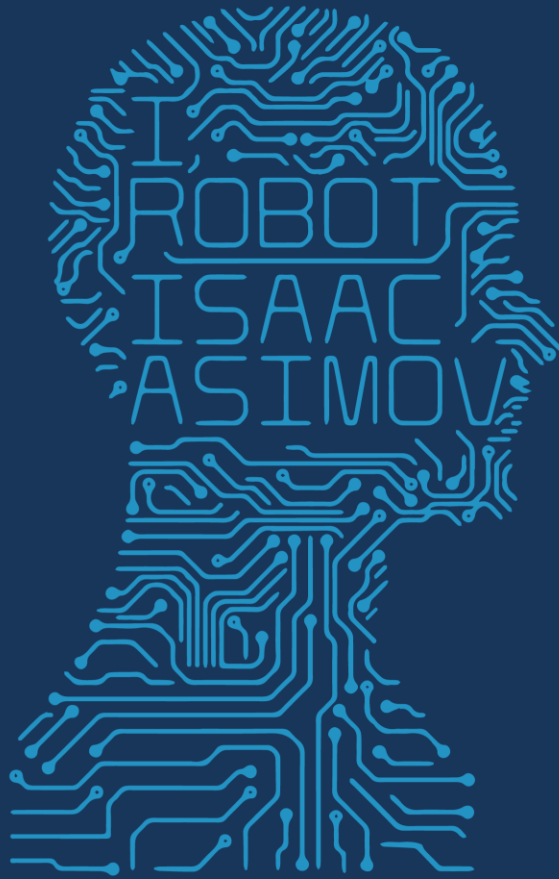




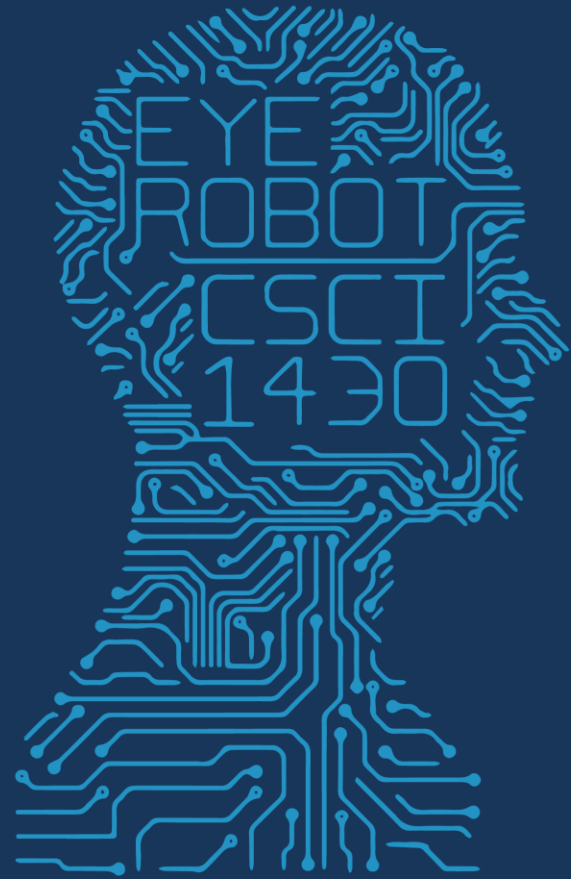
1950

FUTURE VISION



1950

FUTURE VISION



23 JANUARY 2019

COMPUTER VISION

Have you ever used computer vision?

How? Where?

Think-Pair-Share

Jitendra Malik, UC Berkeley
Three 'R's of Computer Vision



Jitendra Malik, UC Berkeley
Three 'R's of Computer Vision



“The classic problems of computational vision:
reconstruction
recognition
(re)organization.”

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), Google Tango (3D reconstruction), Night Sight (Pixel)

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

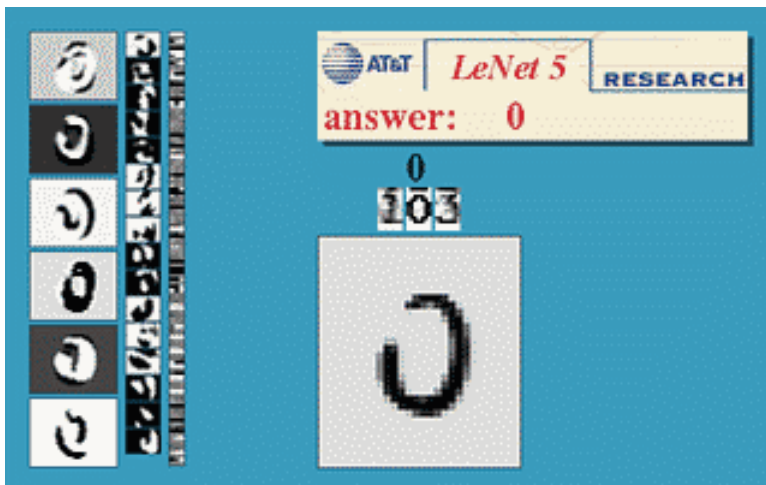
Transportation: Assisted driving (everything), face tracking/iris dilation for drunkenness, drowsiness, automated distribution (all modes)

