



CSE-408 (Digital Image Processing)
Instructor: Dr. Muhammad Abeer Irfan
DCSE

Have you ever used Image processing and computer vision?

Where?

How?

Have you ever used computer vision?

How? Where?

Reconstruction? Recognition? (Re)organization?

Think-Pair-Share

Laptop: Biometrics auto-login (face recognition, 3D), OCR

Smartphones: QR codes, computational photography (Android Lens Blur, iPhone Portrait Mode), panorama construction (Google Photo Spheres), face detection, expression detection (smile), Snapchat filters (face tracking), FaceID (iPhone), Night Sight (Pixel), iPhone 12 Pro (LiDAR)

Web: Image search, Google photos (face recognition, object recognition, scene recognition, geolocalization from vision), Facebook (image captioning), Google maps aerial imaging (image stitching), YouTube (content categorization)

VR/AR: Outside-in tracking (HTC VIVE), inside out tracking (simultaneous localization and mapping, HoloLens), object occlusion (dense depth estimation)

Motion: Kinect, full body tracking of skeleton, gesture recognition, virtual try-on

Medical imaging: CAT / MRI reconstruction, assisted diagnosis, automatic pathology, connectomics, endoscopic surgery

Industry: Vision-based robotics (marker-based), machine-assisted router (jig), automated post, ANPR (number plates), surveillance, drones, shopping

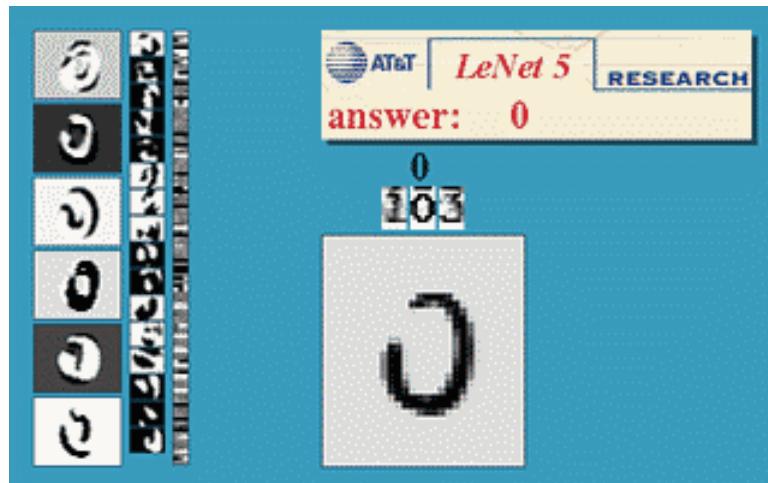
Transportation: Assisted driving (everything), face tracking/iris dilation for drunkenness, drowsiness, automated distribution (all modes)

Media: Visual effects for film, TV (reconstruction), virtual sports replay (reconstruction), semantics-based auto edits (reconstruction, recognition)

Optical character recognition (OCR)

Technology to convert images of text into text

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



Mail digit recognition, AT&T labs
<http://www.research.att.com/~yann/>



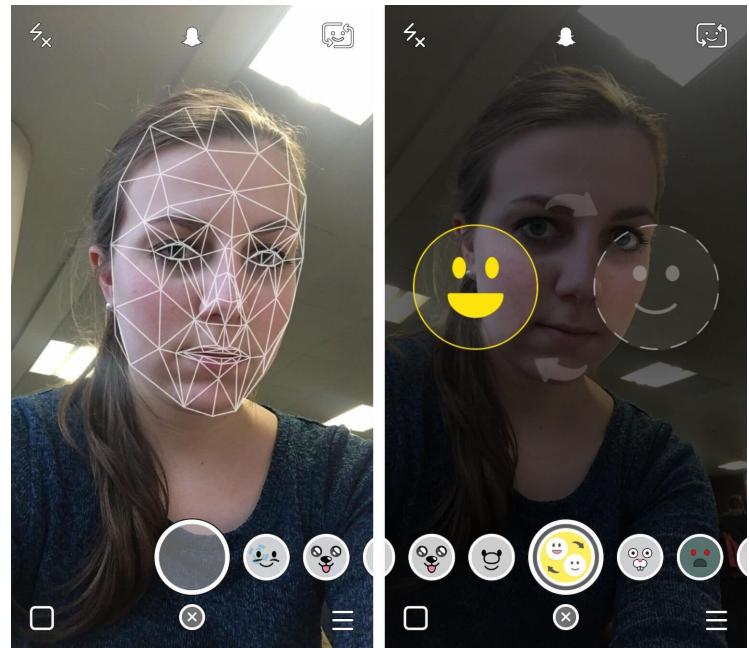
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition



