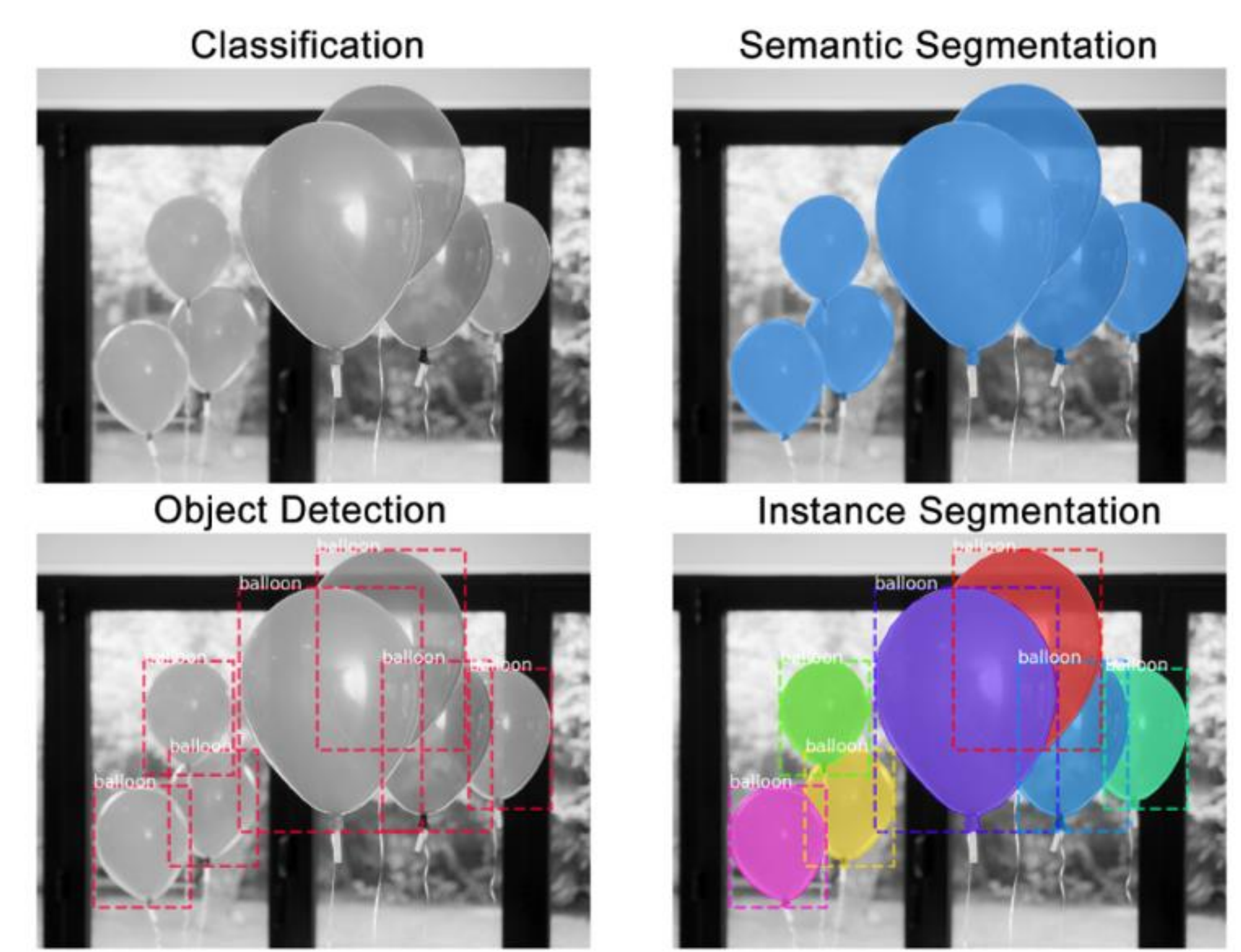
BIRD IDENTIFICATION



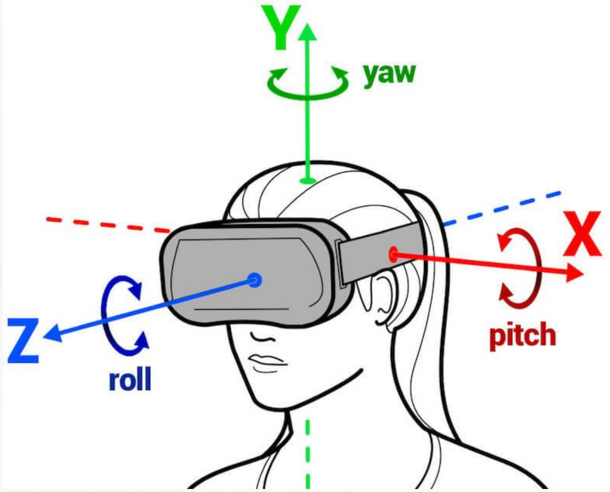
Merlin Bird ID (based on Cornell Tech technology!)

CLASSIFICATION AND SEGMENTATION

- Semantic Instance Segmentation
- City planning, building development
- Combining with AR
- Surveillance and monitoring