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

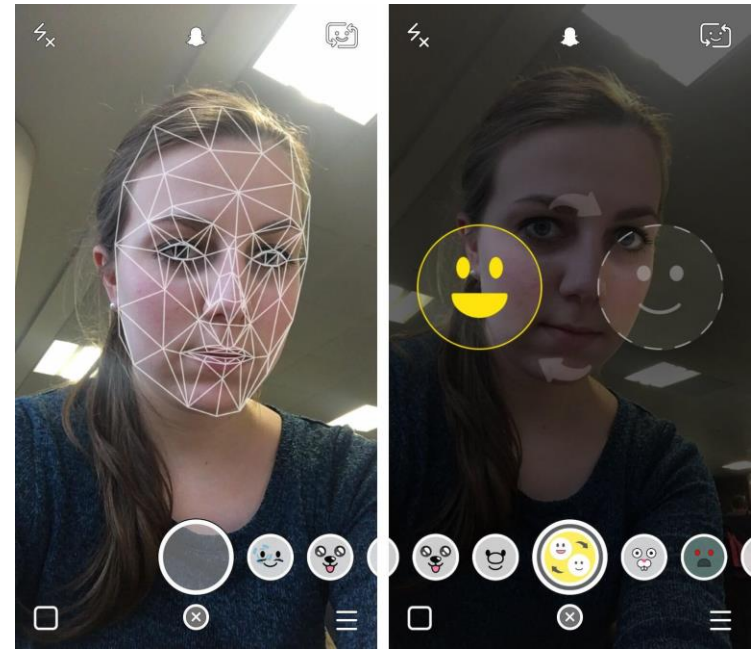
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition



Live
Camera
Translation



Face detection



- Almost all digital cameras detect faces
- Snapchat face filters

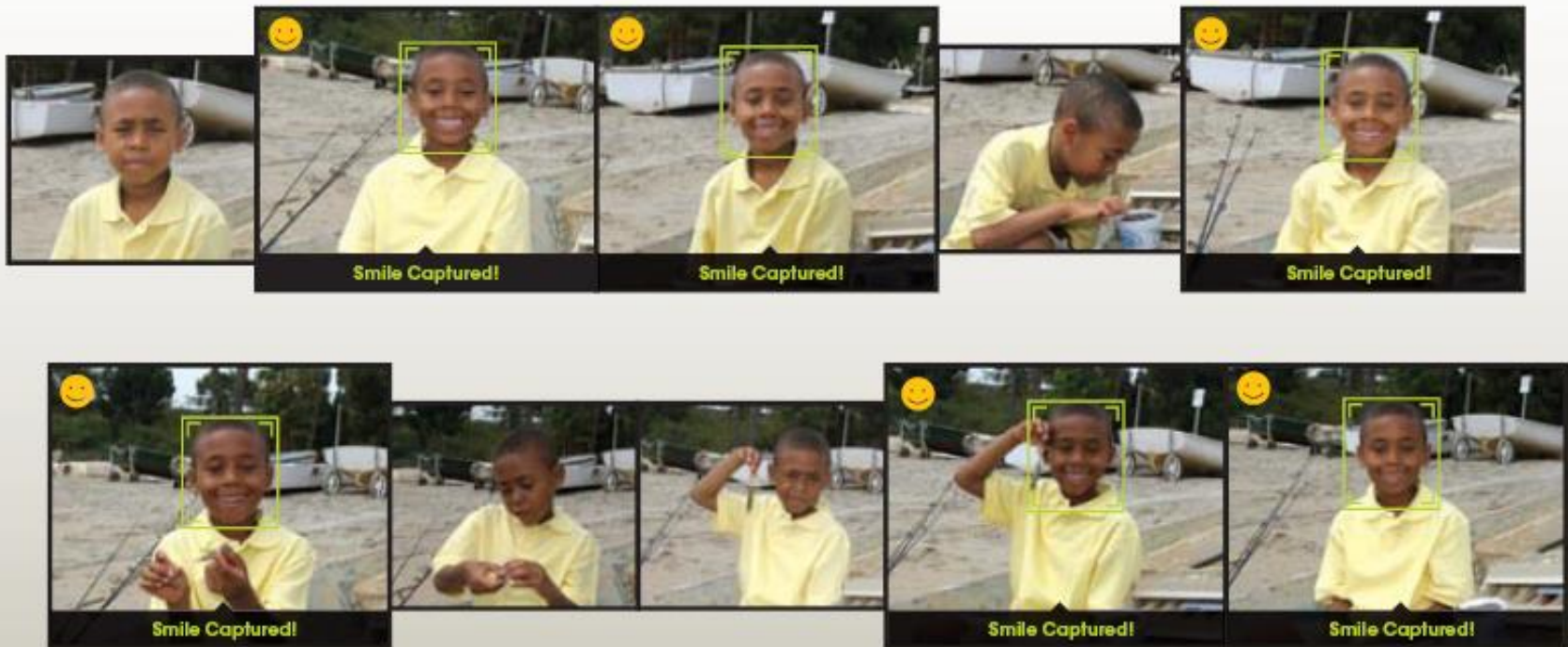




Smile detection

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](#)

Object recognition (in supermarkets)



How does it work?

Think-Pair-Share



How does it work?