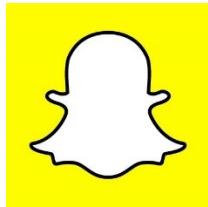
Live
Camera
Translation

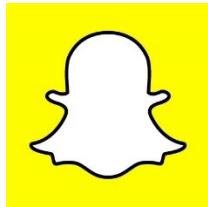


Face detection

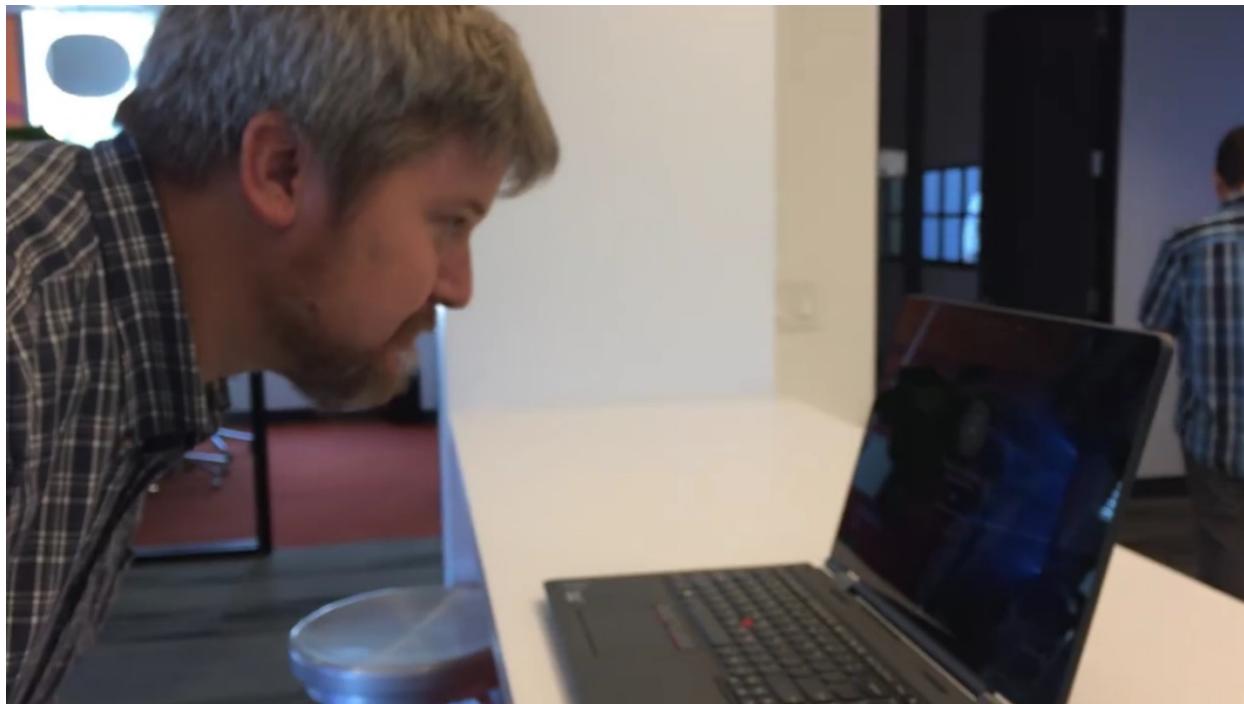


- Almost all digital cameras detect faces
- Snapchat face filters





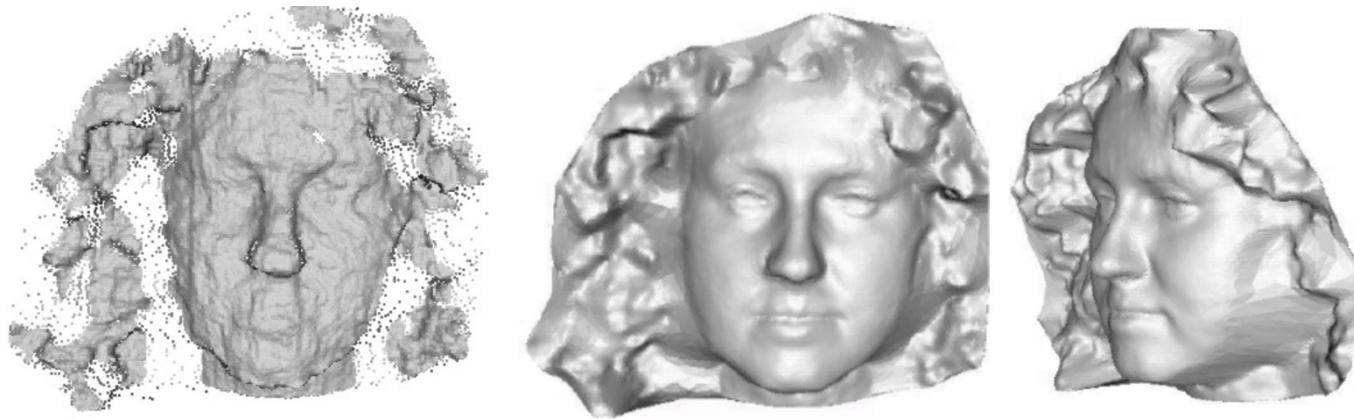
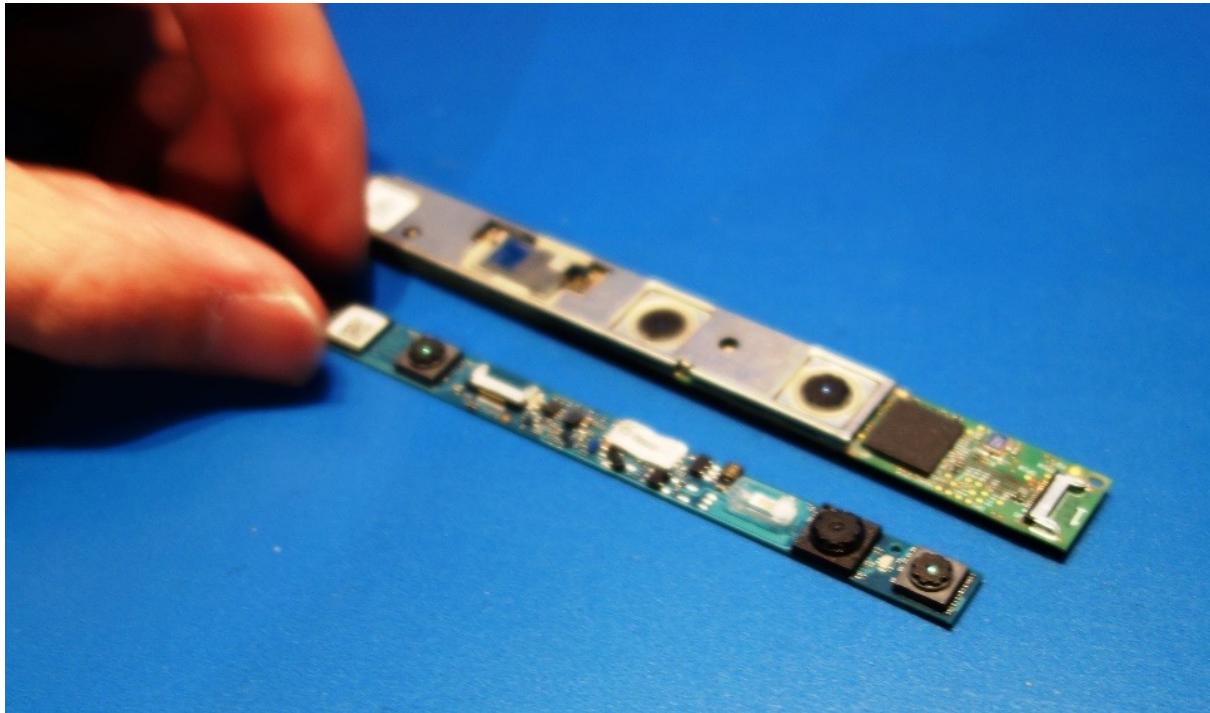
Facial login without a password...



Facial login without a password...



Facial login without a password...

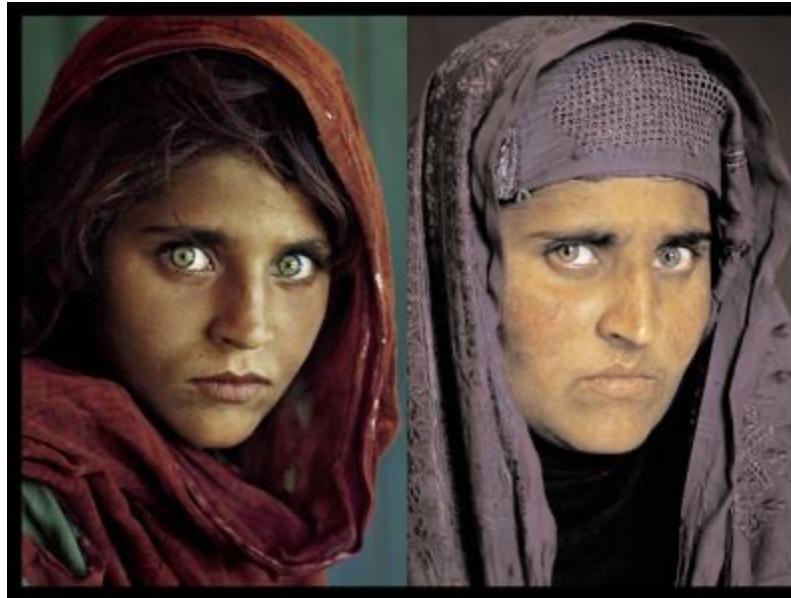


Single depth frame

Reconstructed 3D mesh

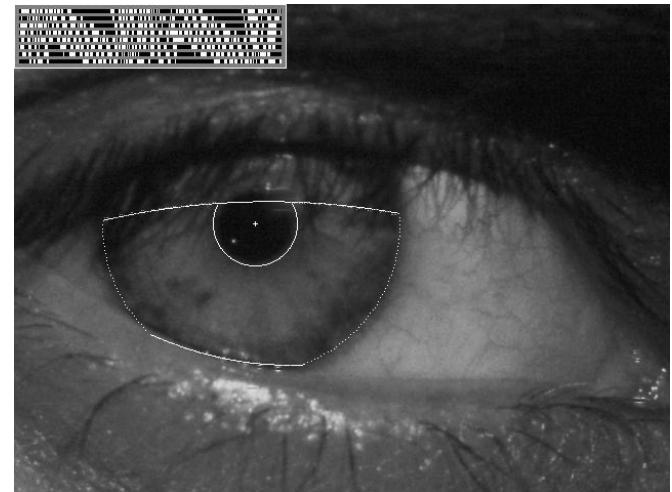
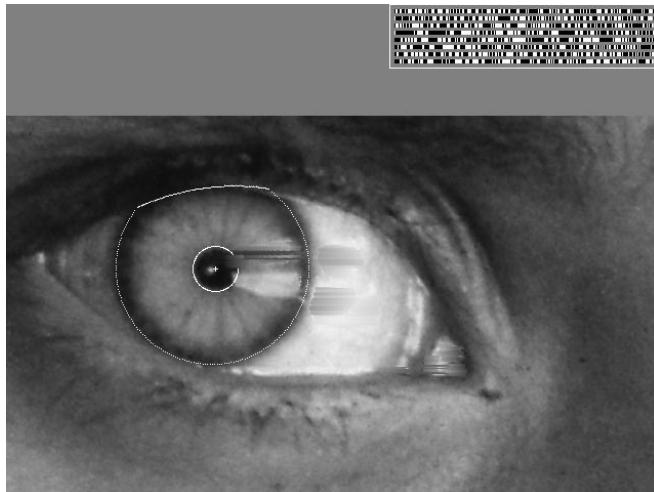
Liang et
al. 2014

Vision-based biometrics



“How the Afghan Girl was Identified by Her Iris Patterns”

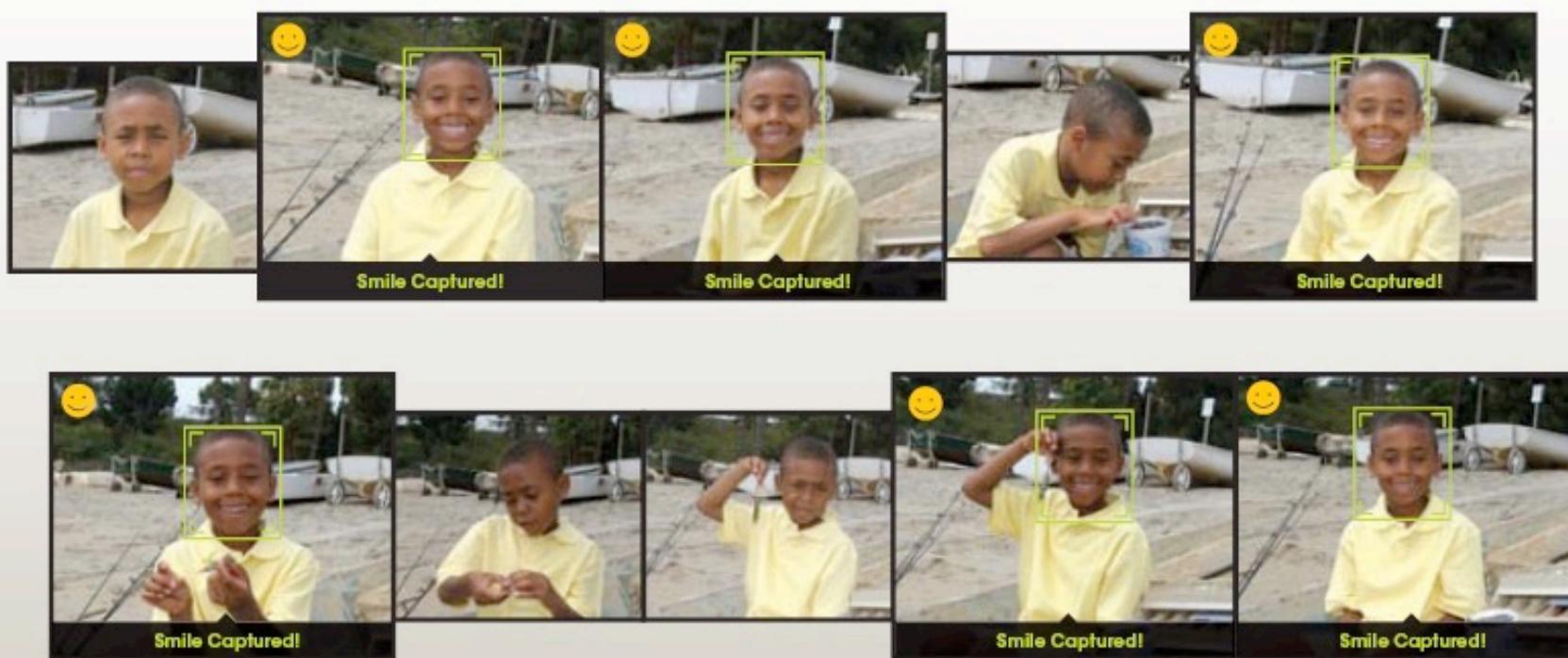
Read the [story \(Wikipedia\)](#)



