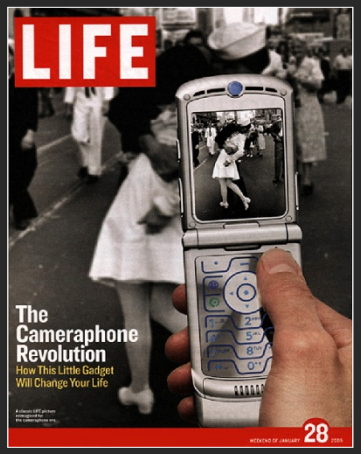
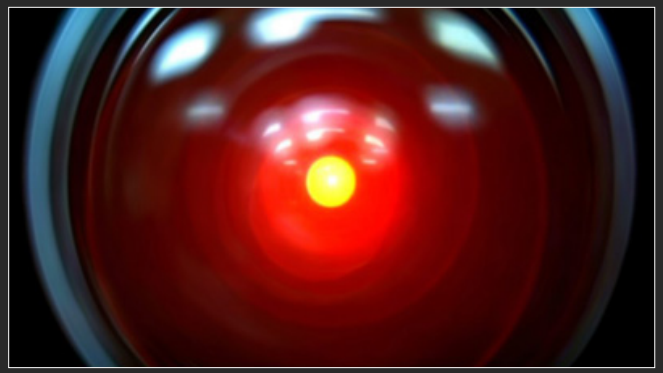
Introduction

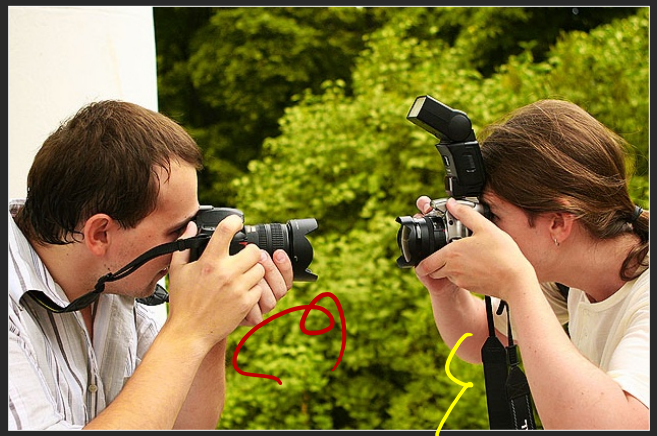
<Vision System>

Department of Robot Engineering
Prof. Younggun Cho

Computer Vision - Introduction



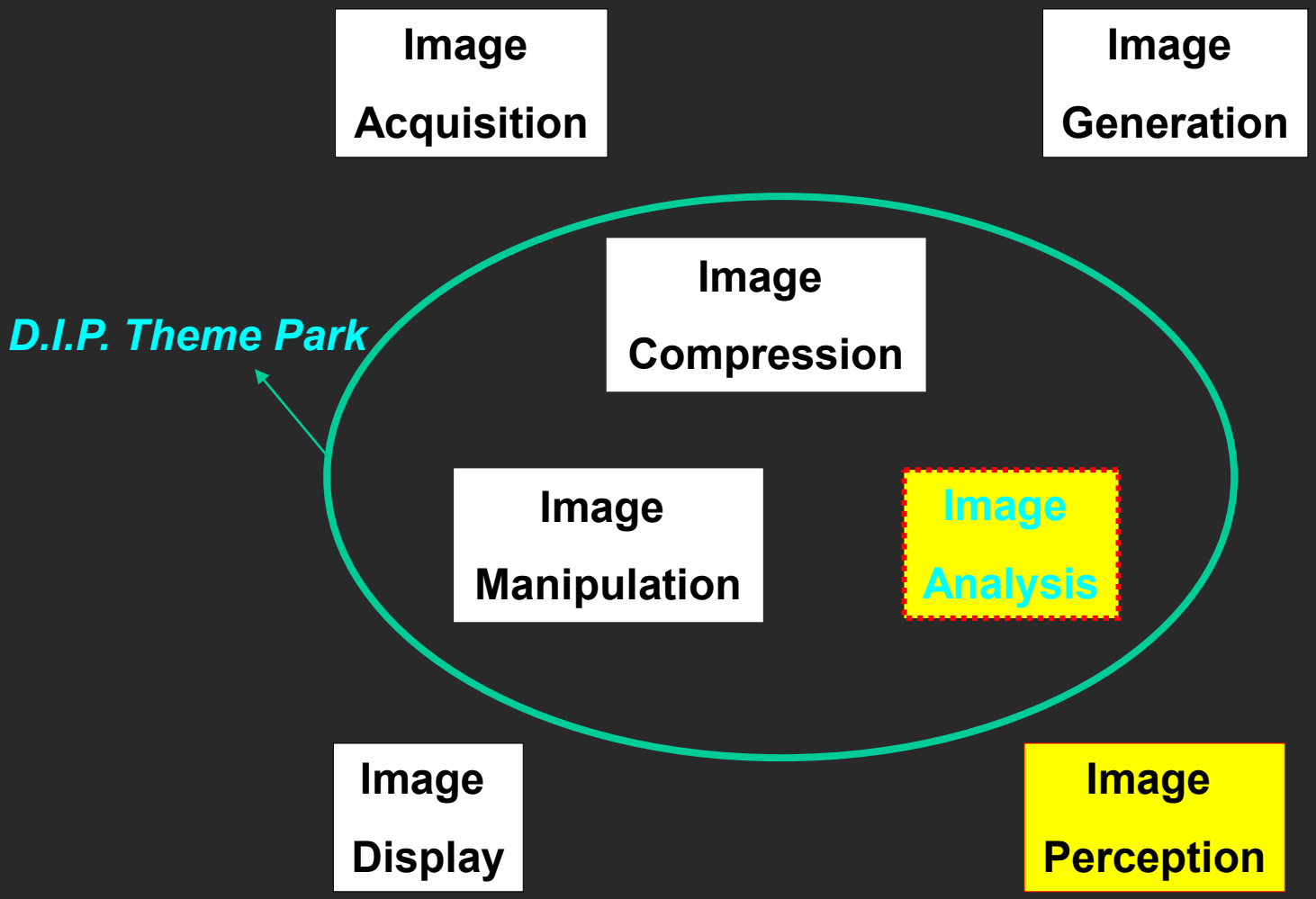




Today

- Introduction to computer vision
- Course overview

cf. Digital Image Processing



Scope of DIP

- Reasons for compression
  - Image data need to be accessed at a different time or location
  - Limited storage space and transmission bandwidth
- Reasons for manipulation
  - Image data might experience nonideal acquisition, transmission or display (e.g., restoration, enhancement and interpolation)
  - Image data might contain sensitive content (e.g., fight against piracy, conterfeit and forgery)
  - To produce images with artistic effect (e.g., pointellism)
- Reasons for Analysis
  - Image data need to be analyzed automatically in order to reduce the burden of human operators
  - To teach a computer to “see” in A.I. tasks