VIRTUAL & AUGMENTED REALITY



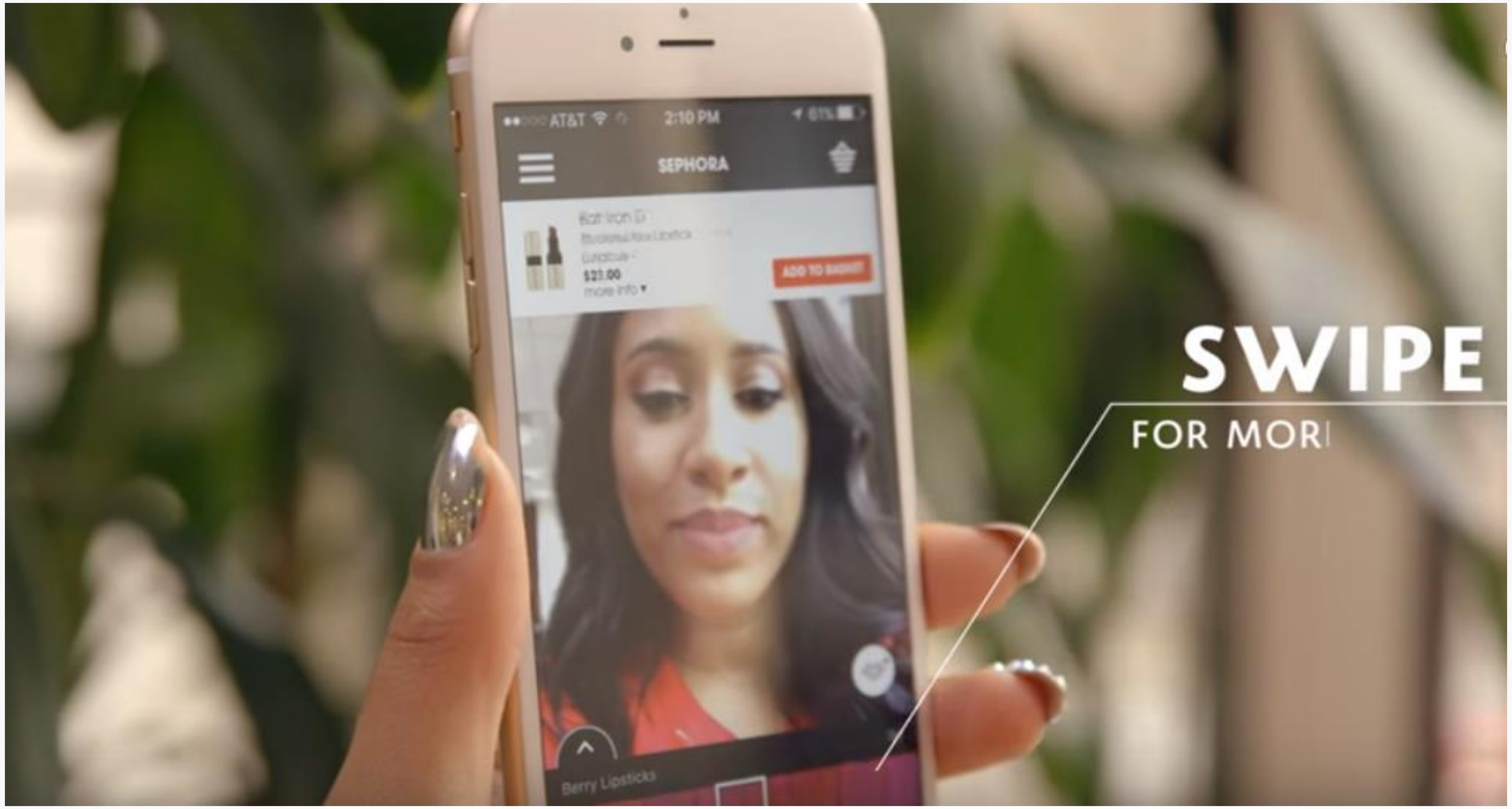
6DoF head tracking

Hand & body tracking

3D-360 video capture

- VR and AR should be developed into Merged Reality (MR)
  - External cameras, other sensors => perceive environment, detect eye and body movement, provide guidance

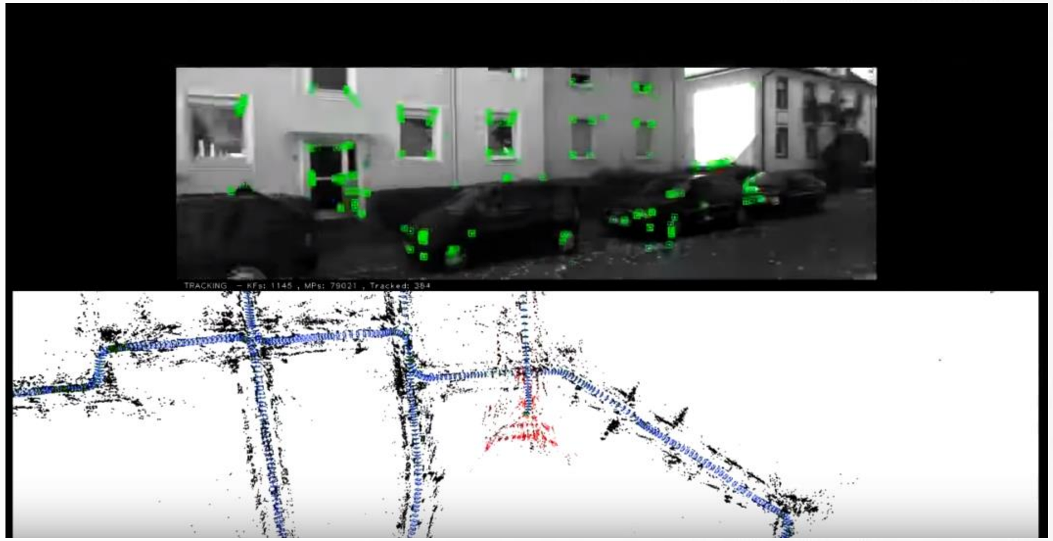
SEPHORA VIRTUAL APP



https://www.youtube.com/watch?v=NFApcSocFDM

SIMULTANEOUS LOCALIZATION AND MAPPING (SLAM)

- SLAM produces simultaneously
  - a map of the unknown environment
  - while positioning the user in this newfound map



https://www.youtube.com/watch?v=8DISRmsO2YQ

AUTONOMOUS DRIVING

- Vision is a key technology for automated driving
- Different cameras
  - Monocular observing the environment; objects and **lane marks**
  - Stereo for observing pedestrians
  - Laser scanner for 3D object recognition

Camera x 4

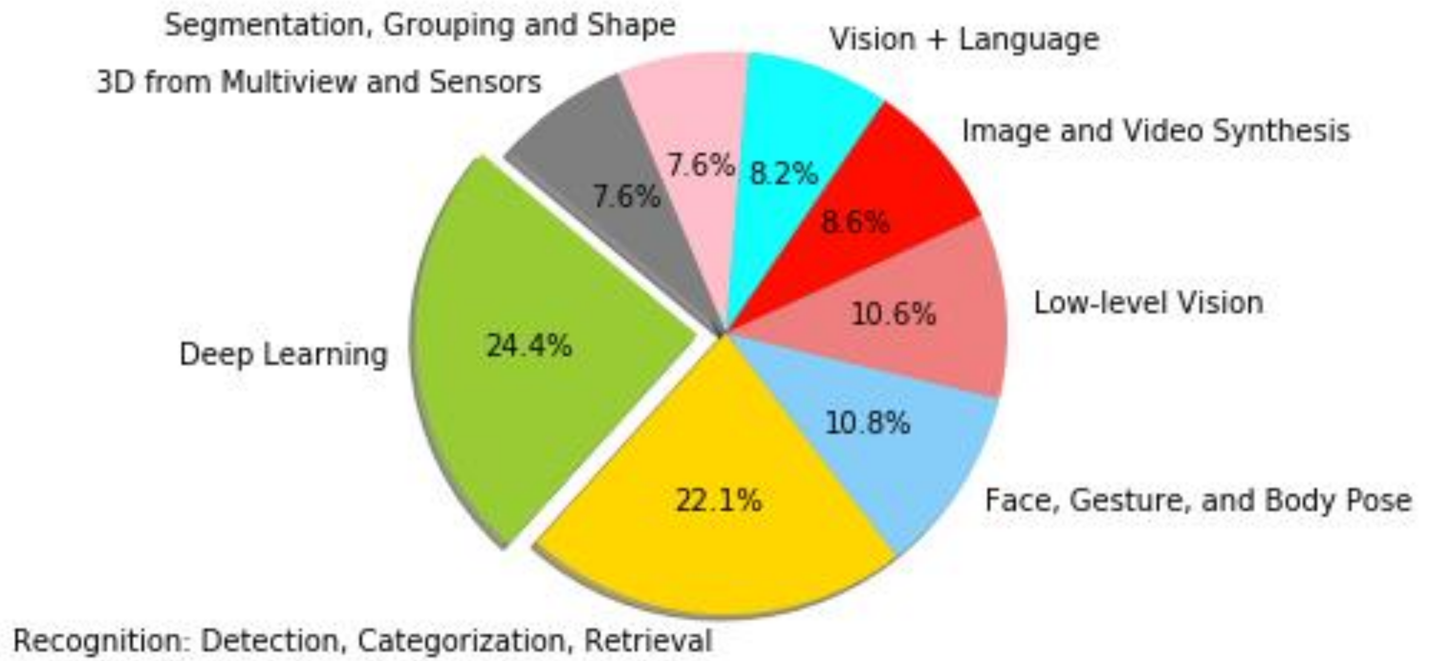
Tri objective camera = stereo + monocular

camera

https://www.volvocars.com/intl/buy/explore/intellisafe/autonomous-driving

COMPUTER VISION RESEARCH NOW

- Best research from 2018: https://www.topbots.com/most-important-ai-computer-vision-research/
- Coolest 10 papers from CVPR 2018 towardsdatascience.com