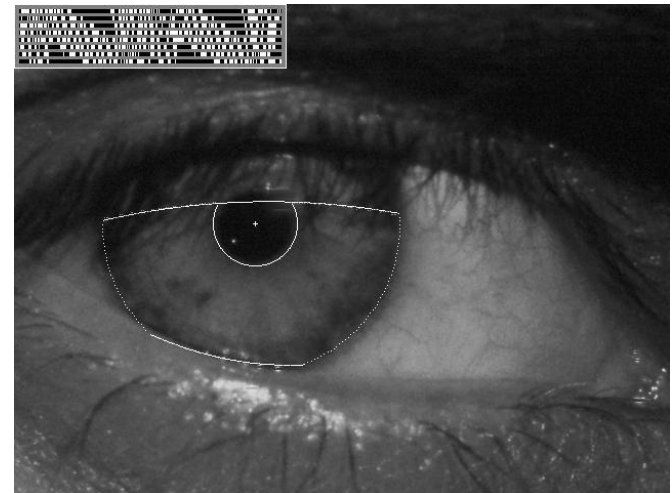
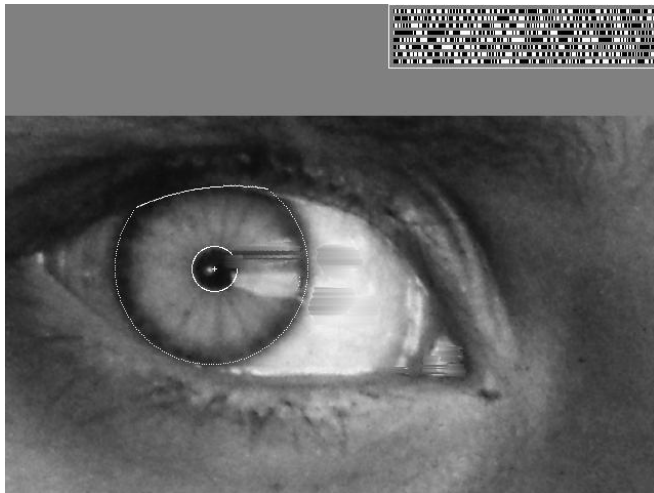
Thanks to Vivek Ramanujan

Vision-based biometrics



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

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



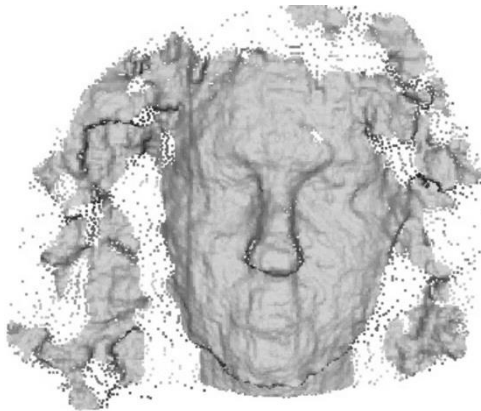
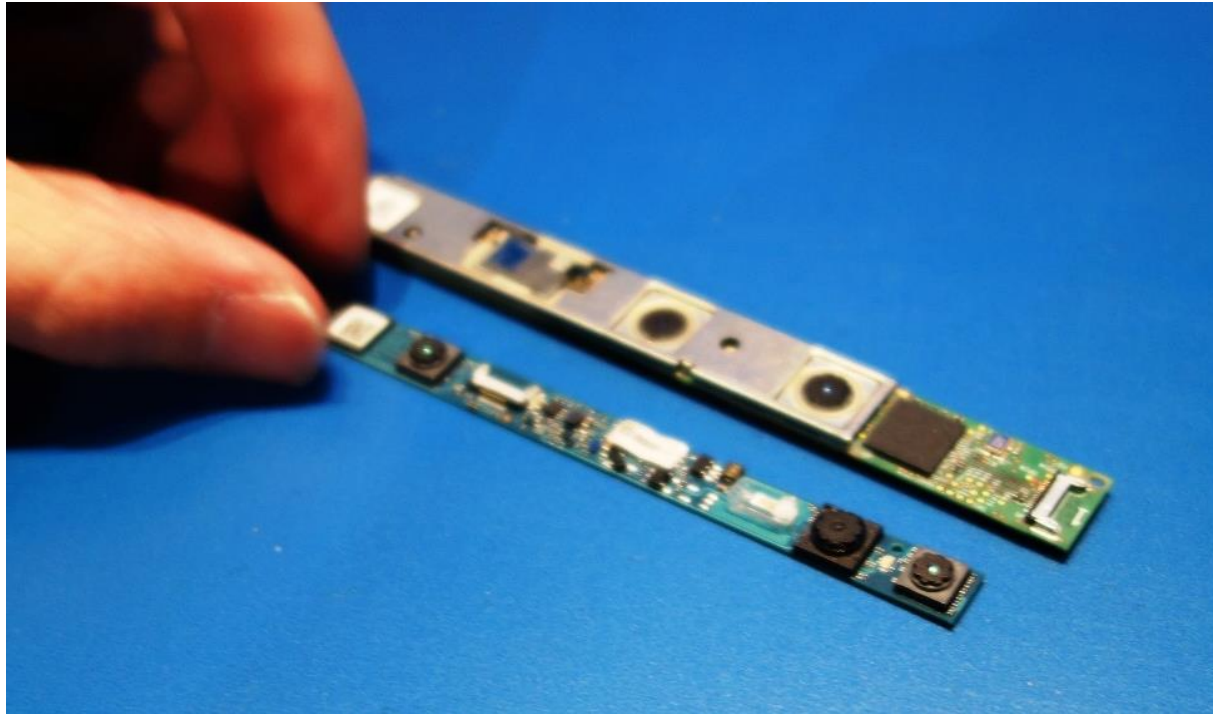
Facial login without a password...



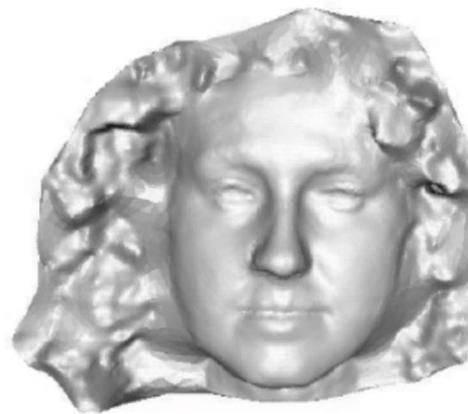
Facial login without a password...



Facial login without a password...