Smile detection

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



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

Video call eye gaze correction

Kuster et al., SIGGRAPH Asia 2012

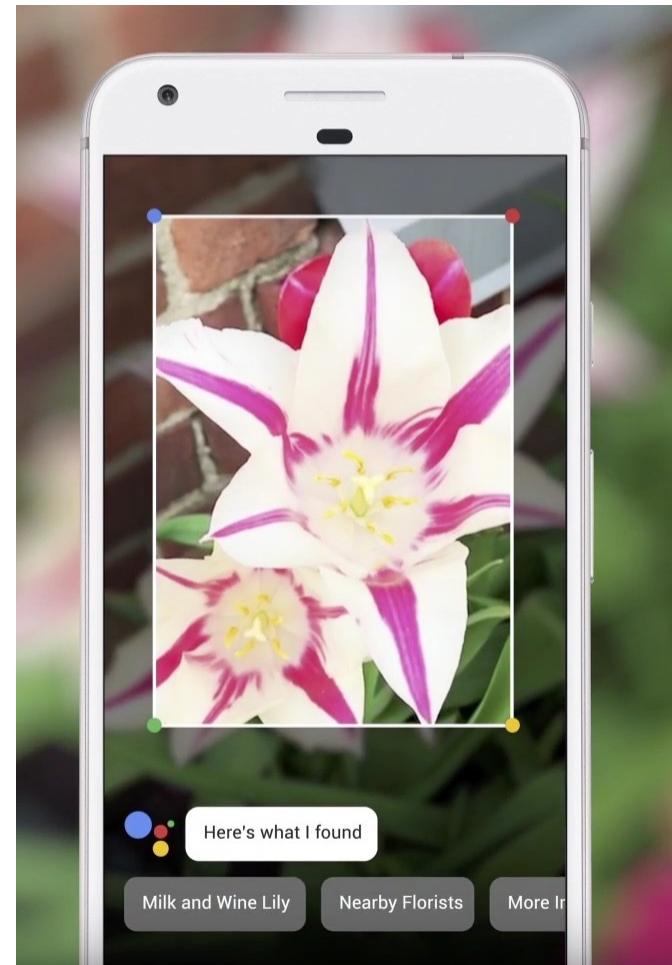
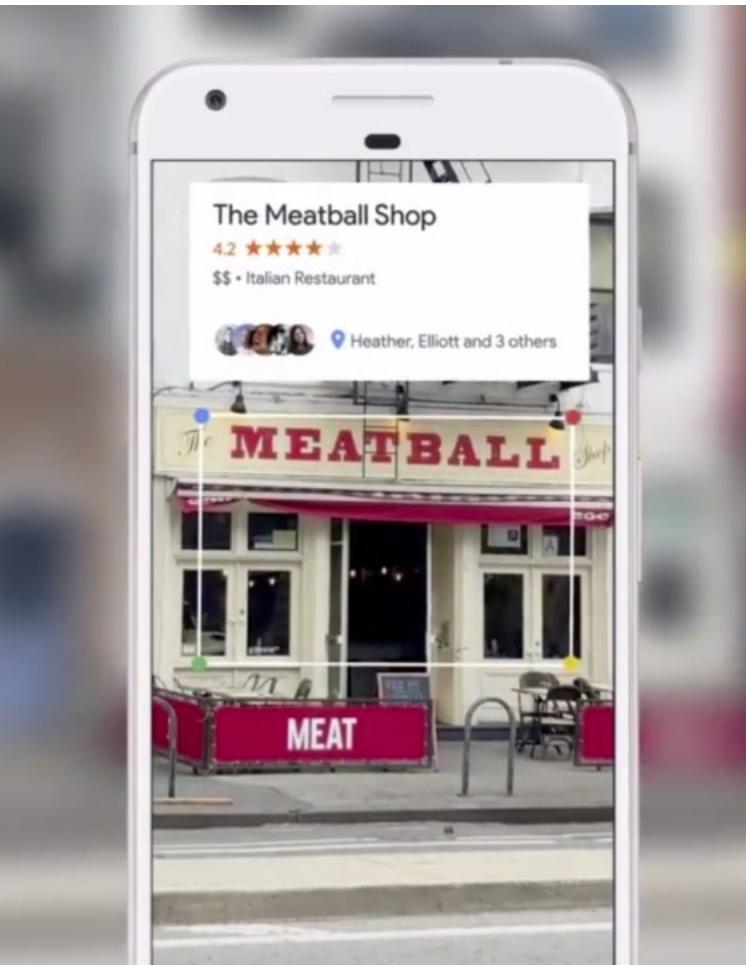
- <https://cgl.ethz.ch/publications/papers/paperKus12.php>

Apple FaceTime
Attention Correction



Object recognition (in mobile phones)

e.g., Google Lens



Object recognition (in supermarkets)



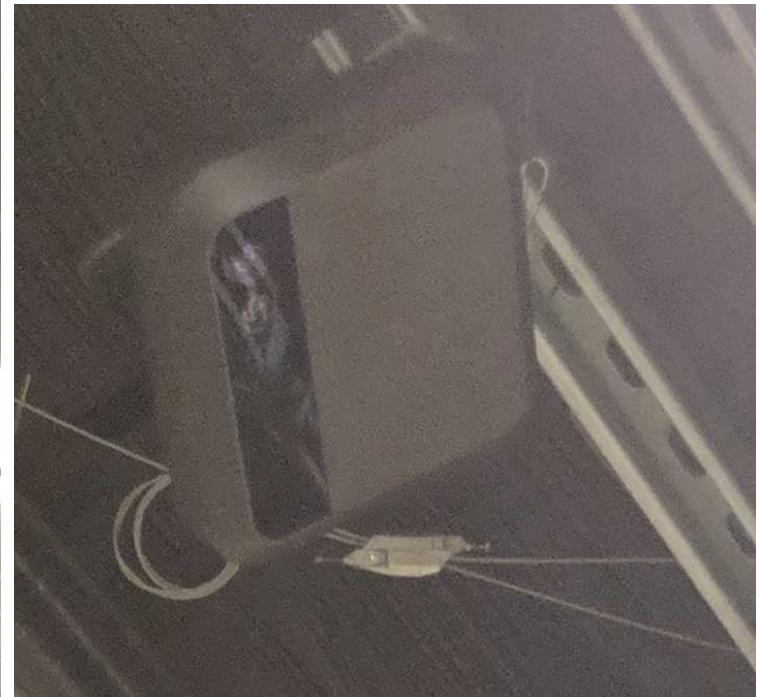
How does it work?

Think-Pair-Share



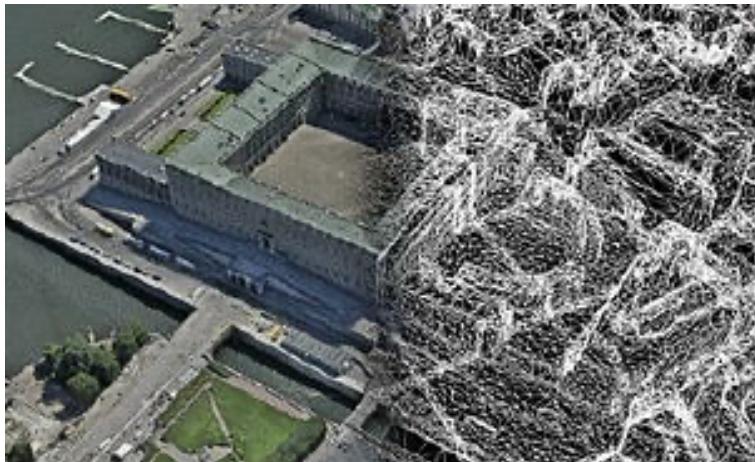
How does it work?