COMPUTER VISION RESEARCH NOW

- Development of cheap but effective cameras and toolkits
  - Intel RealSense
  - Google ARCore https://developers.google.com/ar/discover/, Apple AR https://developer.apple.com/augmented-reality/
- Deep learning, edge computing provides means for effective computing
- Cheaper lasers and LiDARs, however this course consentrates on passive cameras (RGB)

Microsoft Kinect

Intel Realsense

COMPUTER VISION RESEARCH NOW

- Great benchmark repositories for doing research, e.g. KITTI http://www.cvlibs.net/datasets/kitti/index.php
- OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library https://opencv.org/
  - supports: C++, Java, Python, Matlab
  - Windows, Linux, Android, MAC
- Matlab has many toolboxes for computer vision, still mainly used in teaching at universities around the world

BASICS OF IMAGE UNDERSTANDING

0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0
0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0
0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8	0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7	3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6	2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5	5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4	4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3	7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2	6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1	9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0	8	7	6	5	4	3	2	1	0

What a computer sees

What a person sees

WHY IS COMPUTER VISION DIFFICULT?

Viewpoint variation

Illumination

Scale

Slide credit: Noah Snavely

WHY IS COMPUTER VISION DIFFICULT?

Intra-class variation

Motion (Source: S. Lazebnik)

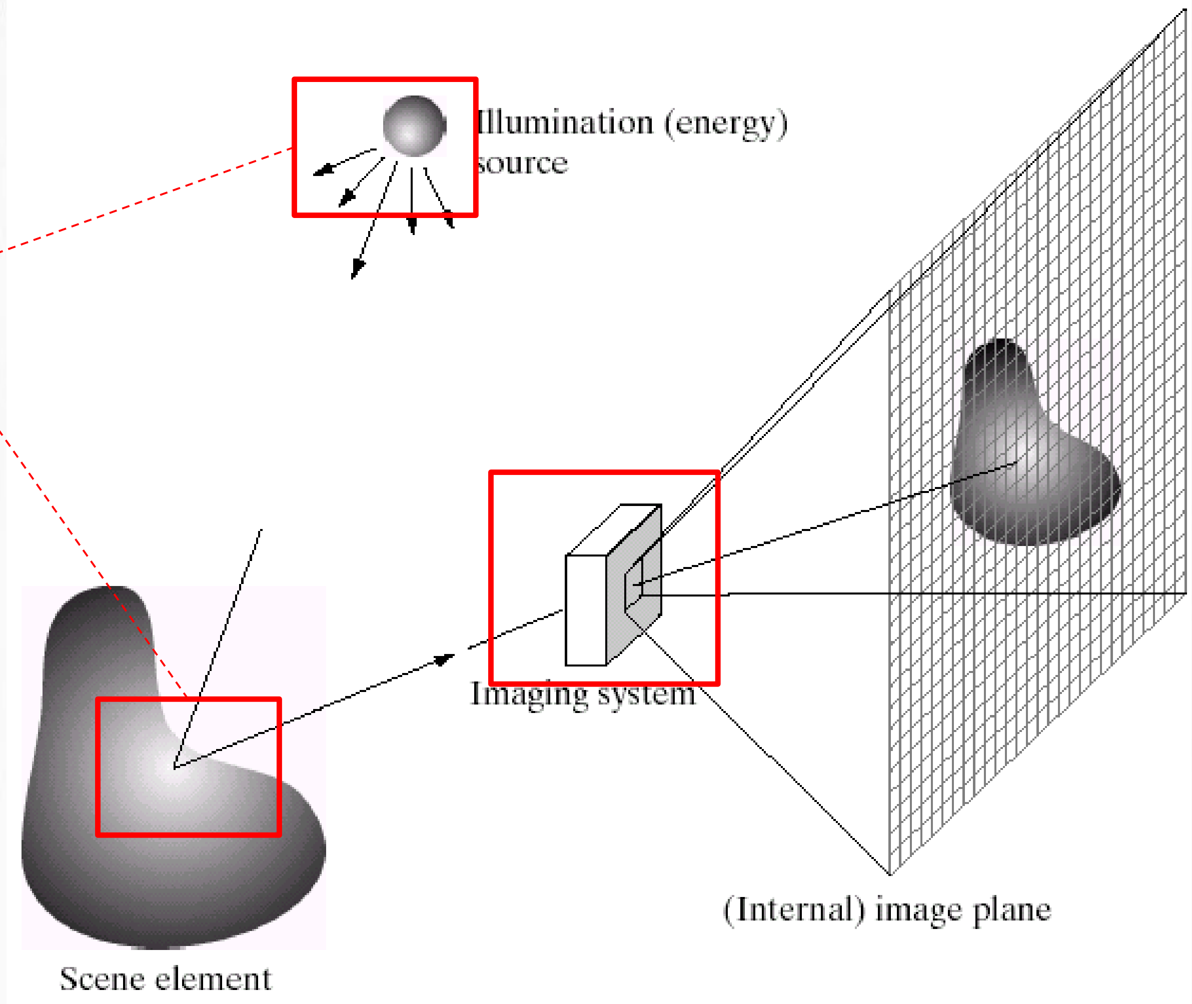
Background clutter

Occlusion

Slide credit: Noah Snavely

WHAT IS AN IMAGE?

Photometric image formation

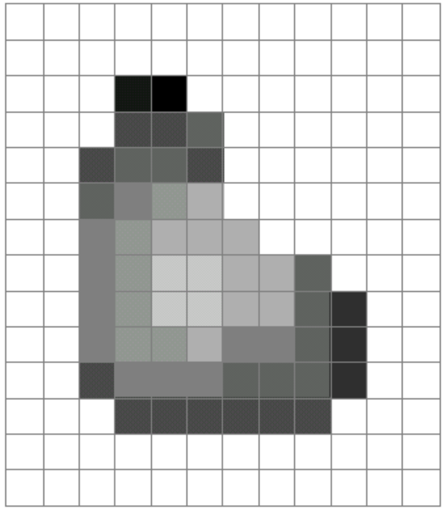


A camera is a mapping between 3D world and a 2D image

- Geometry and transformations
- Camera models
- Camera calibration

WHAT IS AN IMAGE?