The goal of computer vision

- To perceive the story behind the picture

What we see

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

What a computer sees

Source: S. Narasimhan

The goal of computer vision

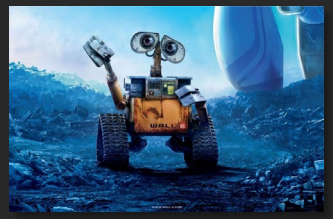
- To perceive the story behind the picture

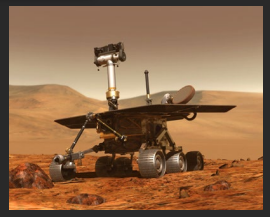
- What exactly does this mean?
  - Vision as a source of metric 3D information
  - Vision as a source of semantic information

The goal of computer vision

Vision as measurement device

Real-time stereo

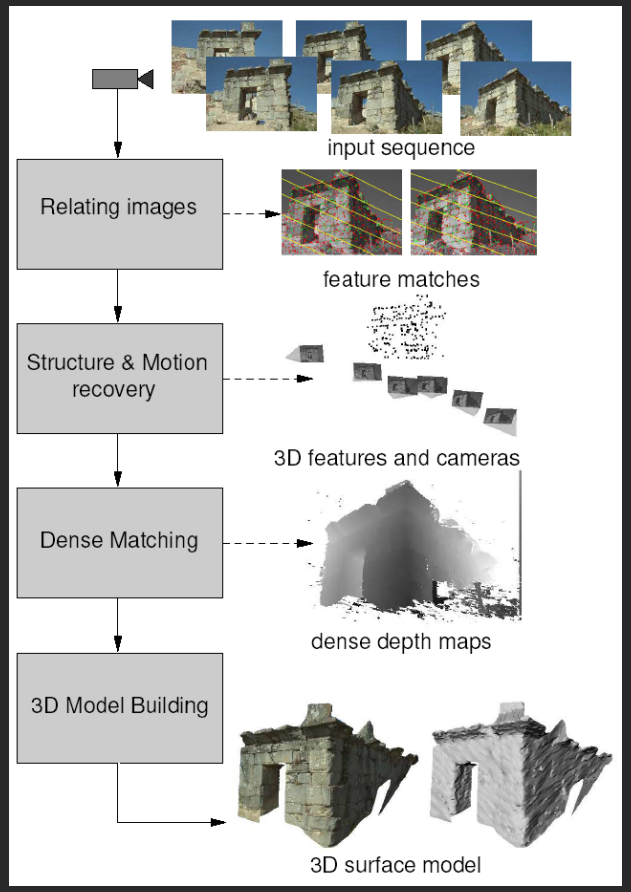




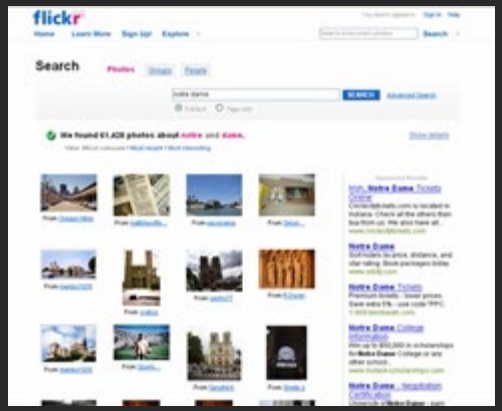
NASA Mars Rover

Pollefeys et al.

Structure from motion



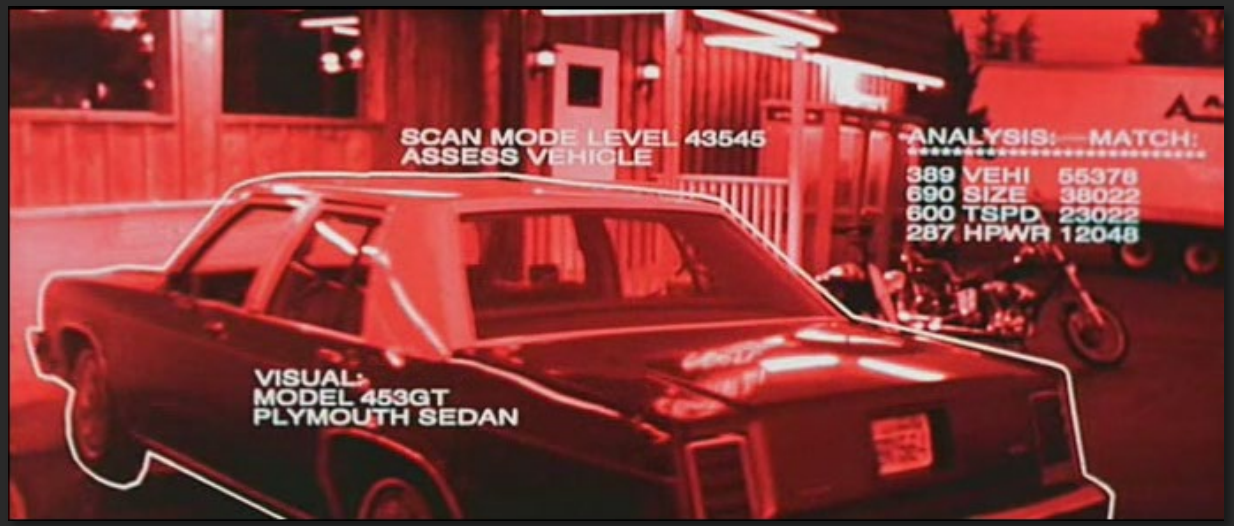
Multi-view stereo for community photo collections



Goesele et al.

The goal of computer vision

- Recognize objects and people



Terminator 2, 1991

Object categorization

The goal of computer vision

- Forensics

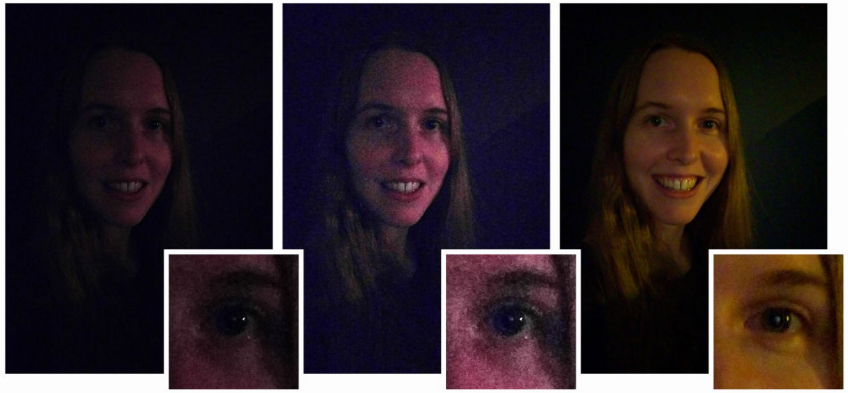
Source: Nayar and Nishino, “Eyes for Relighting”

Source: Nayar and Nishino, “Eyes for Relighting”

The goal of computer vision

- Improve photos (“Computational Photography”)

Super-resolution (source: 2d3)