Source: Vivek Ramanujan

3D from images

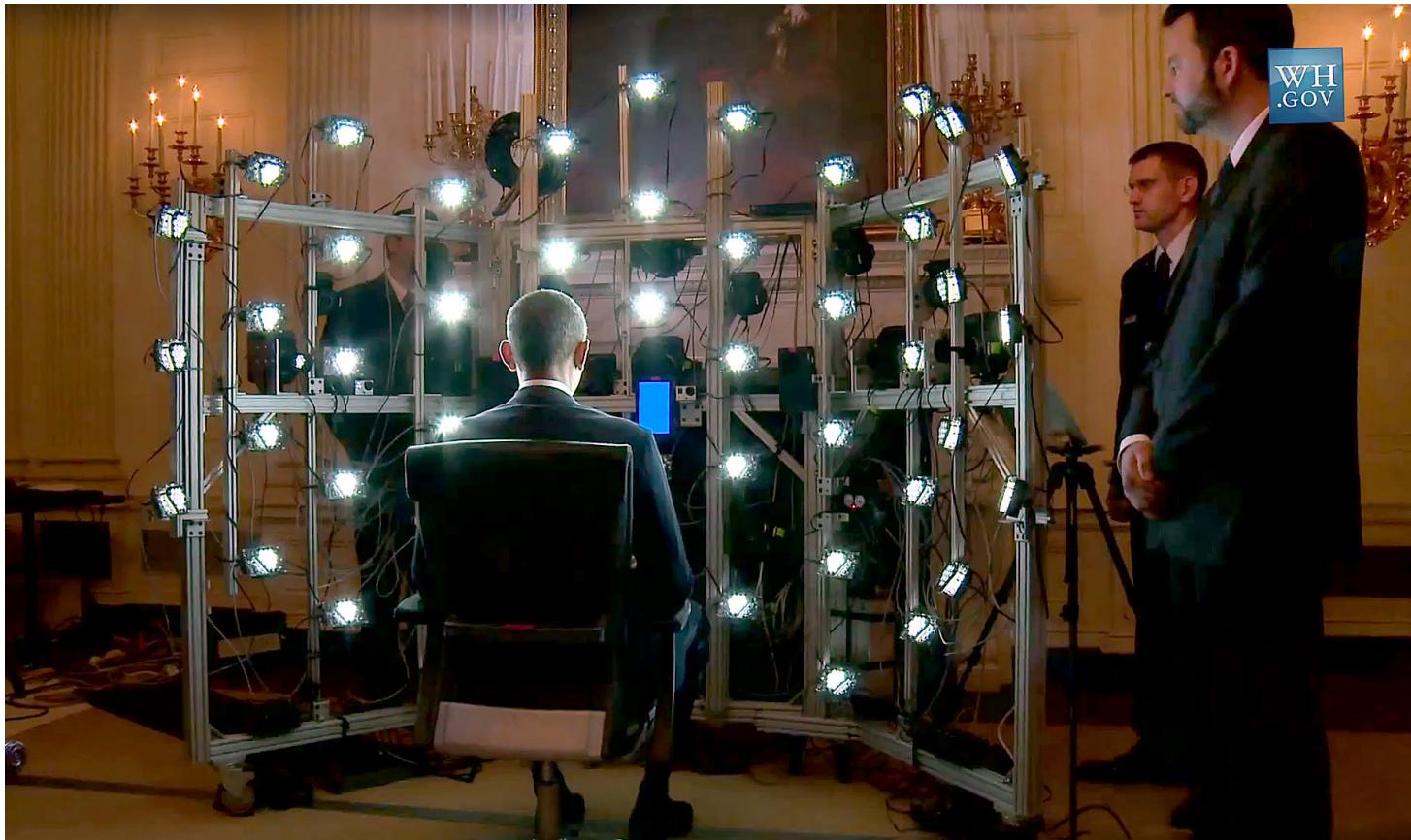


Building Rome in a Day: Agarwal et al. 2009

Human shape capture



Human shape capture



Human shape capture



Human shape capture



Special effects: motion capture



Interactive Games

Object Recognition:

<http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o>

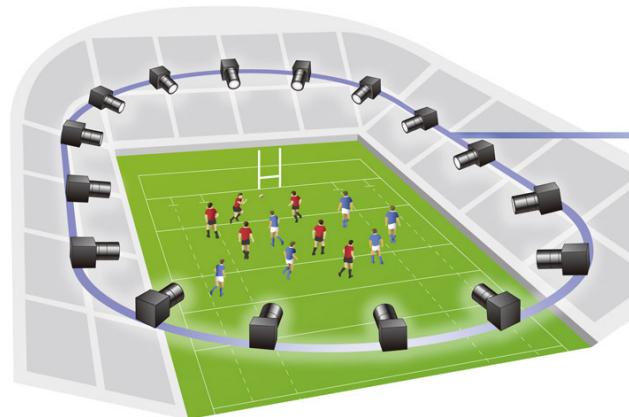
Mario: <http://www.youtube.com/watch?v=8CTJL5IUjHg>

3D: <http://www.youtube.com/watch?v=7QrnwoO1-8A>

Robot: <http://www.youtube.com/watch?v=w8BmgtMKFbY>



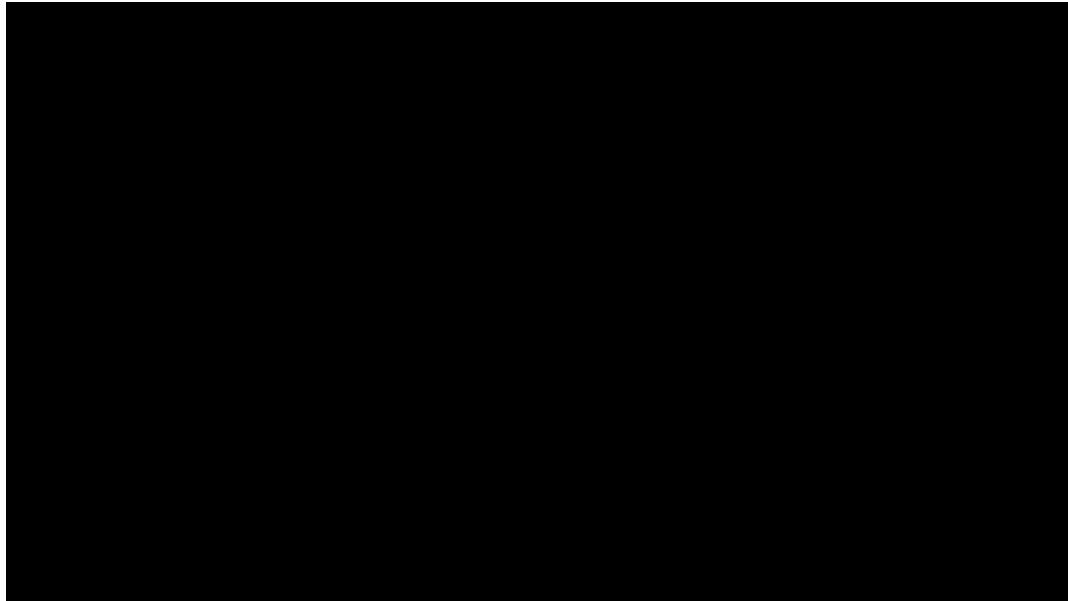
Sports



Virtual pitch markings



Free viewpoint video

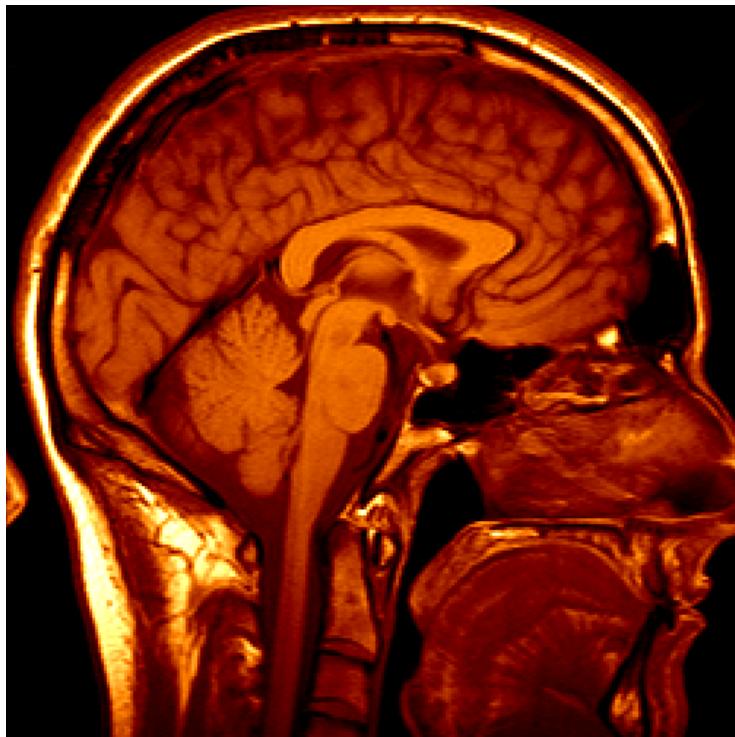


Sportvision first down line

Nice explanation on www.howstuffworks.com

[Canon 2017]