A grid (matrix) of intensity values



=

255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	20	0	255	255	255	255	255	255	255
255	255	255	75	75	75	255	255	255	255	255	255
255	255	75	95	95	75	255	255	255	255	255	255
255	255	96	127	145	175	255	255	255	255	255	255
255	255	127	145	175	175	175	255	255	255	255	255
255	255	127	145	200	200	175	175	95	255	255	255
255	255	127	145	200	200	175	175	95	47	255	255
255	255	127	145	145	175	127	127	95	47	255	255
255	255	74	127	127	127	95	95	95	47	255	255
255	255	255	74	74	74	74	74	74	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255	255

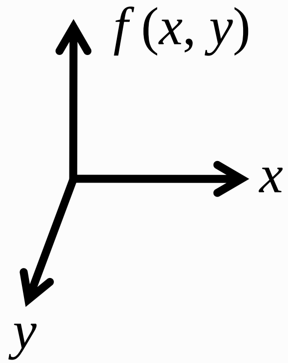
(common to use one byte per value: 0 = black, 255 = white)

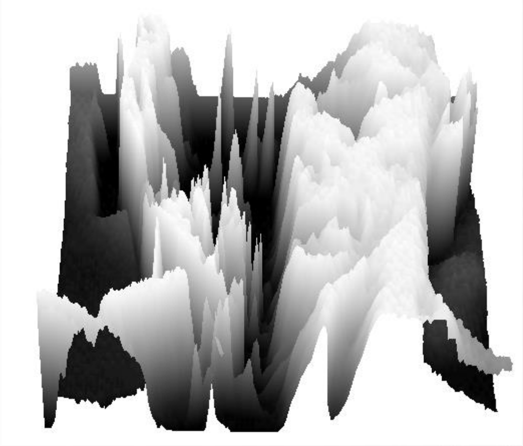
Slide credit:
Noah Snavely

WHAT IS AN IMAGE?

We can think of a (grayscale) image as a **function**, *f*, from R^2 to R:

$f(x,y)$ gives the **intensity** at position (x,y)





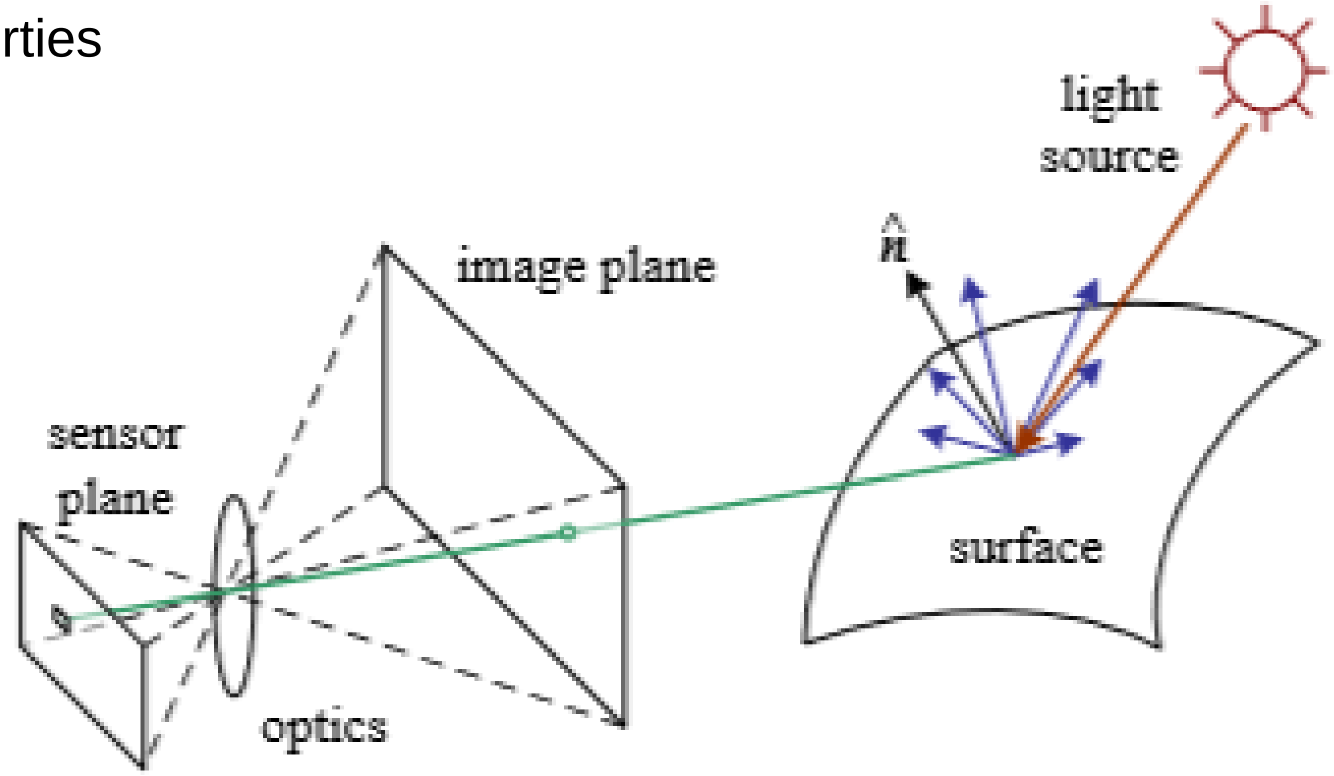
A **digital** image is a discrete (**sampled**, **quantized**) version of this function

Slide credit:
Noah Snavely

PHOTOMETRIC IMAGE FORMATION

- Camera optics
- Sensor properties

- Lighting in the environment
- Surface properties and geometry

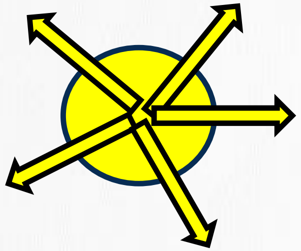


Szeliski

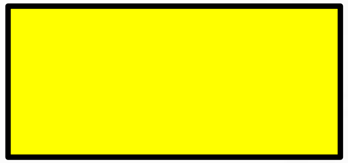
Images are made of discrete intensity values.

PHOTOMETRIC IMAGE FORMATION

- To produce an image the scene must be illuminated with one or more light sources



Sun, light bulb, …



Sky, fluorescent light boxes, …

Point source

- Intensity and color spectrum $L(\lambda)$
- Intensity falls of with the square of distance