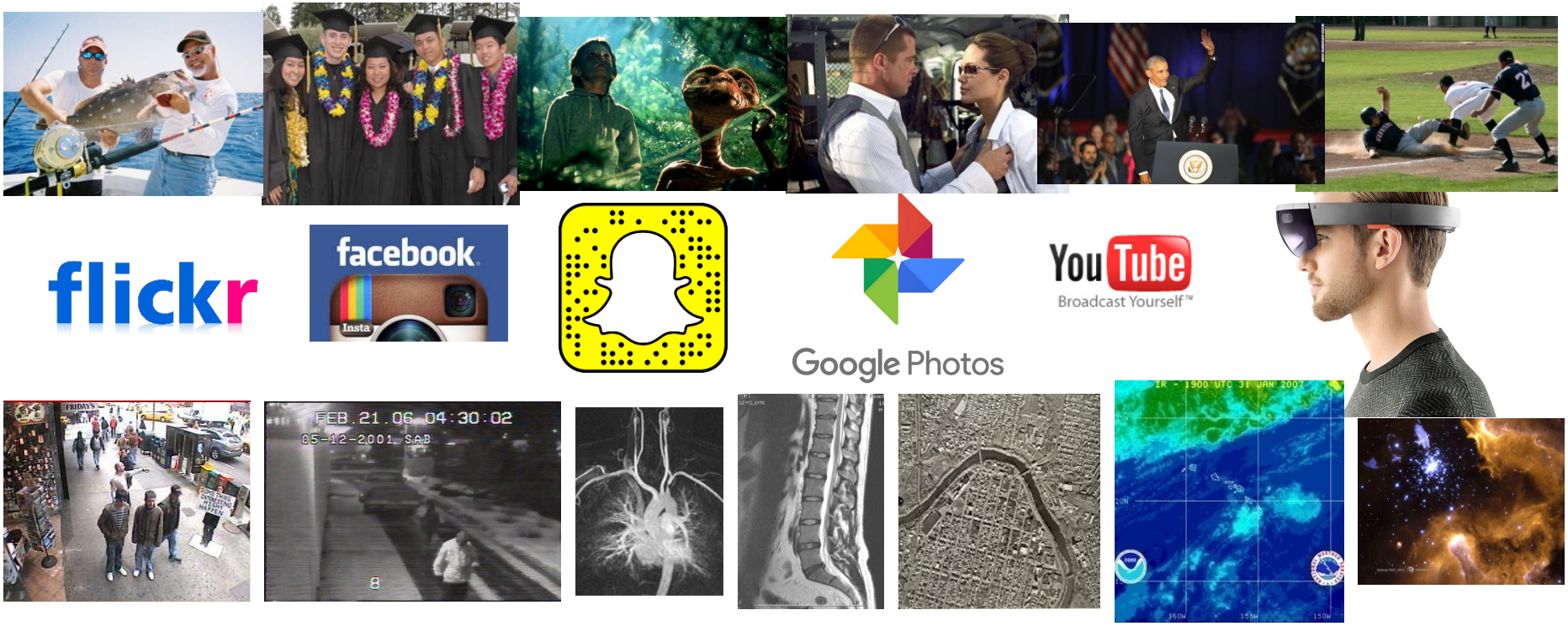
Low-light photography
(credit: Hasinoff et al., SIGGRAPH ASIA 2016)

Depth of field on cell phone camera
(source: Google Research Blog)

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

Why study computer vision?

- Billions of images/videos captured per day



- Huge number of useful applications
- The next slides show the current state of the art

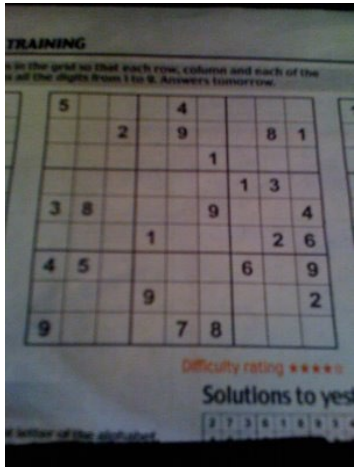
Optical character recognition (OCR)

- 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

Automatic check processing



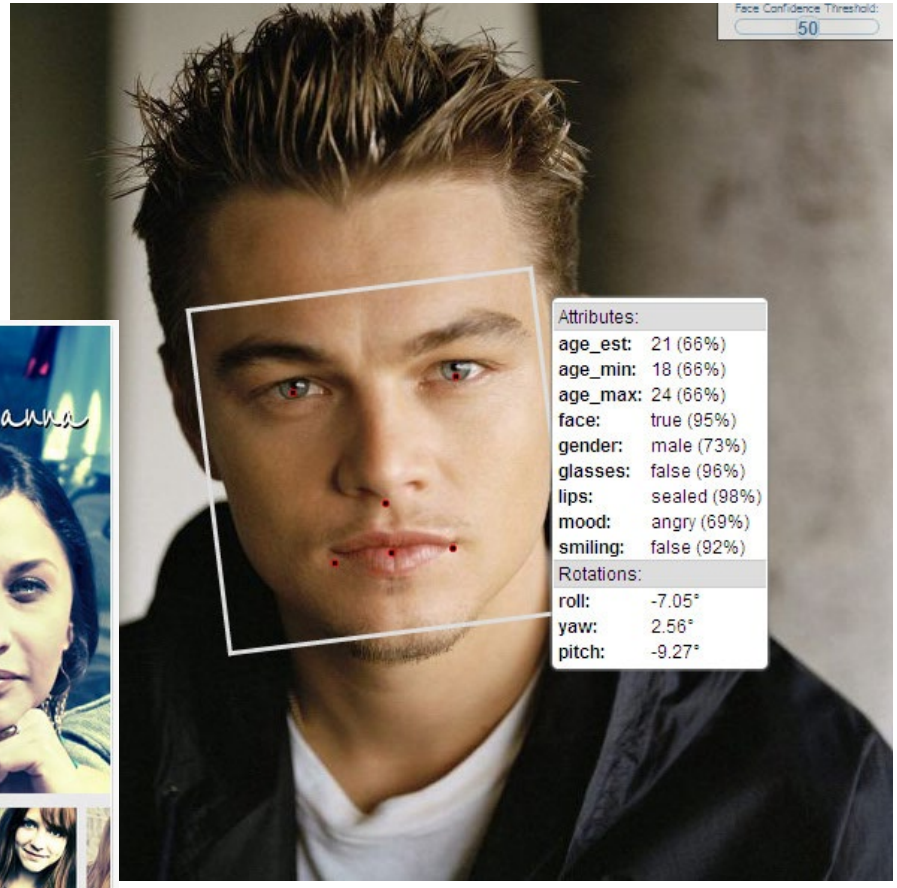
Sudoku grabber
http://sudokugrab.blogspot.com/

Face detection

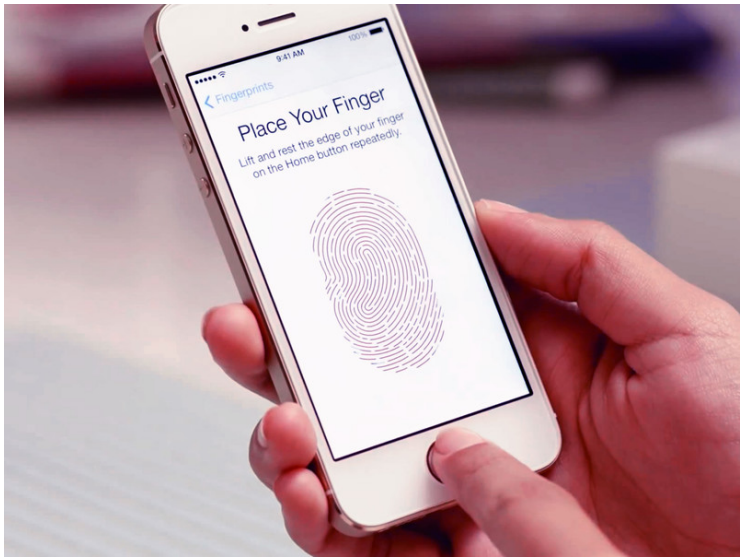
- Nearly all cameras detect faces in real time
  - (Why?)