Medical imaging



3D imaging
MRI, CT



Image guided surgery
Grimson et al., MIT

AutoCars - Uber bought CMU's lab (2015) Then sold it (2020)



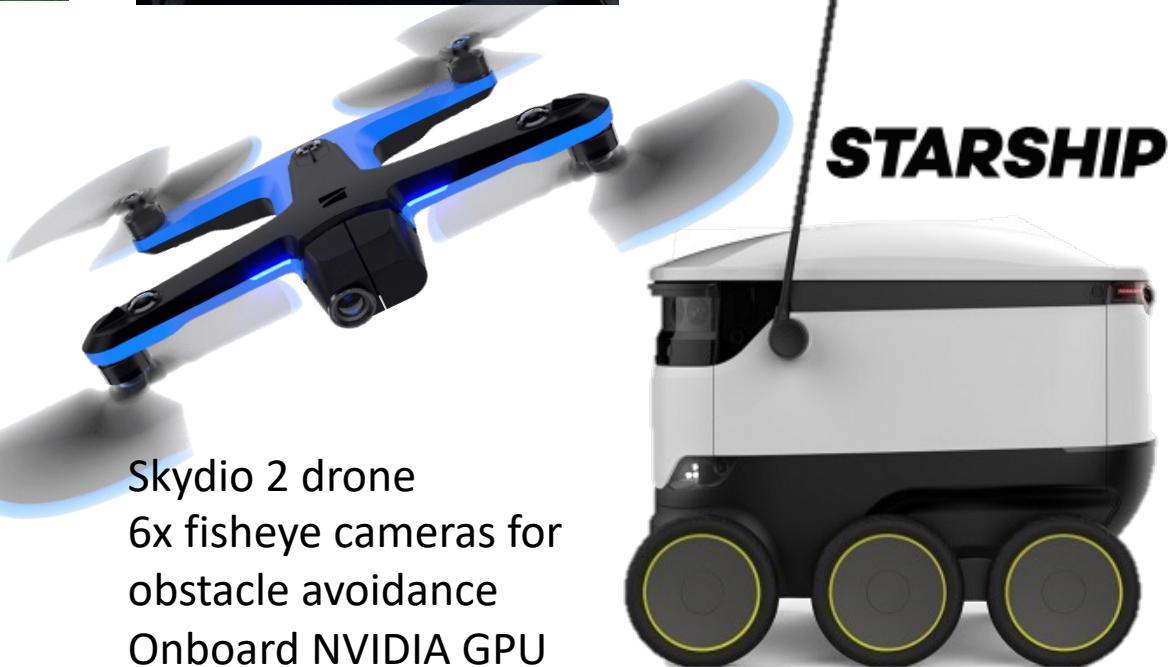


Mobile robots

<http://www.robocup.org/>

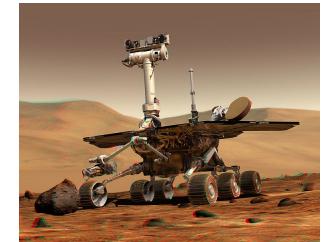


Saxena et al. 2008
[STAIR](#) at Stanford



Skydio 2 drone
6x fisheye cameras for
obstacle avoidance
Onboard NVIDIA GPU

Vision in space

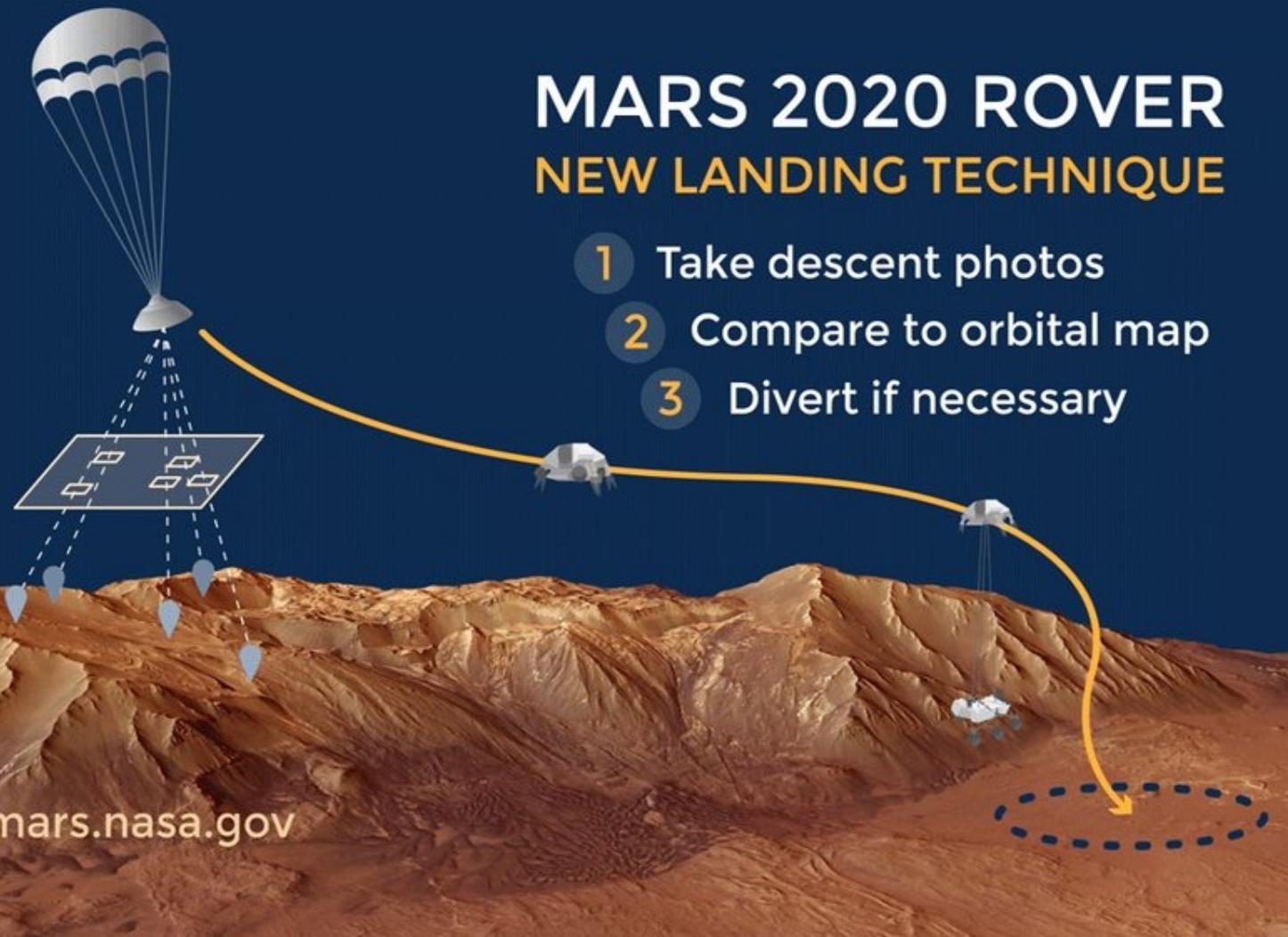


[NASA'S Mars Exploration Rover Spirit](#) captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

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.