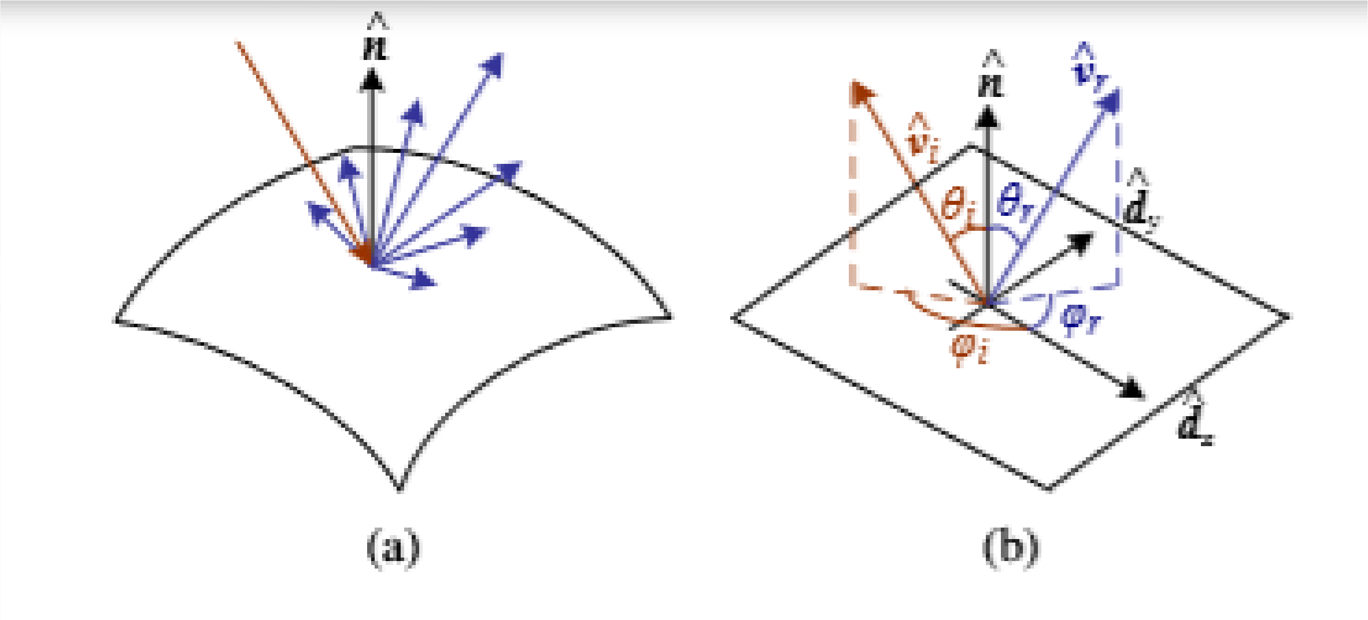
Area source

- Uniform illumination, does not change with distance

- Light source described by its emitted radiance and source’s geometry

PHOTOMETRIC IMAGE FORMATION

- When light hits an object's surface it is scattered and reflected
- Bidirectional Reflectance Distribution Function (BRDF) = how much of each wavelength (λ) arriving at an incident angle (v_i) is emitted in a reflected direction (v_r)



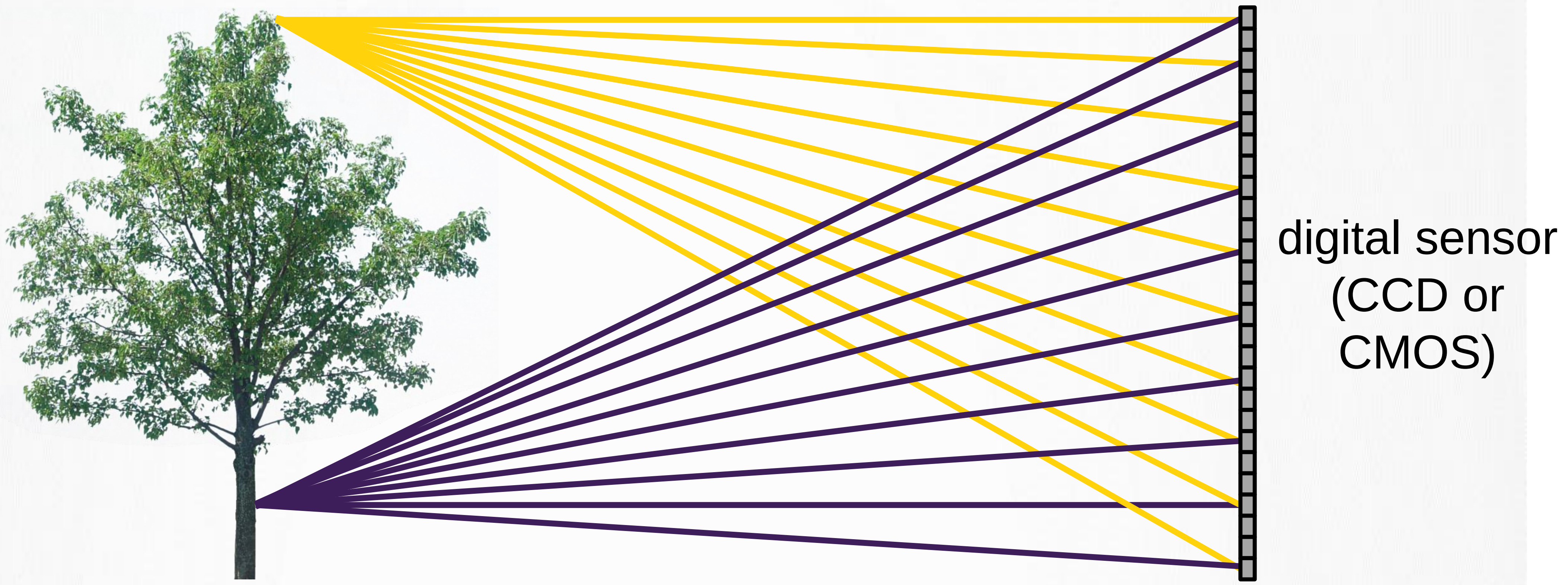
$$f_r(\theta_i, \phi_i, \theta_r, \phi_r; \lambda).$$

Amount of light exiting the surface is computed by integrating product of BRDF and incoming light $L_i(v_i, \lambda)$
=> Convolution

PHOTOMETRIC IMAGE FORMATION

- The simple shading models assume that light rays leave the light source, bounce off surfaces visible to the camera, change in intensity or colour and arrive at the camera
- However, rays are in reality occluded or shadowed and rays can bounce multiple times from surroundings
- Effects may be modelled
  - Ray tracing, mainly when scene is specular
  - Radiosity, when uniform albedo simple geometry illuminators and surfaces