Single depth frame

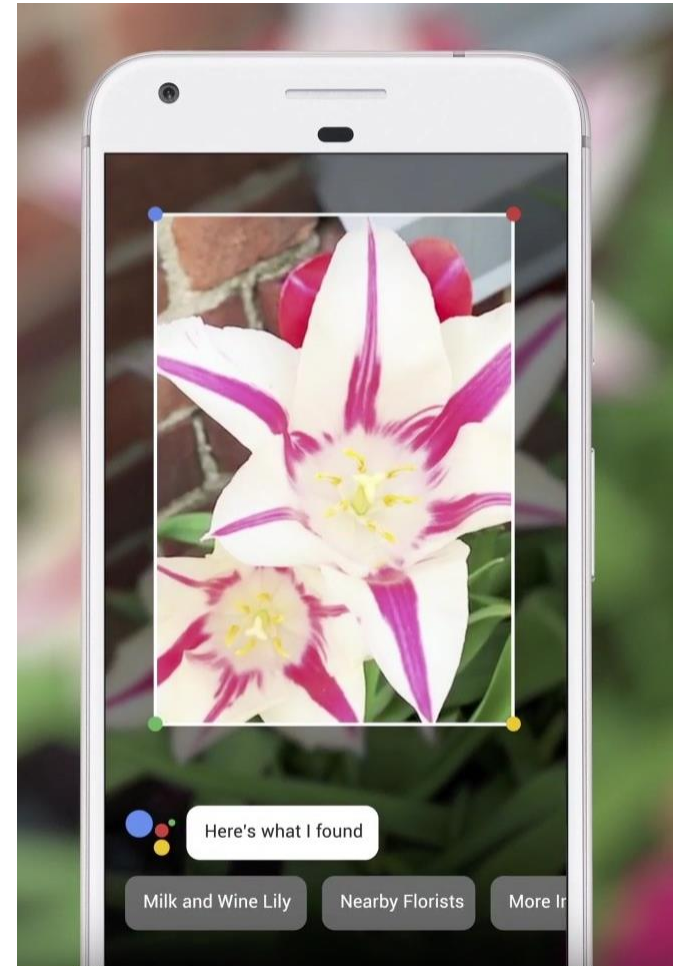
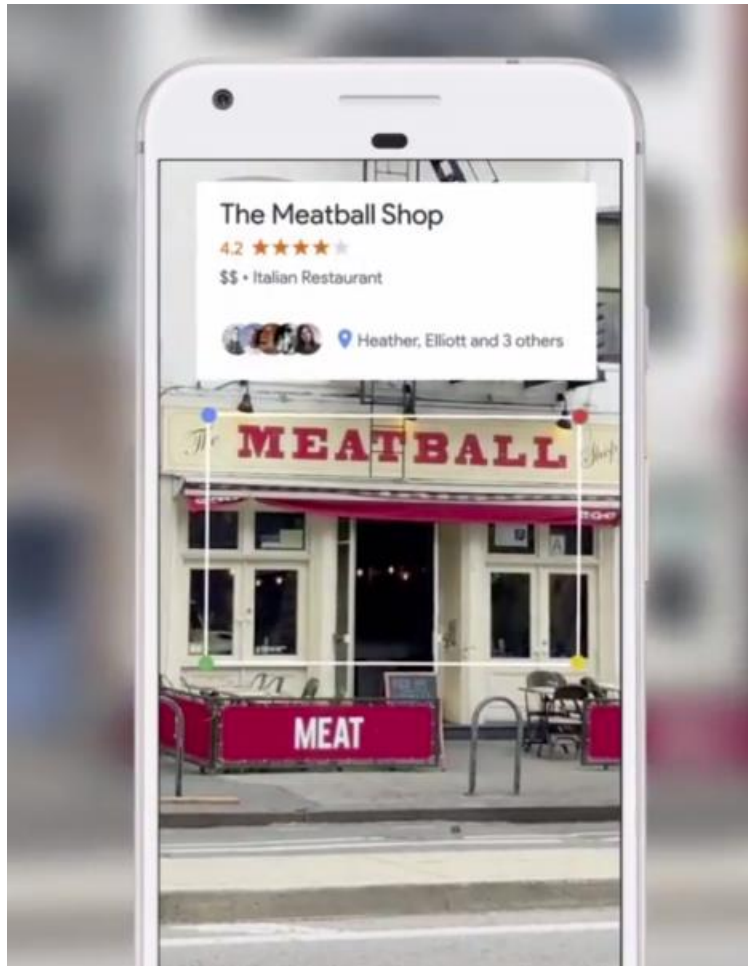