Face Recognition

FACE++



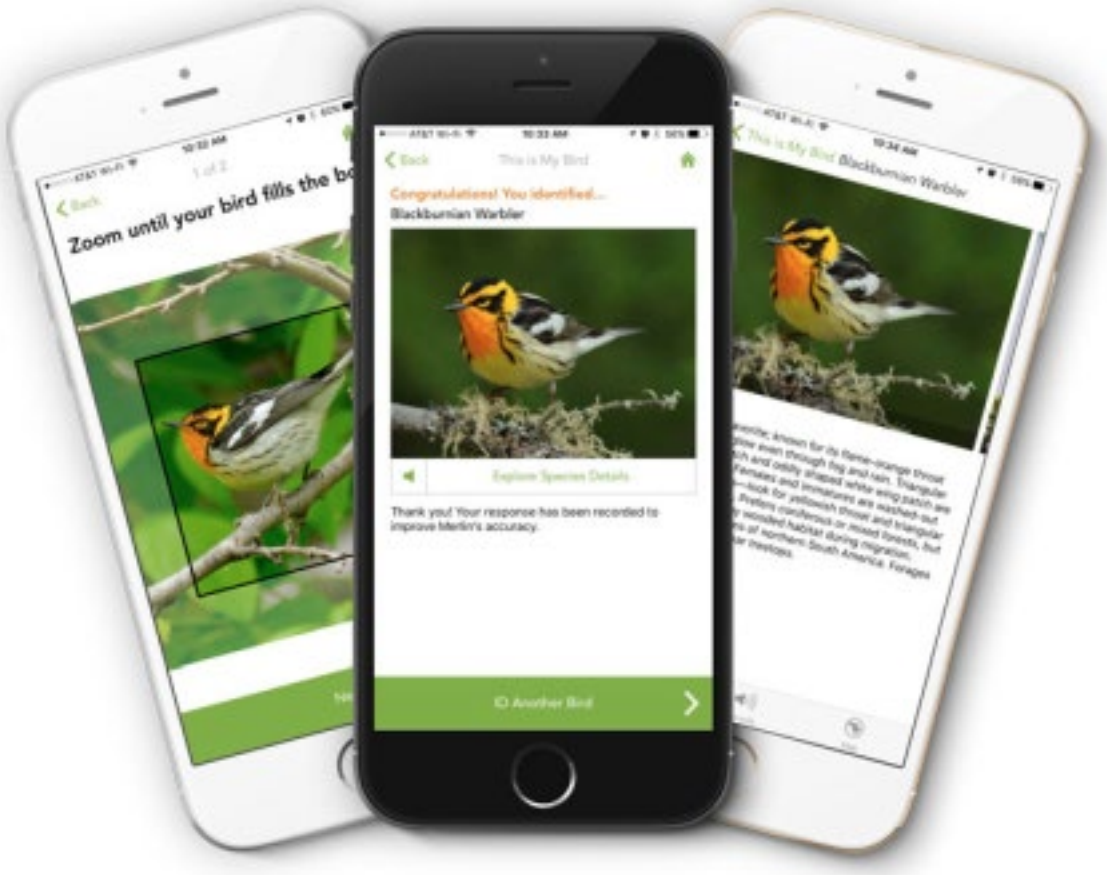
Login without a password



Fingerprint scanners on many new smartphones and other devices

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

Bird Identification



Merlin Bird ID (based on Cornell Tech technology!)