BARE-SENSOR IMAGING



All scene points contribute to all sensor pixels

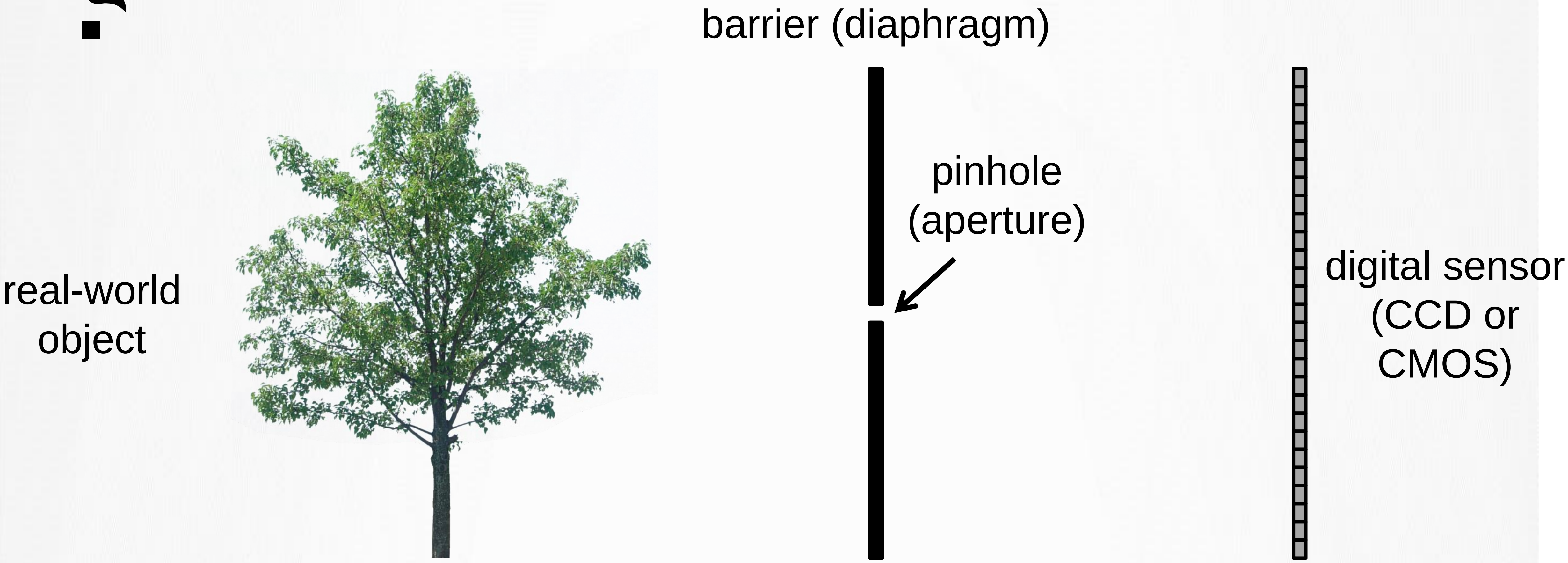
Slide credit: Kris Kitani

BARE-SENSOR IMAGING

All scene points contribute to all sensor pixels

Slide credit:
Kris Kitani

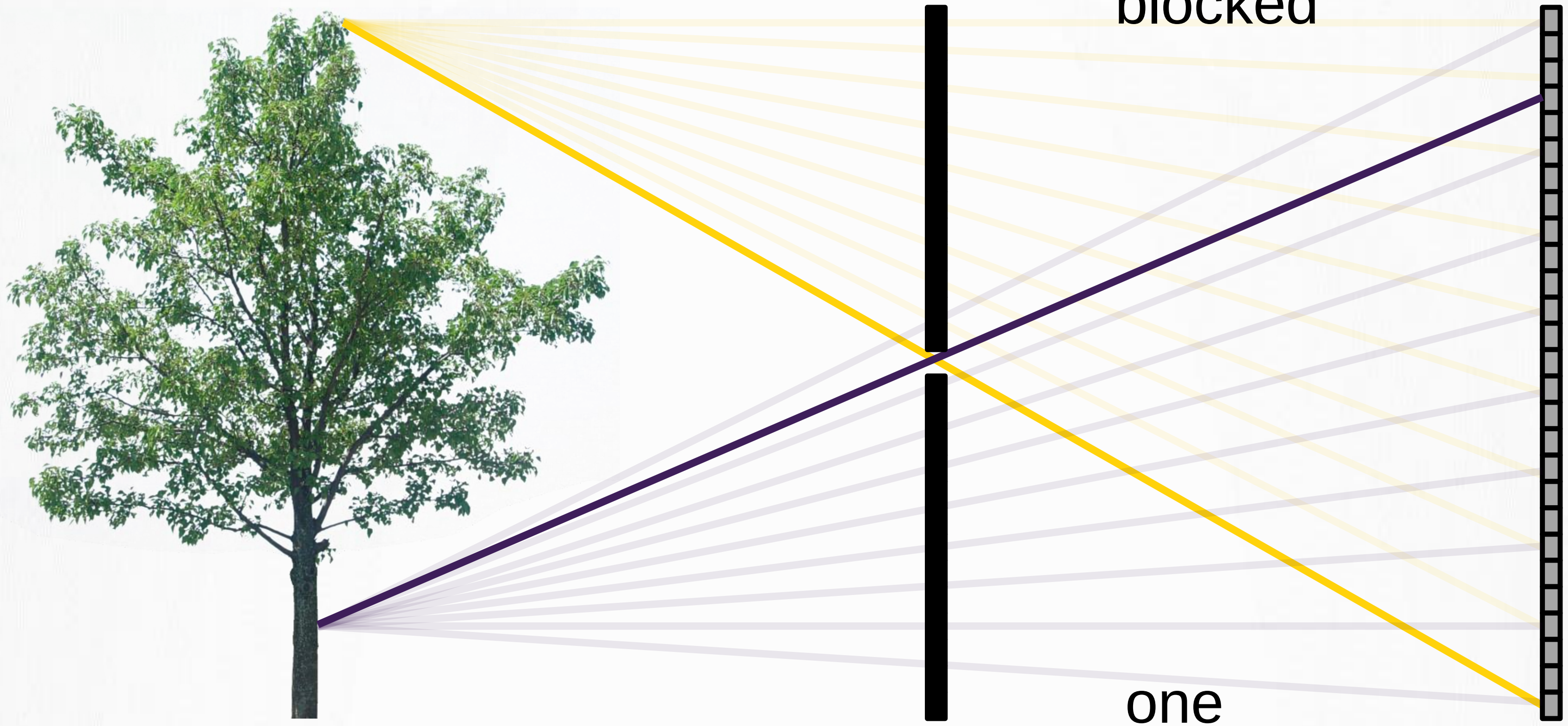
LET’S ADD SOMETHING TO THIS SCENE



A barrier blocking off most of the rays with an **aperture** (an opening) reduces blurring

Slide credit:
Kris Kitani

PINHOLE IMAGING



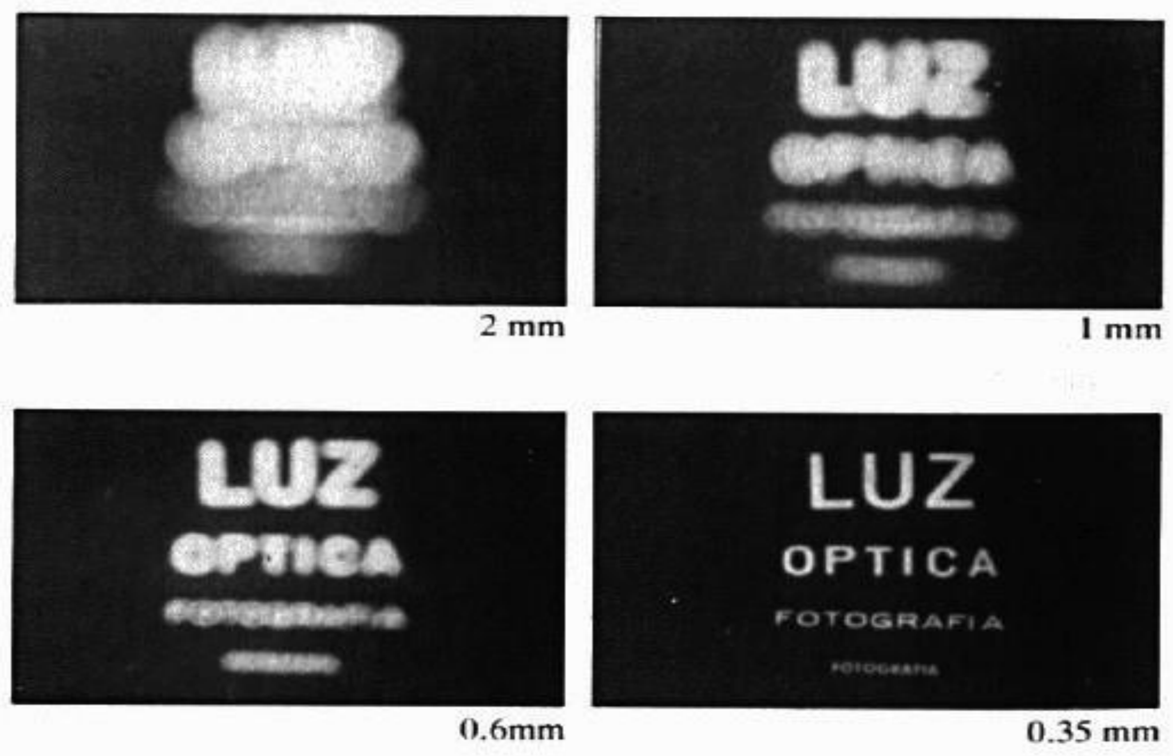
Each scene point contributes to only one sensor pixel