Reconstructed 3D mesh

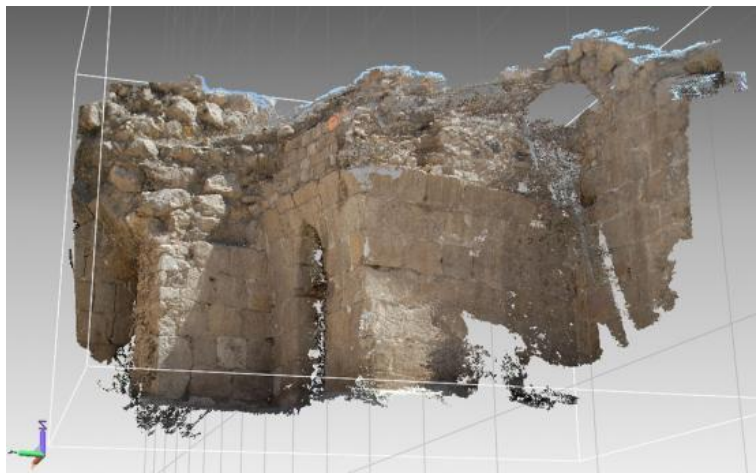
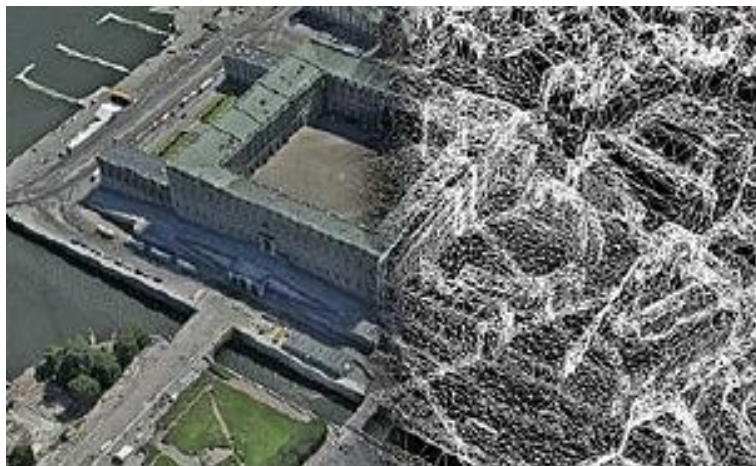


Object recognition (in mobile phones)