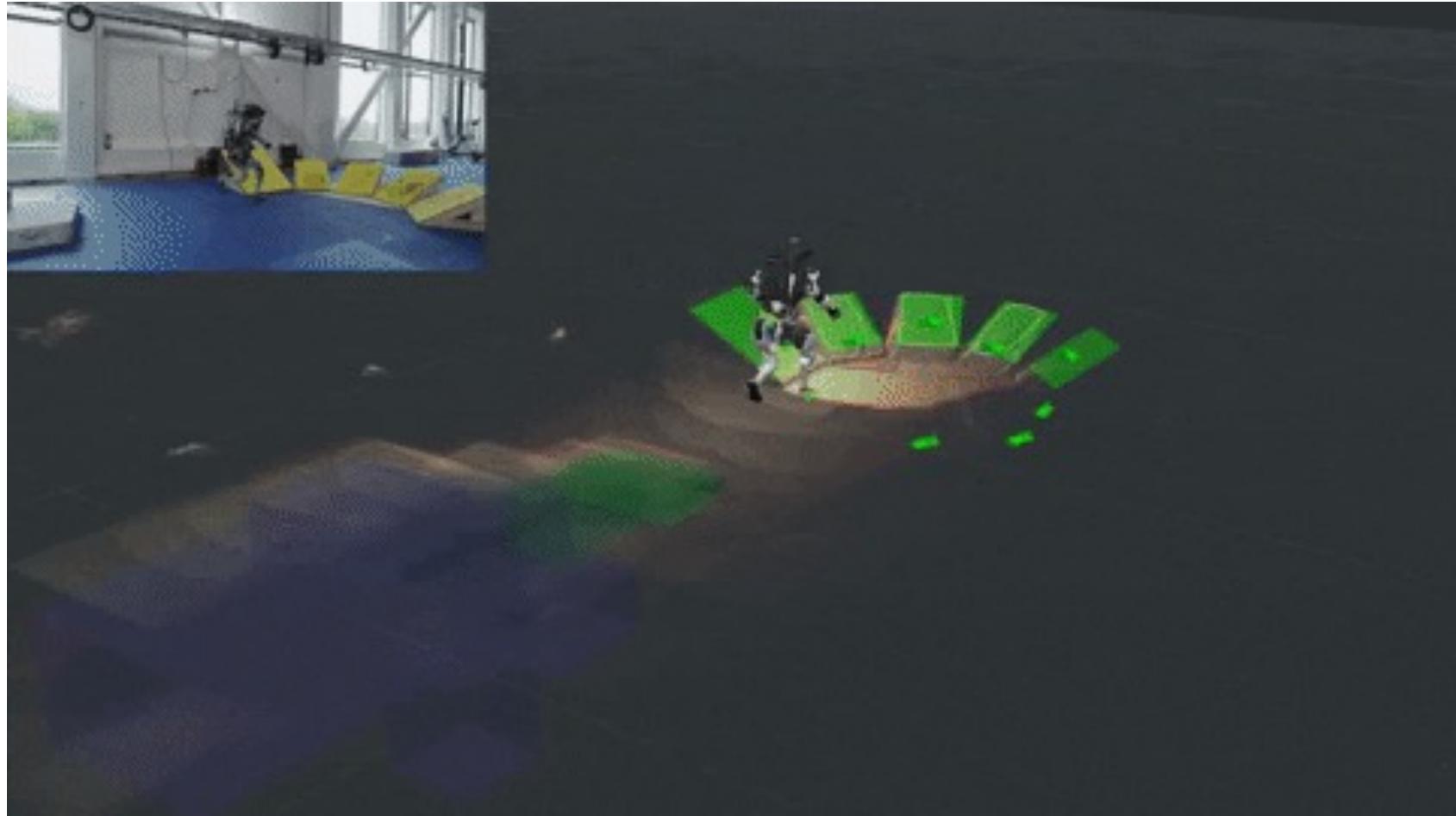
NASA Perseverance lander and rover





Landed 18th February 2021

Humanoid Robots



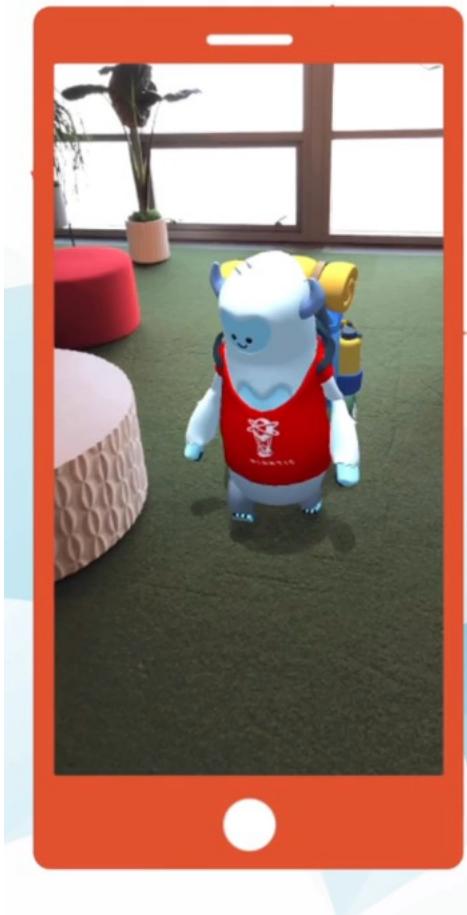
Augmented Reality and Virtual Reality



MS HoloLens, Oculus, Magic Leap,
ARCore / ARKit

Augmented Reality and Virtual Reality

Real-time monocular depth estimation and camera tracking



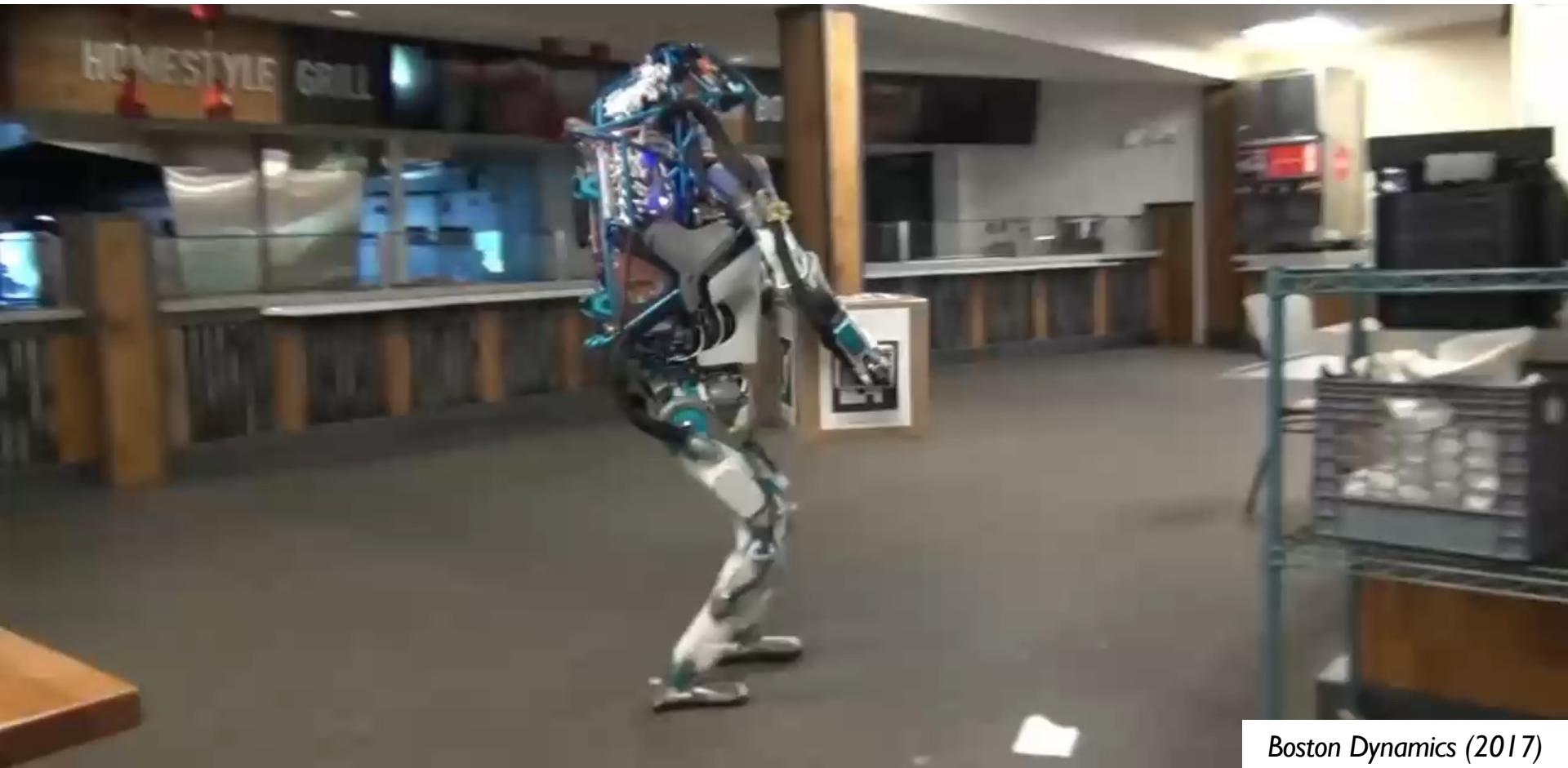
Real-time 3D hand pose estimation



Oculus (Quest)

Niantic

AI for Physical Interaction



Boston Dynamics (2017)

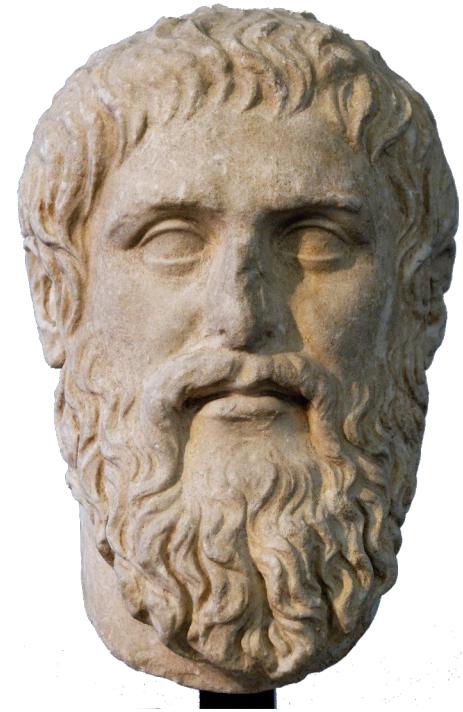
COURSE SCOPE

Vision and Society

“Worldview” vs. “World-sense”

“Vision, in my view, is the cause of the **greatest benefit** to us, inasmuch as none of the accounts now given concerning the Universe would ever have been given if men had not seen the stars or the sun or the heavens.”