Slide credit: Kris Kitani

PINHOLE IMAGING

Slide credit:
Kris Kitani

SHRINKING THE APERTURE 1

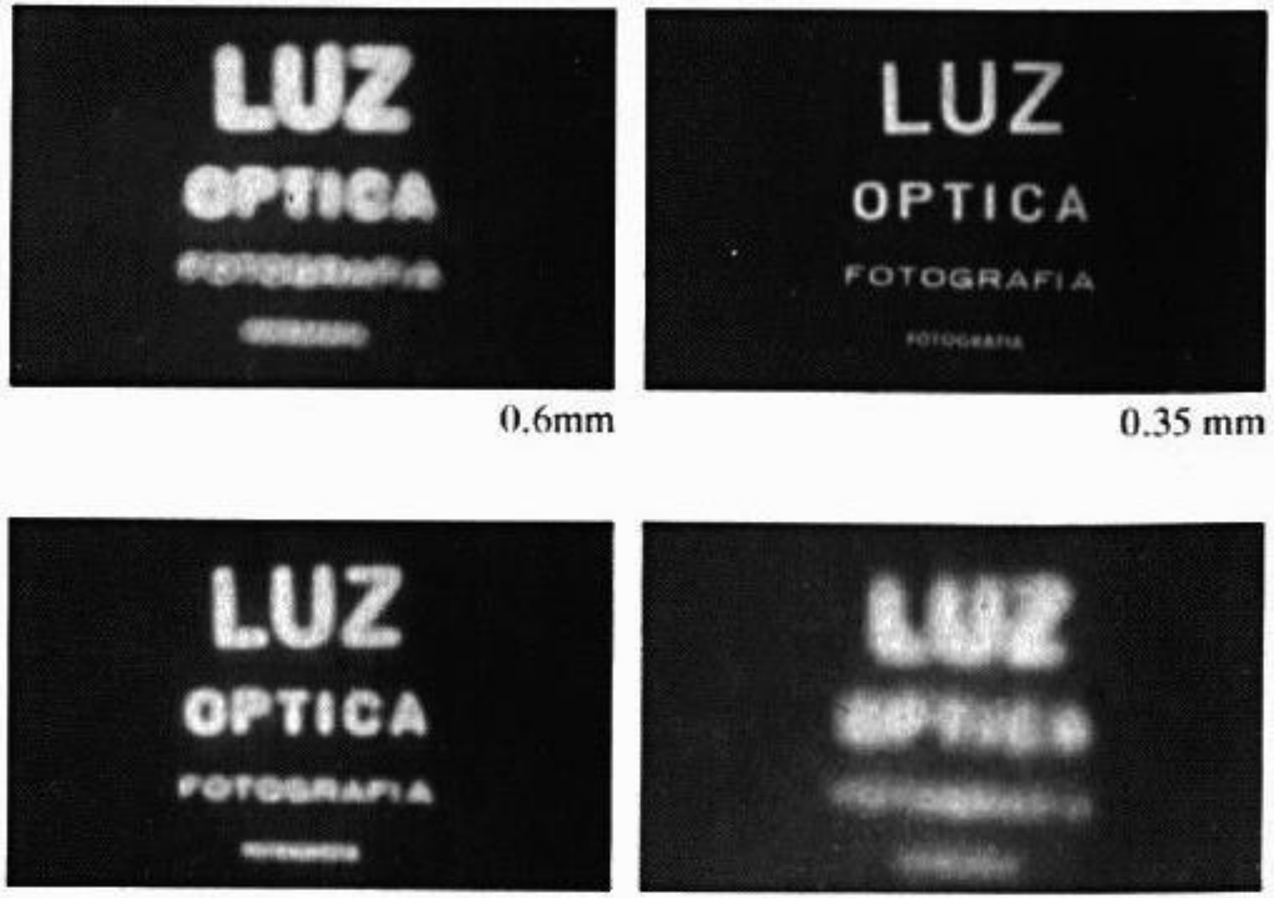


Why not make the aperture as small as possible?