Special effects: camera tracking

Boujou, 2d3

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

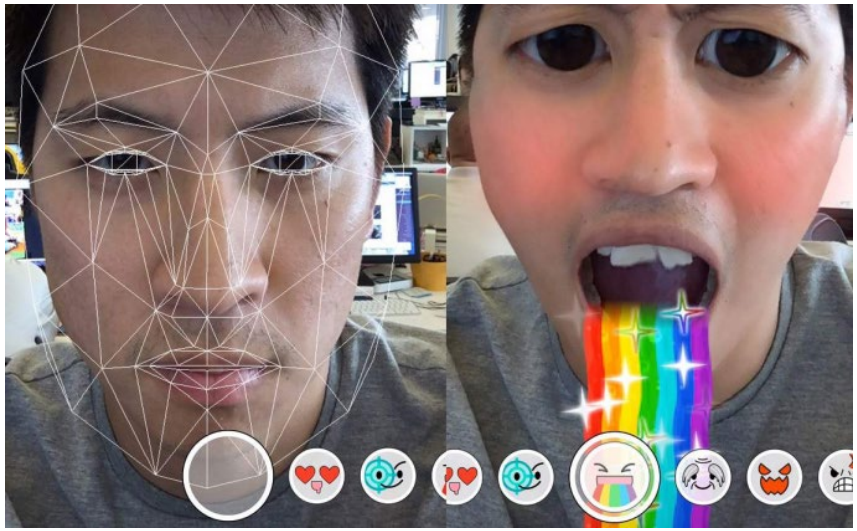
Source: S. Seitz

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

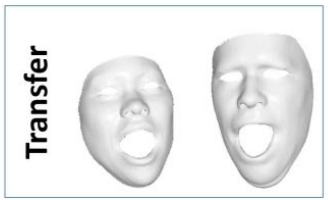
Source: S. Seitz

3D face tracking w/ consumer cameras



Snow

Snapchat Lenses

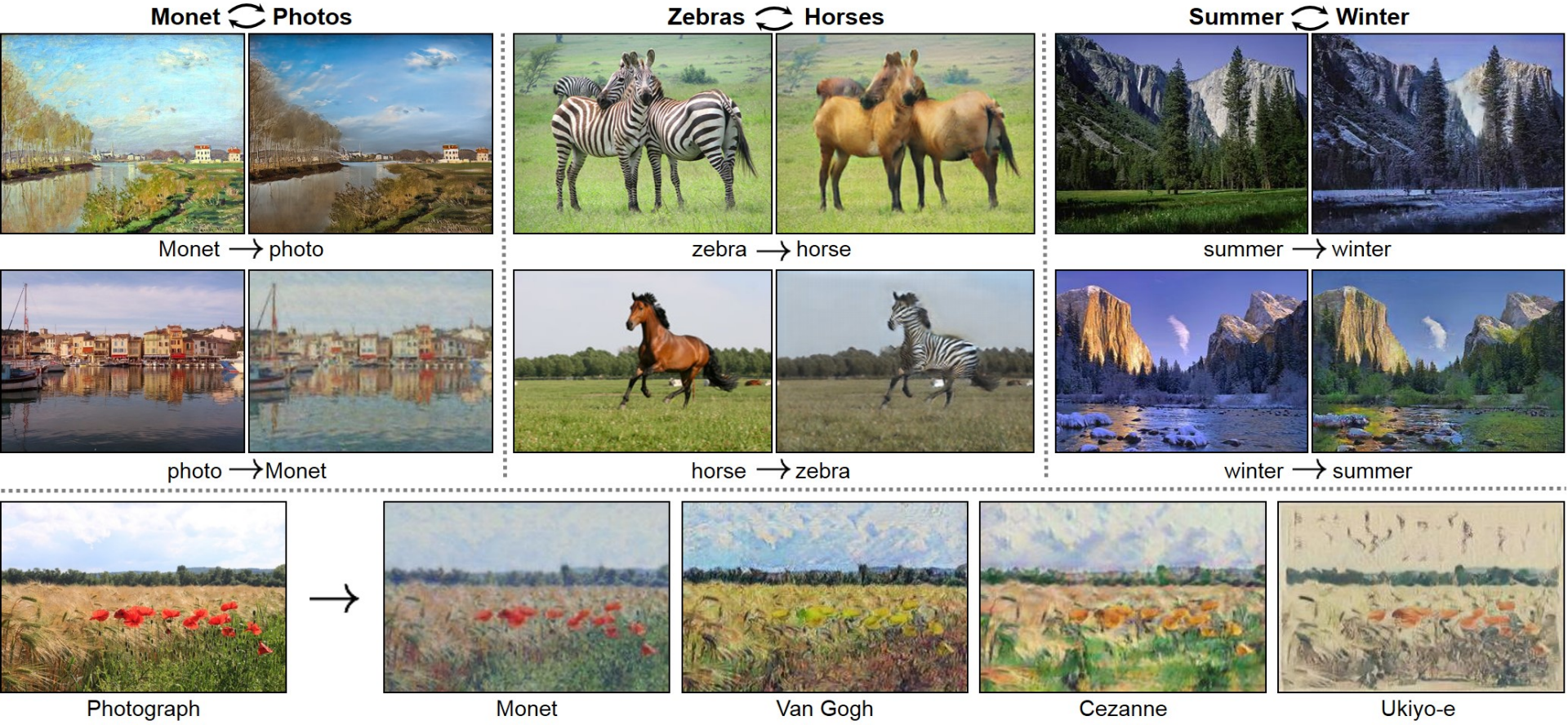


Face2Face system (Thies et al.)

Image synthesis

Karras, et al., *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, ICLR 2018

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

Source: S. Seitz

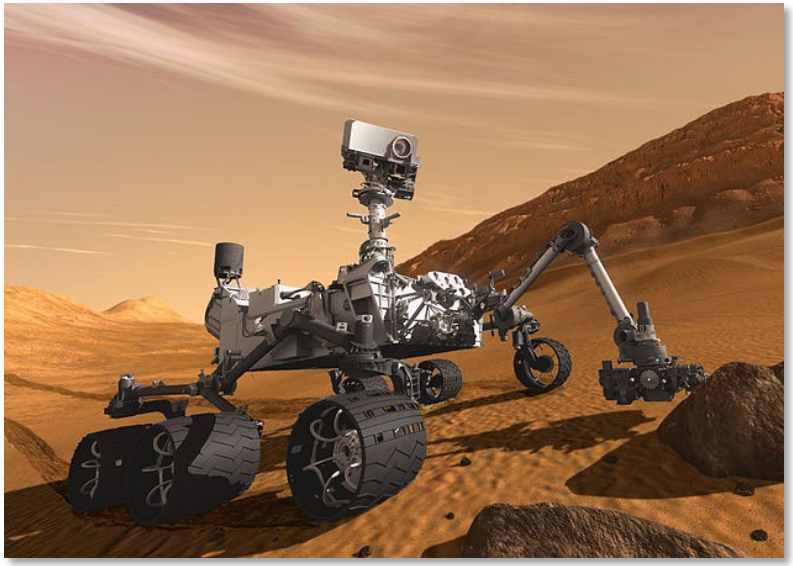
Smart cars

- Mobileye
- Tesla Autopilot
- Safety features in many high-end cars

Self-driving cars

Google Waymo

Robotics

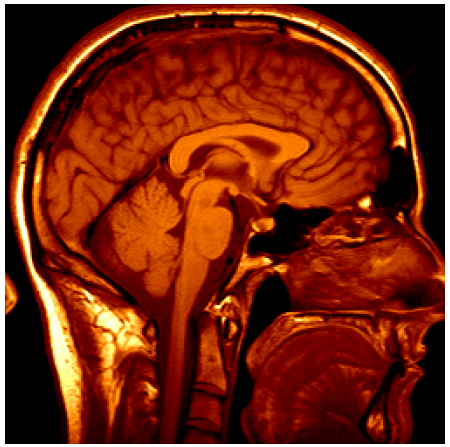


NASA's Mars Curiosity Rover
https://en.wikipedia.org/wiki/Curiosity_(rover)

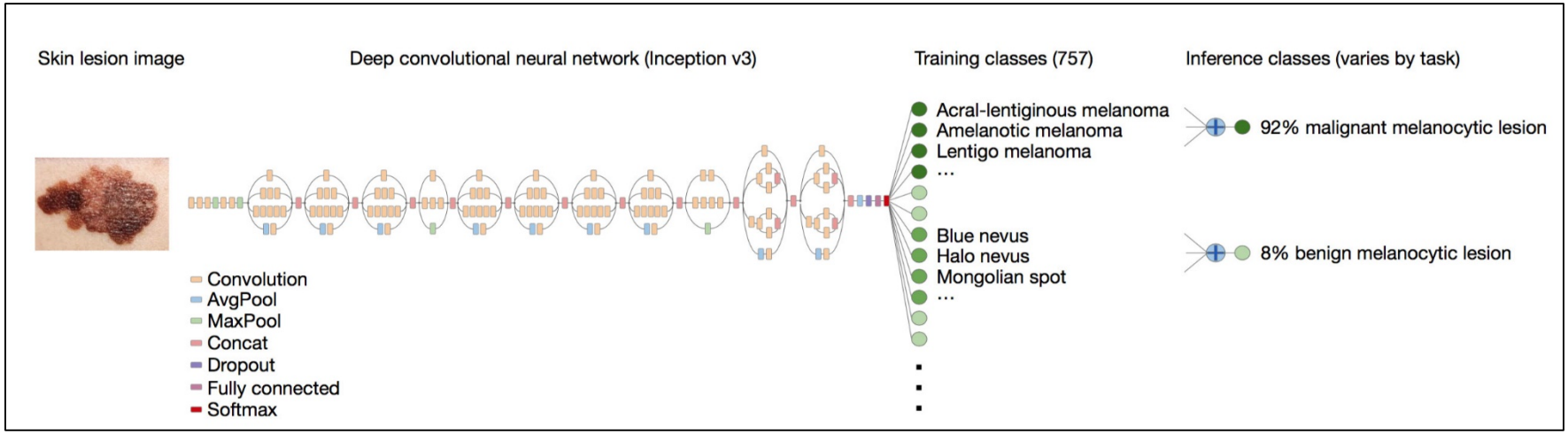
Amazon Picking Challenge
http://www.robocup2016.org/en/events/amazon-picking-challenge/