e.g., Google Lens



3D from images



Human shape capture



Human shape capture



Human shape capture



Human shape capture

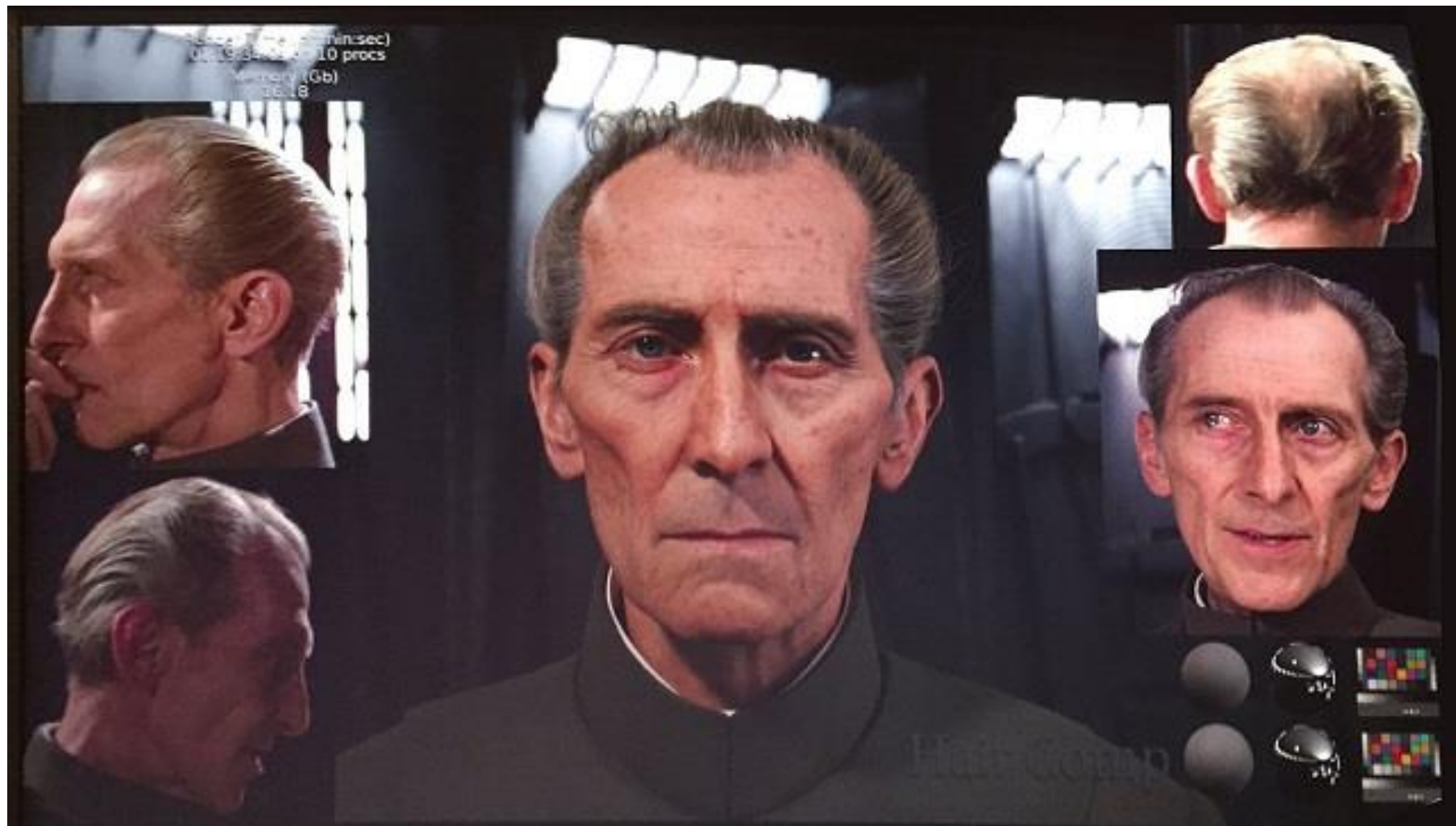


Special effects: shape capture



Star Wars: Rogue One – Peter Cushing / Admiral Tarkin

Special effects: shape capture



Special effects: motion capture



Interactive Games: Kinect

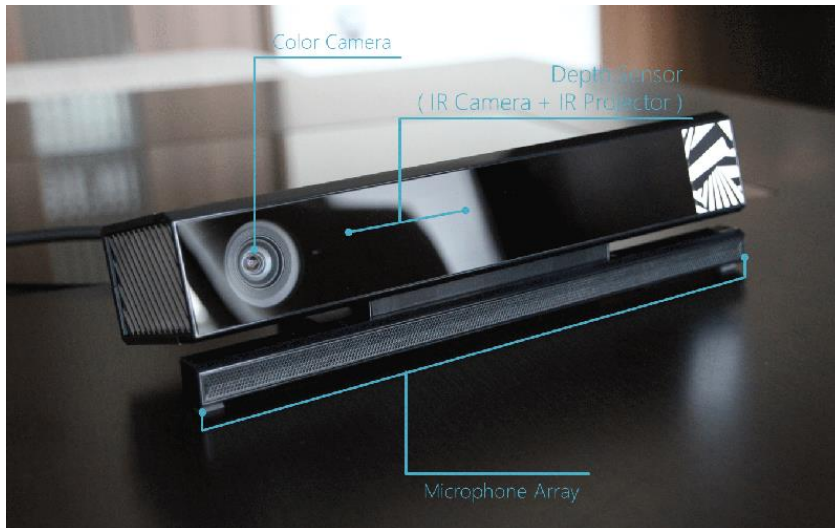
Object Recognition:

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

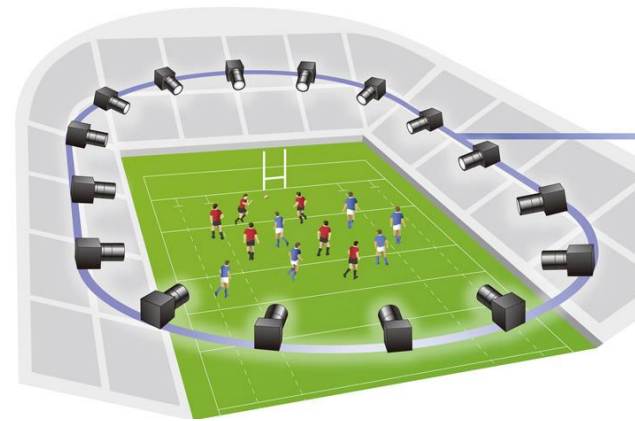
Mario: <http://www.youtube.com/watch?v=8CTJL5IUjHg>

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

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

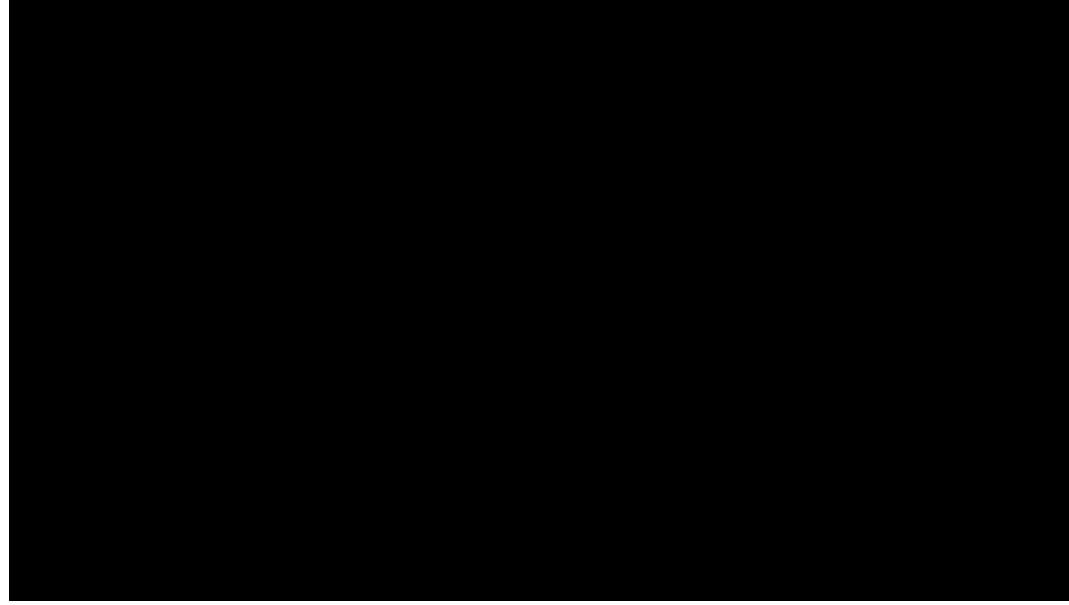


Sports



Virtual pitch markings

Free viewpoint video

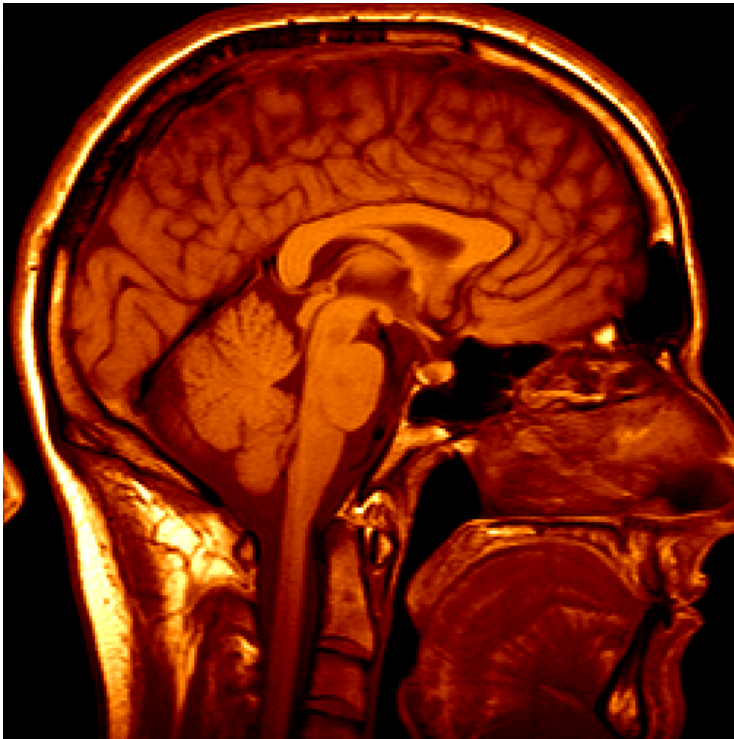


Sportvision first down line

[Canon 2017]