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

Slide credit:
Noah Snavely

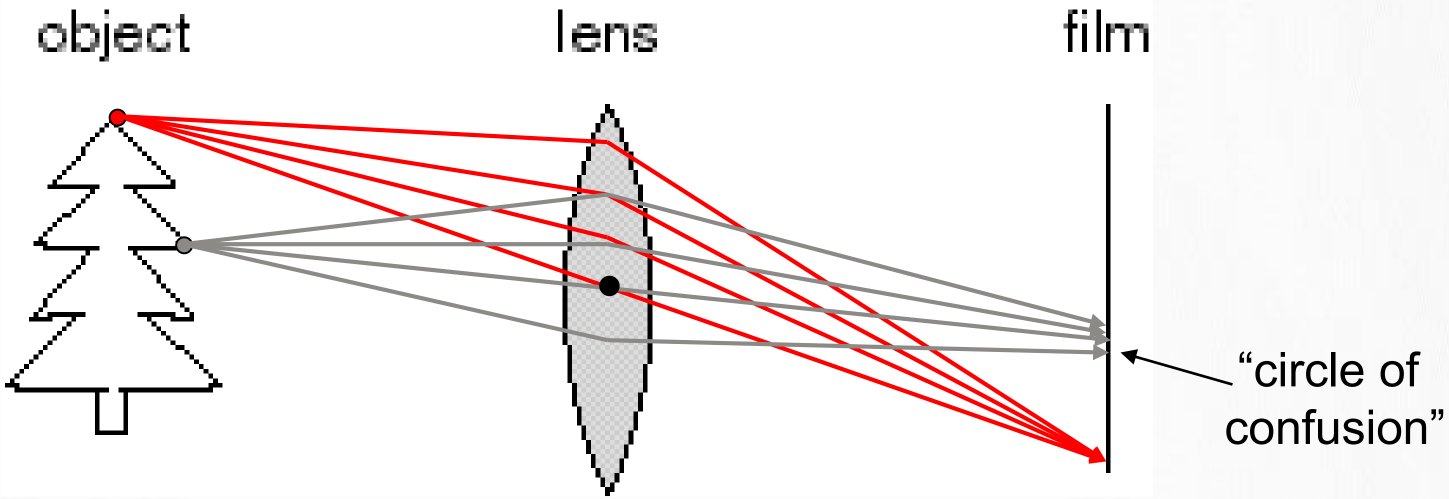
SHRINKING THE APERTURE 2



0.6mm 0.35 mm 0.15 mm 0.07 mm

Slide credit:
Noah Snavely

ADDING A LENS



A lens focuses light onto the film

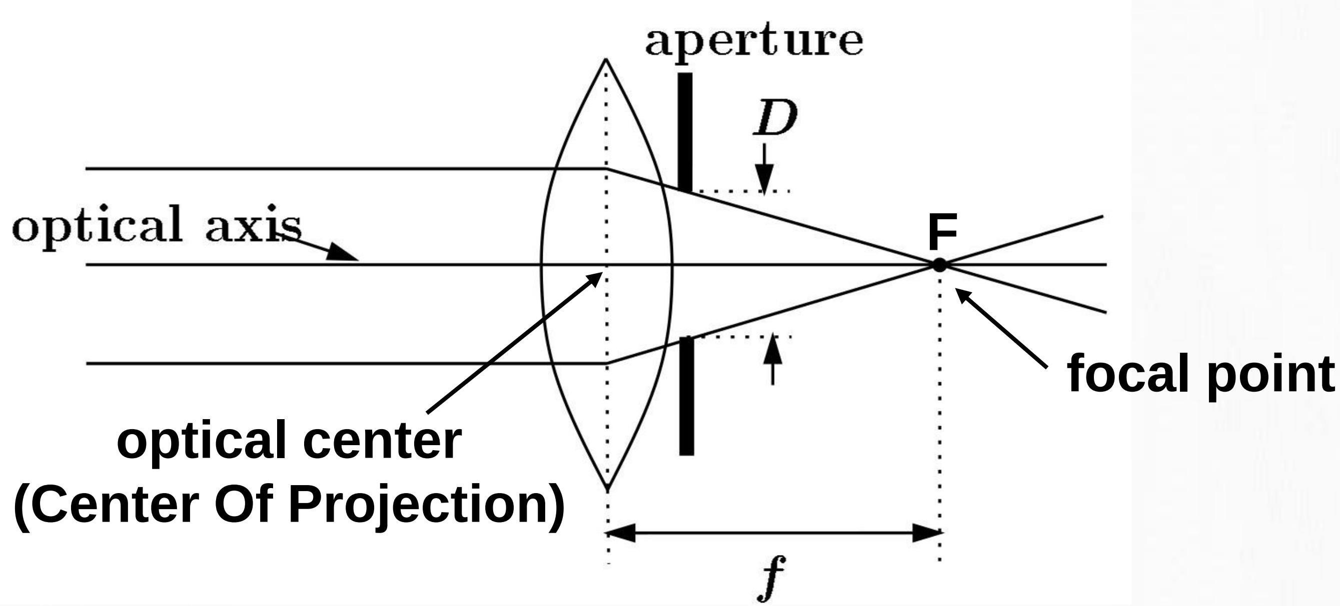
There is a specific distance at which objects are "in focus"

– other points project to a "circle of confusion" in the image

Changing the shape of the lens changes this distance

Slide credit:
Noah Snavely

LENSES



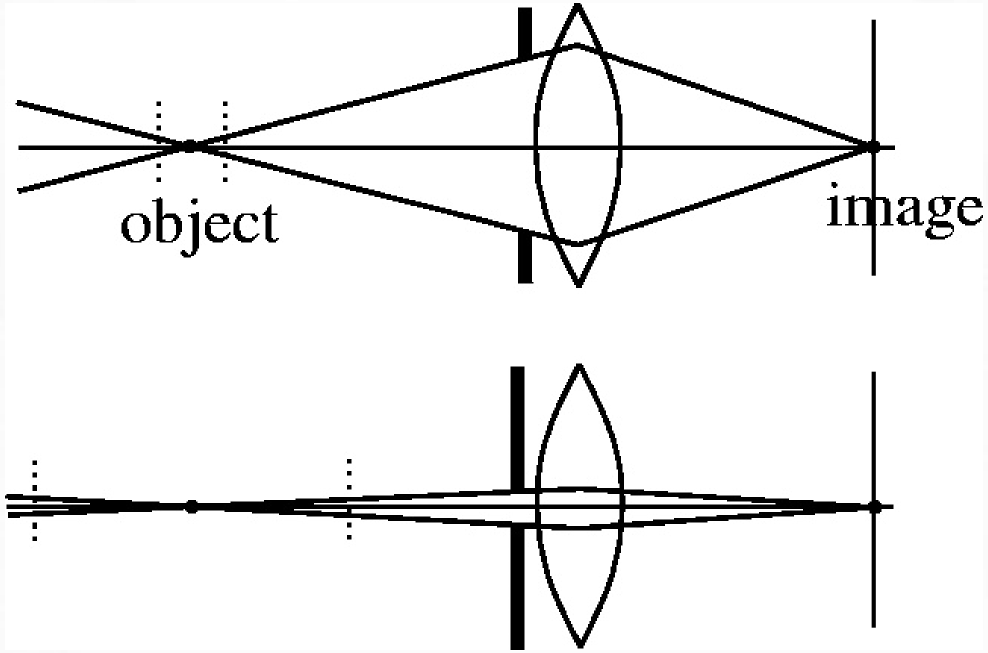
A lens focuses parallel rays onto a single focal point

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

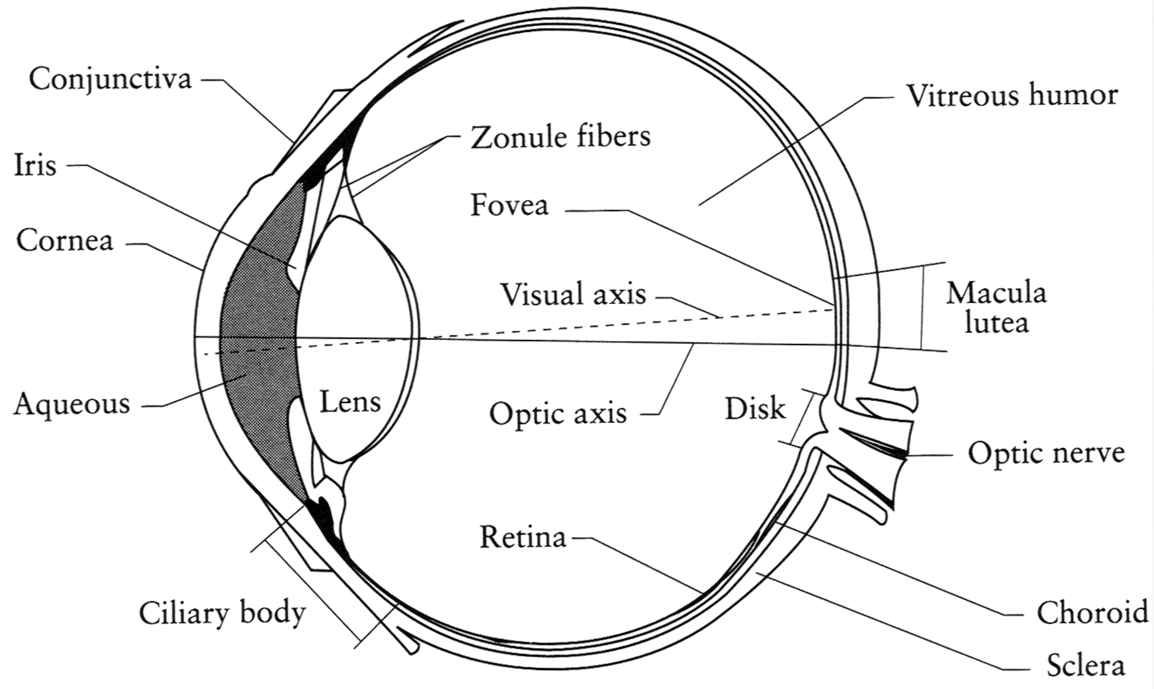
DEPTH OF FIELD



Changing the aperture size affects depth of field

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

THE EYE



The human eye is a camera

Iris - colored annulus with radial muscles

Pupil - the hole (aperture) whose size is controlled by the iris

What's the "film"?

– photoreceptor cells (rods and cones) in the **retina**

And what about the computer vision algorithms and processor!

DIGITAL CAMERA

A digital camera replaces film with a sensor array

Each cell in the array is a **C**harge **C**oupled **D**evice (CCD)

– light-sensitive diode that converts photons to electrons

– other variants exist: CMOS is becoming more popular

– http://electronics.howstuffworks.com/digital-camera.htm

60˚ 10 1.2 N, 24˚ 57 18 E