Amazon Prime Air

Medical imaging



3D imaging
(MRI, CT)



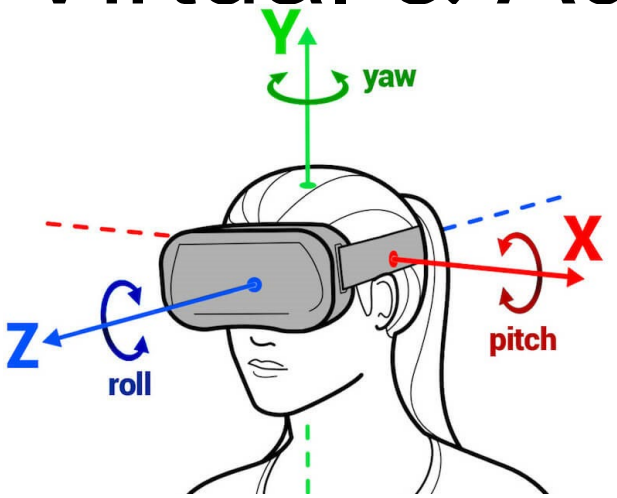
Skin cancer classification with deep learning
https://cs.stanford.edu/people/esteva/nature/

INVESTING 3/25/2014 @ 5:43PM | 70,399 views

Facebook Buys Oculus, Virtual Reality Gaming Startup, For $2 Billion

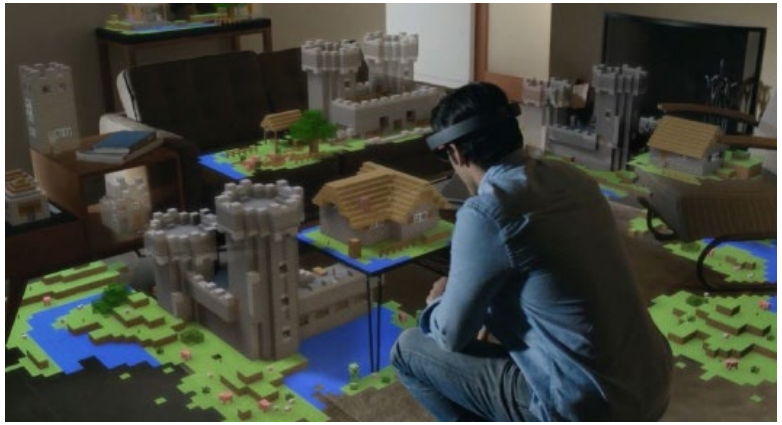
+ Comment Now + Follow Comments

Virtual & Augmented Reality



6DoF head tracking

Hand & body tracking



3D scene understanding

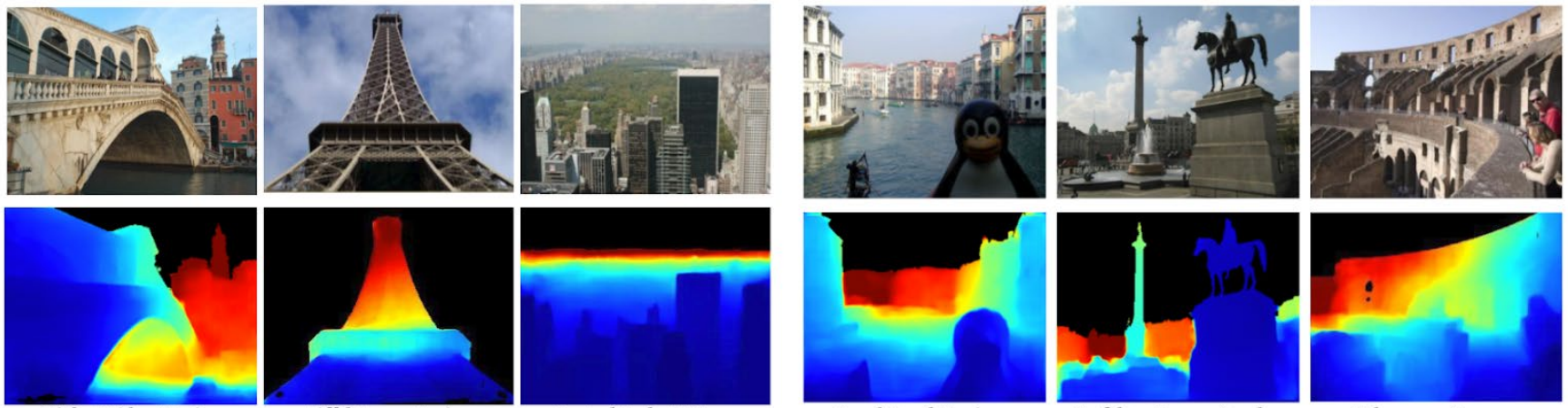
3D-360 video capture

Photosynth

City-scale reconstruction

Reconstruction of Dubrovnik, Croatia, from ~40,000 images

Depth from a single image



Current state of the art

- You just saw many examples of current systems.
  - Many of these are less than 5 years old
- This is a very active research area, and rapidly changing
  - Many new apps in the next 5 years
  - Deep learning powering many modern applications
- Many startups across a dizzying array of areas
  - Deep learning, robotics, autonomous vehicles, medical imaging, construction, inspection, VR/AR, …

Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

Why is computer vision difficult?

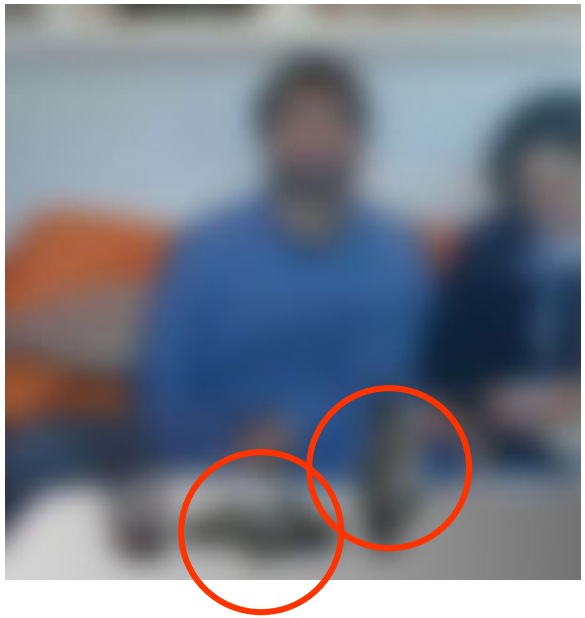
Intra-class variation

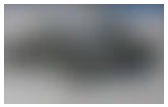
Motion (Source: S. Lazebnik)