Nice [explanation](http://www.howstuffworks.com) on www.howstuffworks.com

Medical imaging



3D imaging
MRI, CT



Image guided surgery
[Grimson et al., MIT](#)

AutoCars - Uber bought CMU's lab





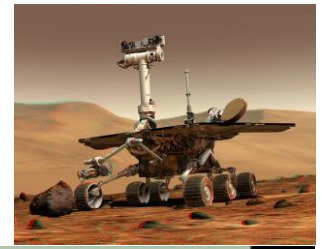


Industrial robots



Vision-guided robots position nut runners on wheels

Vision in space



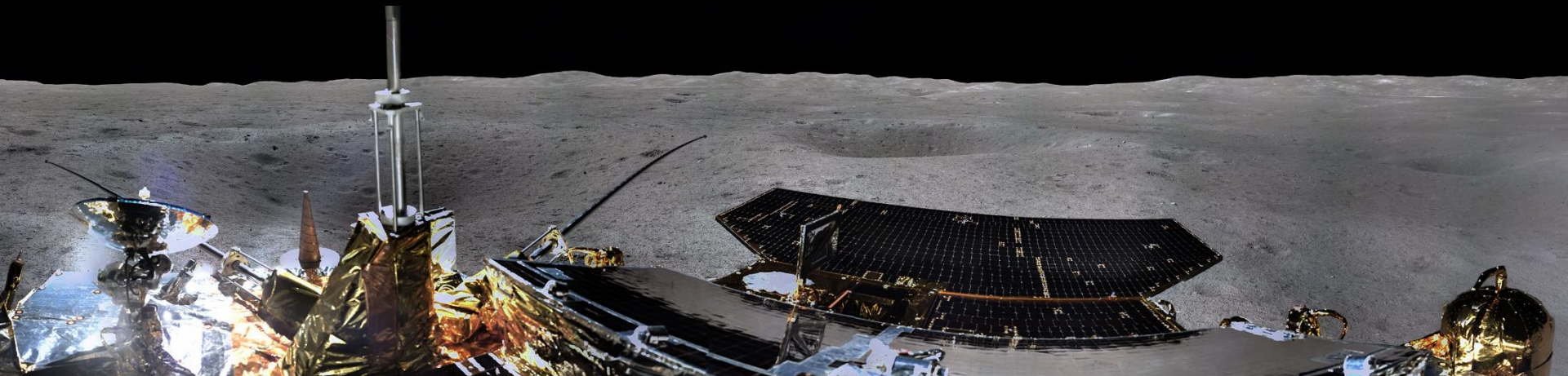
[NASA'S Mars Exploration Rover Spirit](#) captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

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.

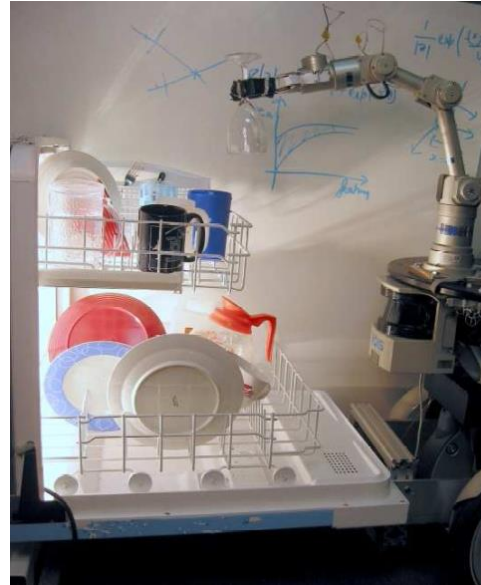
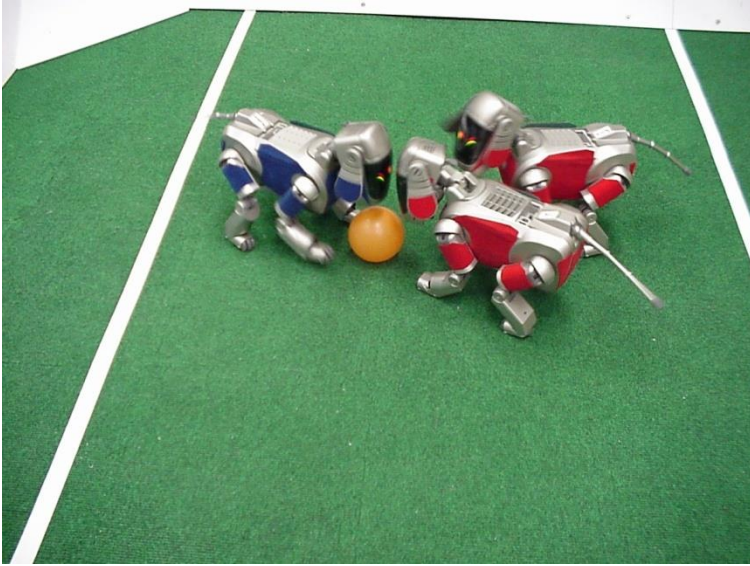