- Plato (Timeus, 360 BC)



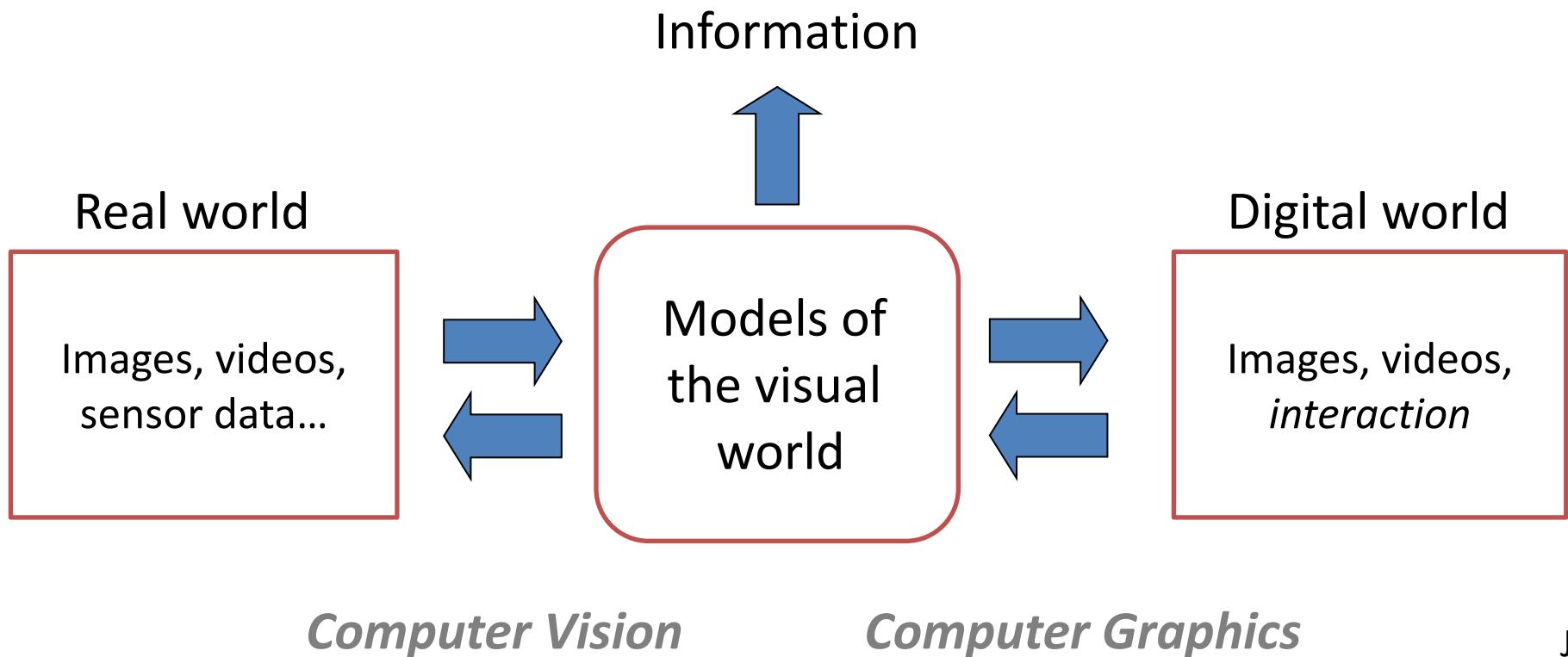
Computer Vision and Nearby Fields

Derogatory summary of computer vision:
“Machine learning applied to visual data.”

Computer Vision and Nearby Fields

Derogatory summary of computer vision:

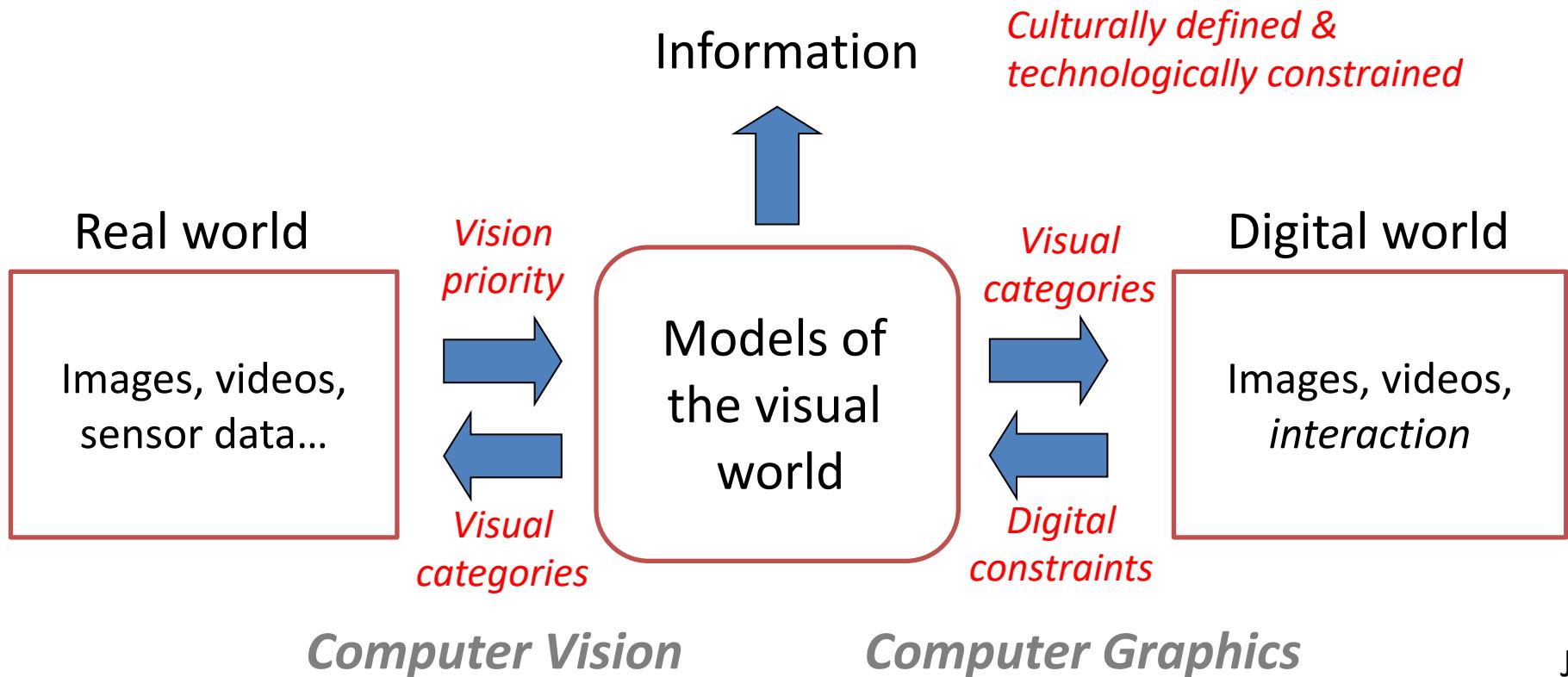
~~“Machine learning applied to visual data.”~~



Computer Vision and Nearby Fields

Derogatory summary of computer vision:

“Machine learning applied to visual data.”



Superhuman state of the art?

Deep learning is an enormous disruption to the field.
Since 2012, rapid expansion and commercialization.

Why?

“With enough data, computer vision matches or even outperforms human vision at most recognition tasks.”

WHAT.

Vision and Society

Lots of data = lots of potential bias in the data.

Needs understanding of possible failures.

+

Responsible approach.

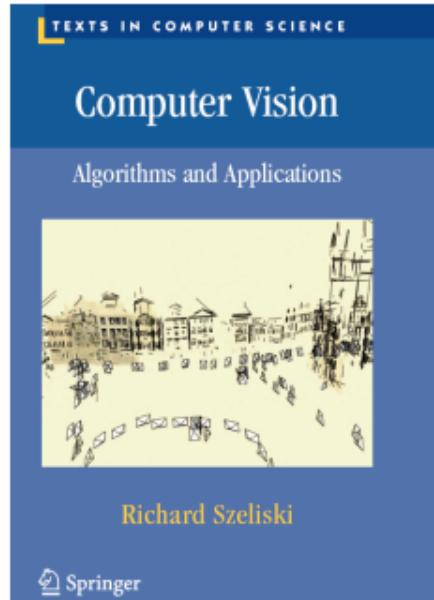
+

Techniques to overcome bias.