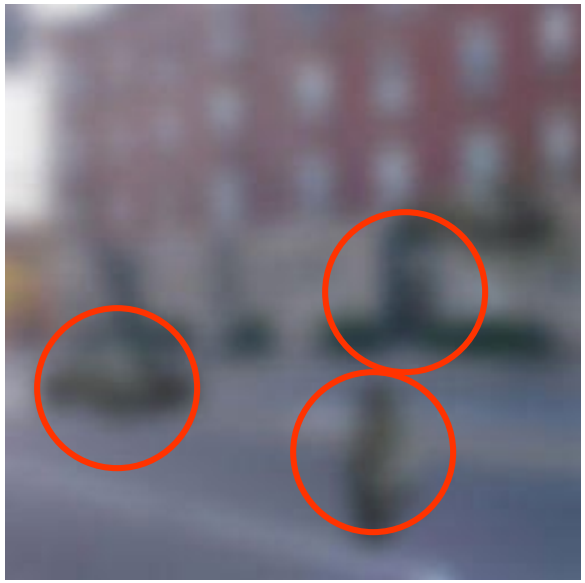
Background clutter

Occlusion

Challenges: local ambiguity







slide credit: Fei-Fei, Fergus & Torralba

But there are lots of cues we can exploit…

Source: S. Lazebnik

But there are lots of cues we can exploit…

Source: S. Lazebnik

Depth cues: Aerial perspective

Depth ordering cues: Occlusion

Source: J. Koenderink

Shape cues: Texture gradient

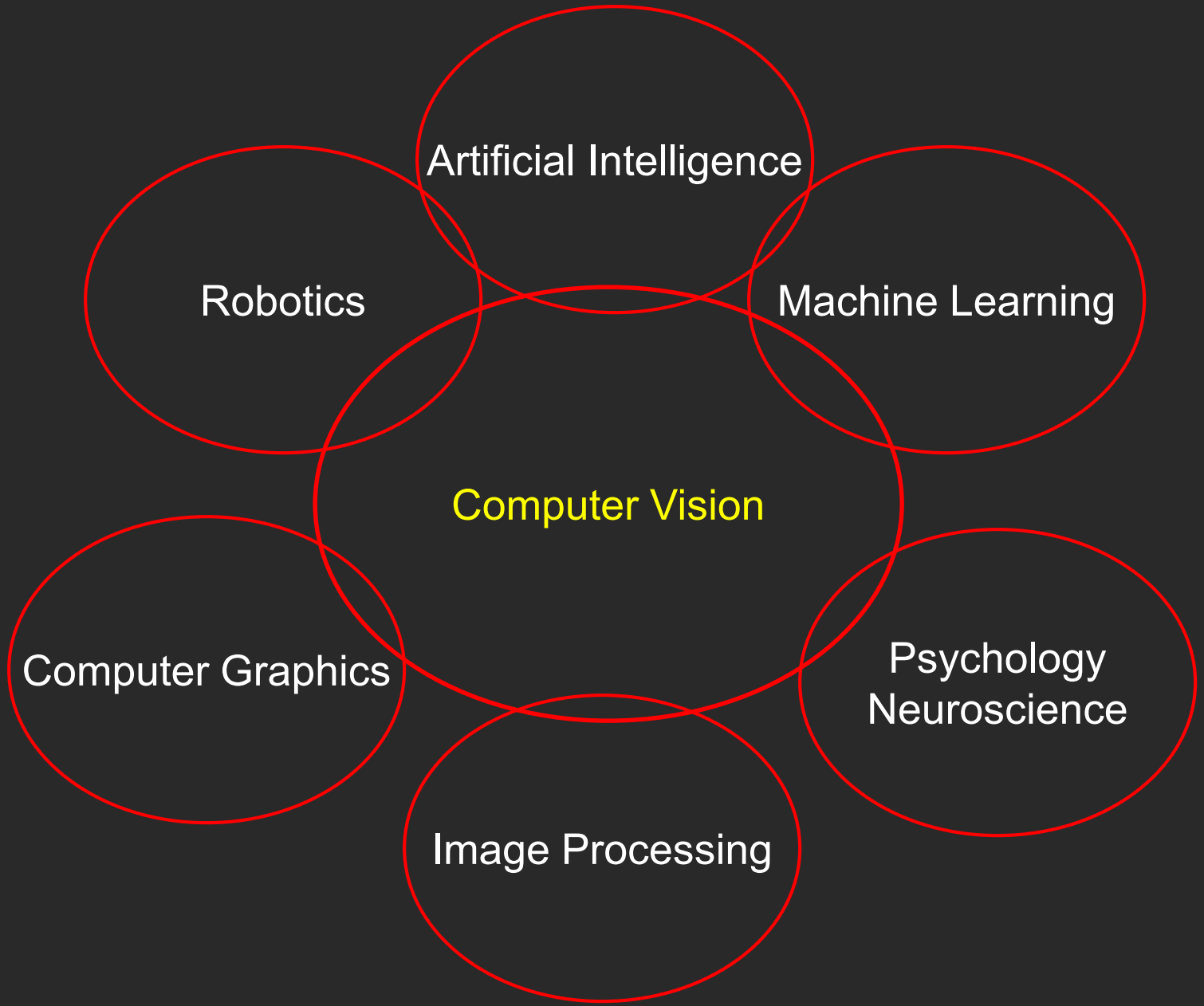
Bottom line

- Perception is an inherently ambiguous problem
  - Many different 3D scenes could have given rise to a particular 2D picture

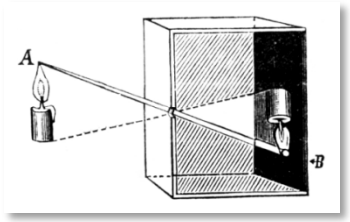
- Possible solutions
  - Bring in more constraints (more images)
  - Use prior knowledge about the structure of the world
- Need a combination of different methods

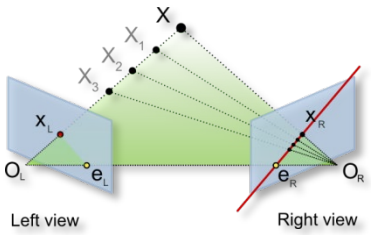
Image source: F. Durand

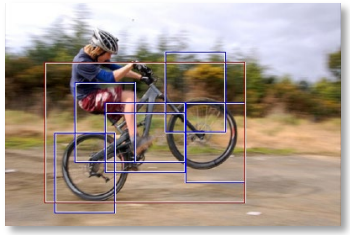
Connections to other disciplines



Course overview



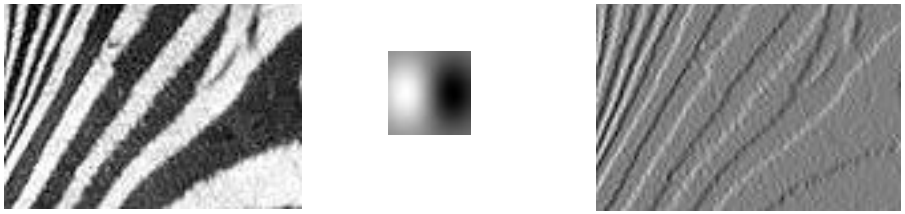




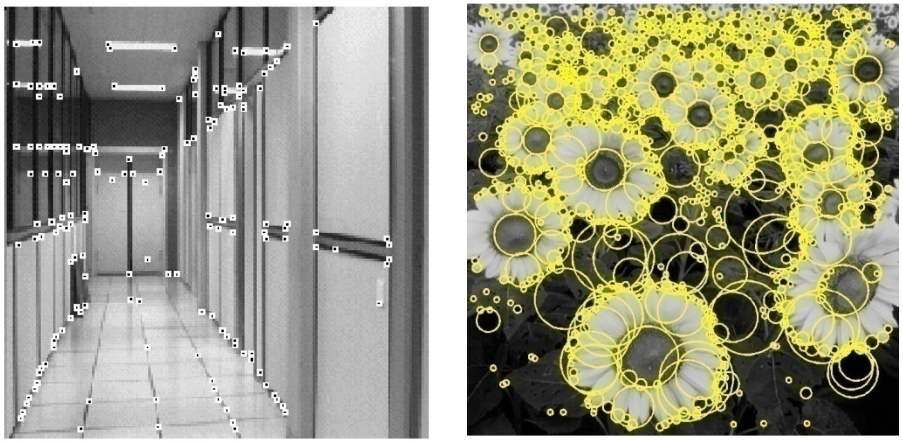
1. Low-level vision
   - image processing, edge detection, feature detection, cameras, image formation
2. Geometry and algorithms
   - projective geometry, stereo, structure from motion, optimization
3. Recognition
   - face detection / recognition, category recognition, segmentation

1. Low-level vision

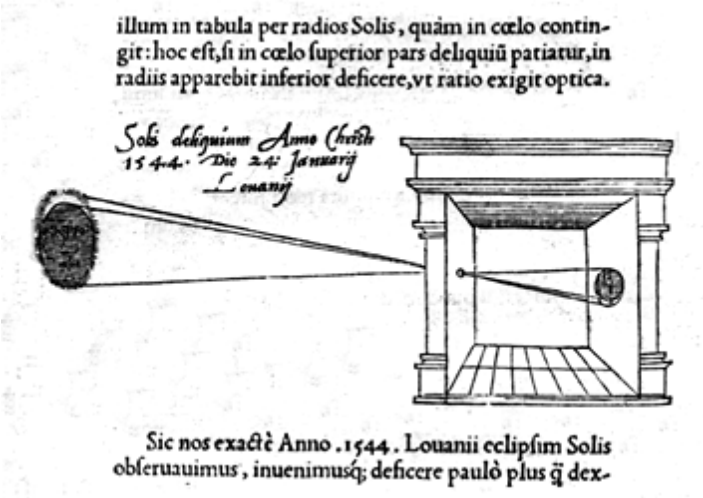
- Basic image processing and image formation

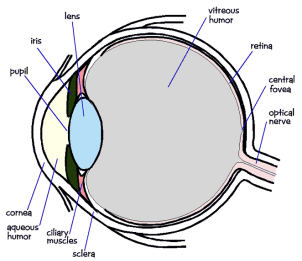


Filtering, edge detection



Feature extraction





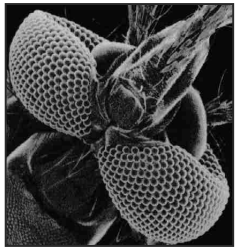
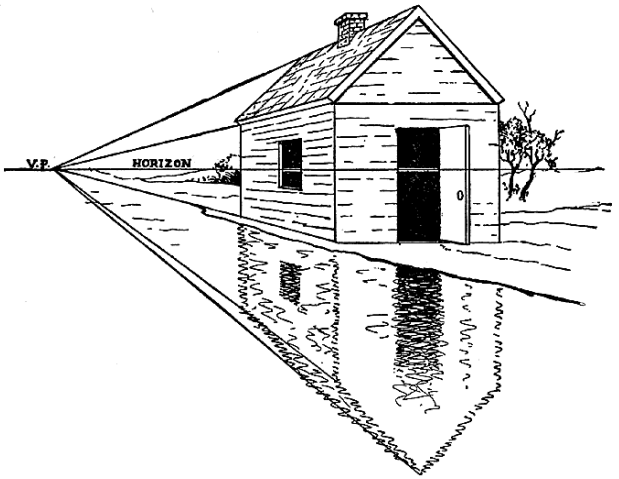
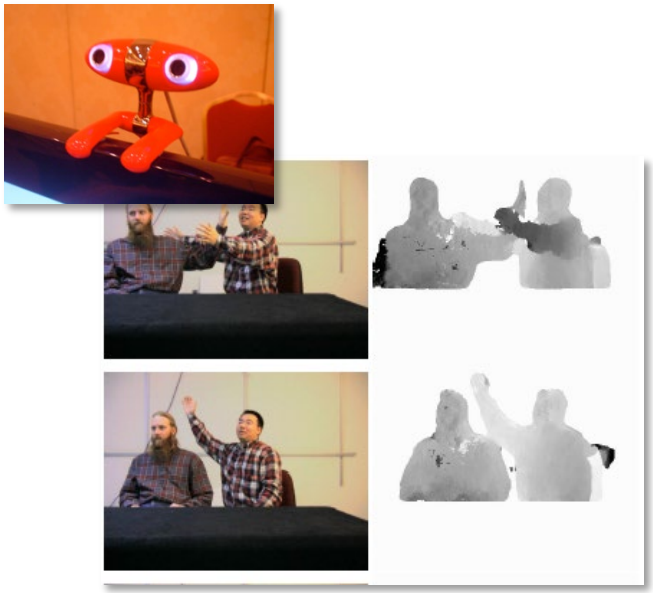


Image formation

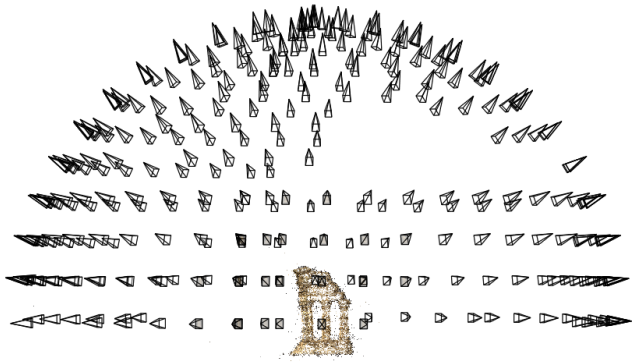
2. Geometry



Projective geometry



Stereo

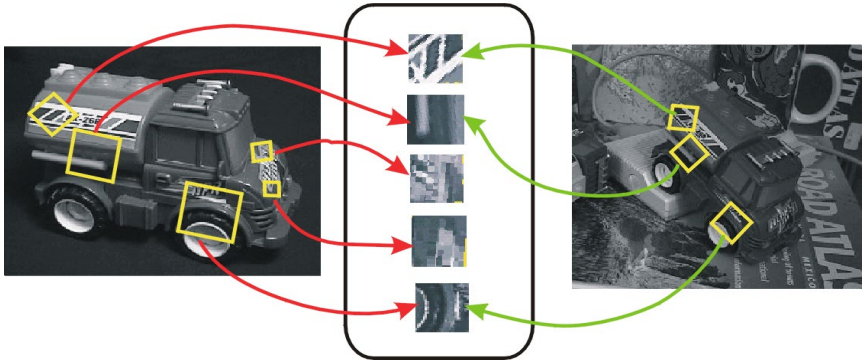


Multi-view stereo

Structure from motion

3. Recognition

Face detection and recognition



Single instance recognition

Category recognition