Textbooks

Computer Vision: Algorithms and Applications

© 2010 [Richard Szeliski](#), Microsoft Research



<http://szeliski.org/Book/>

Textbooks

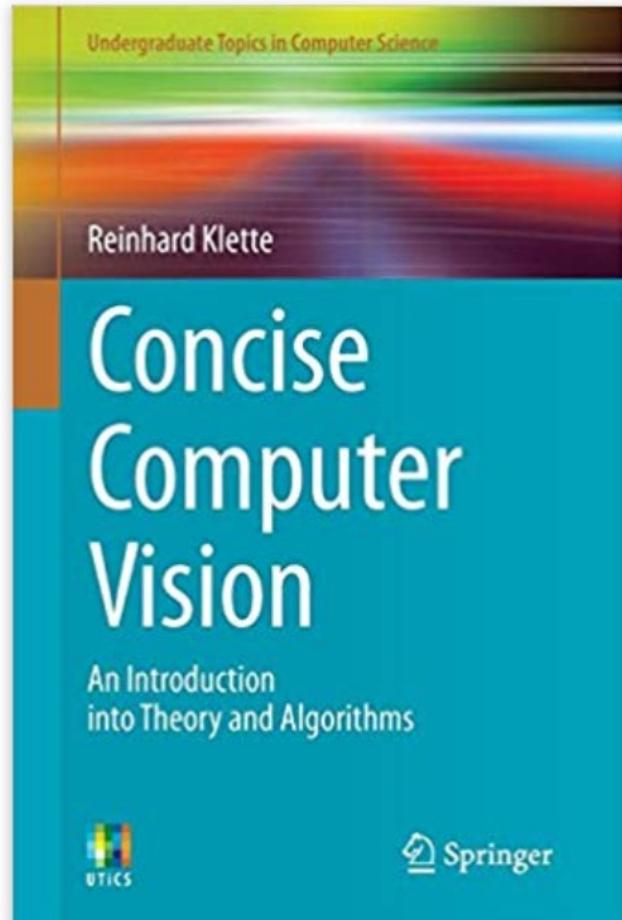
<https://ccv.wordpress.fos.auckland.ac.nz/>

Klette

Concise Computer Vision

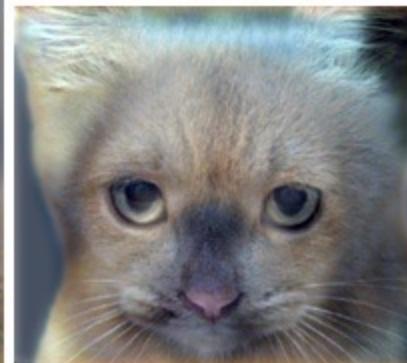
Digital copy at our library

More introductory than
Szeliski.



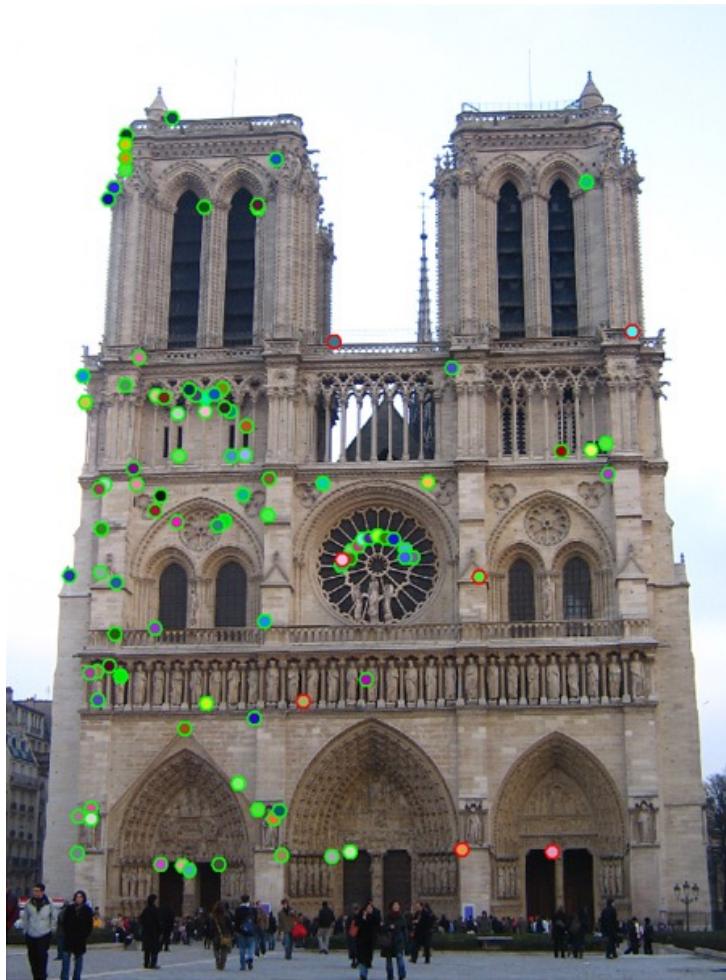
HW 1: Image Filtering and Hybrid Images

- Implement image filtering to separate high and low frequencies.
- Combine high frequencies and low frequencies from different images to create a scale-dependent image.



HW 2: Local Feature Matching

- Implement interest point detector, SIFT-like local feature descriptor, and simple matching algorithm.



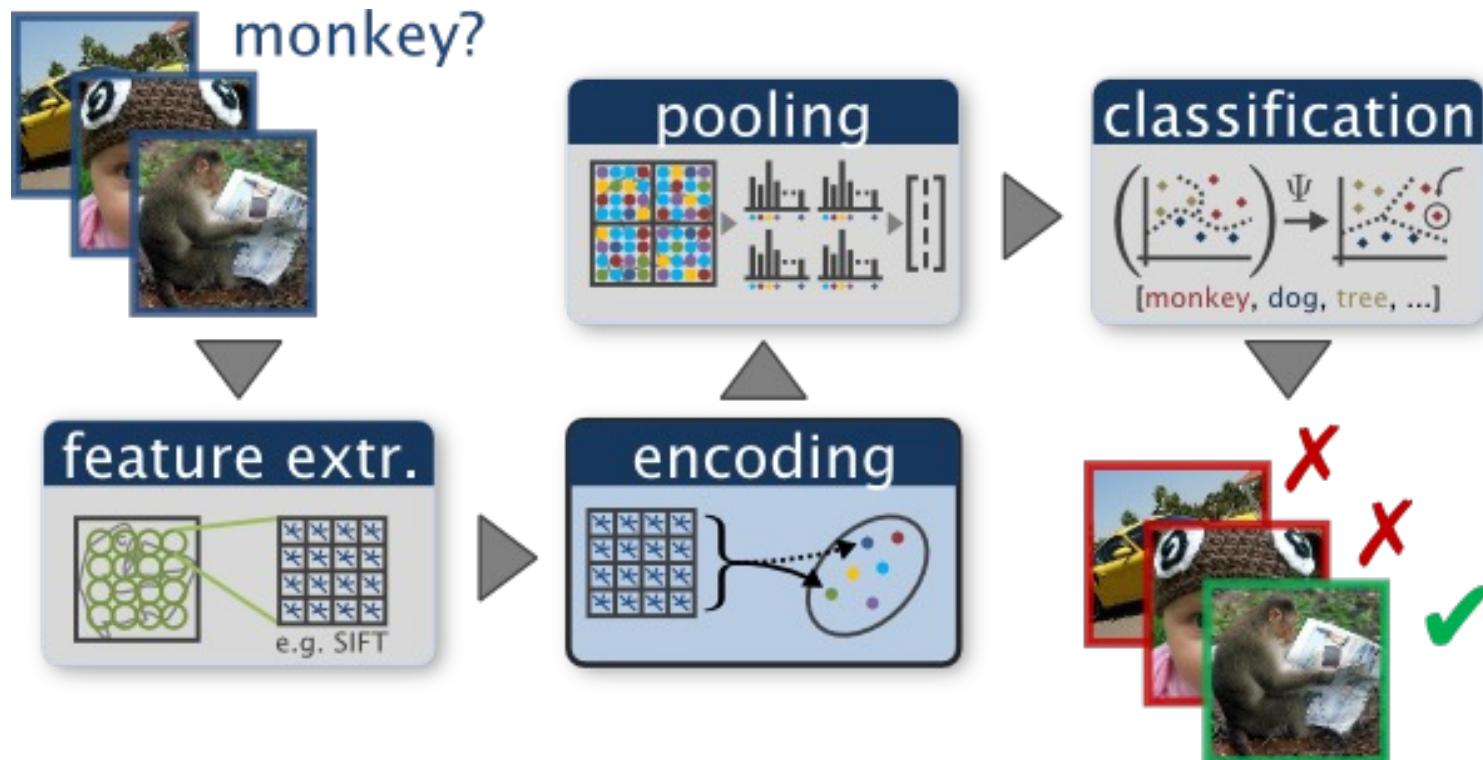
HW 3: Multi-view Geometry

- Recover camera calibration from feature point matches.
- Foundation for almost all measurement in computer vision.



HW 4: Scene Recognition with Bag of Words

- Quantize local features into a “vocabulary”, describe images as histograms of “visual words”, train classifiers to recognize scenes based on these histograms.



Group Final Project

- Free choice with a set of suggested projects
- Up to four people
- Go wild

Project examples (more shown later!)

- Real-time eye tracking
- Multi-view geometry reconstruction
- Computational photography
- Style Transfer

Hyperlapse Stabilization - Michael Mao, Jiaju Ma, James Li

Baseline Video



Result Video



Card Recognition for Bridge Bidding



4:18



Results

Spades: A10

Hearts: AKQ5

Diamonds: K108

Clubs: AKQ9

Bid: 2c

Points: 25

[Take another photo](#)



Slow-Motion Video Interpolation