2019 Chang-e 4 Moon Lander



Mobile robots

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



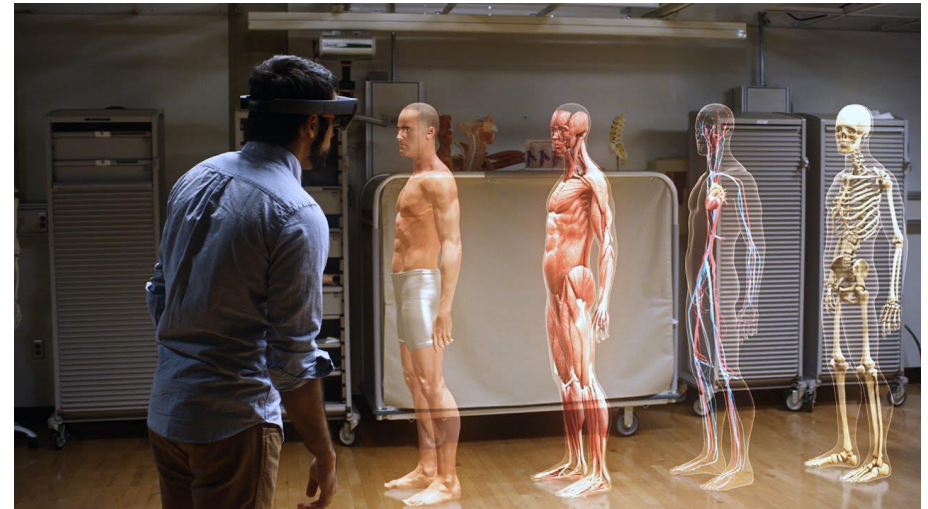
Saxena et al. 2008
STAIR at Stanford



amazon
Prime Air



Augmented Reality and Virtual Reality



MS HoloLens, Oculus, Magic Leap,
ARCore / ARKit

Jitendra Malik, UC Berkeley

Three 'R's of Computer Vision



“[Further progress in] the classic problems of computational vision:

reconstruction

recognition

(re)organization

[requires us to study the interaction among these processes].”

Scope of CSCI 1430

Computer Vision

Robotics

Human Computer Interaction

Machine Learning

Image Processing
Recognition
Deep Learning
Geometric Reasoning

Medical Imaging

Graphics

Neuroscience

Computational Photography

Optics

Computer Vision and Nearby Fields

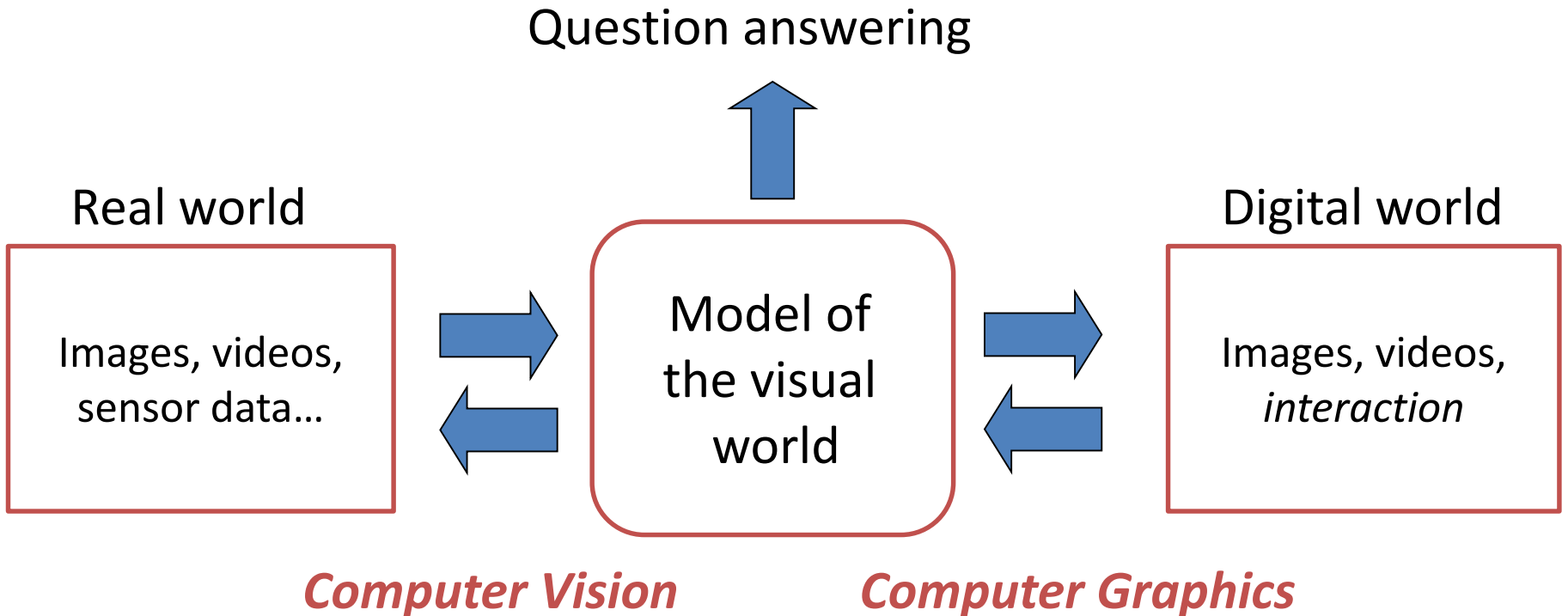
Derogatory summary of computer vision:

“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.

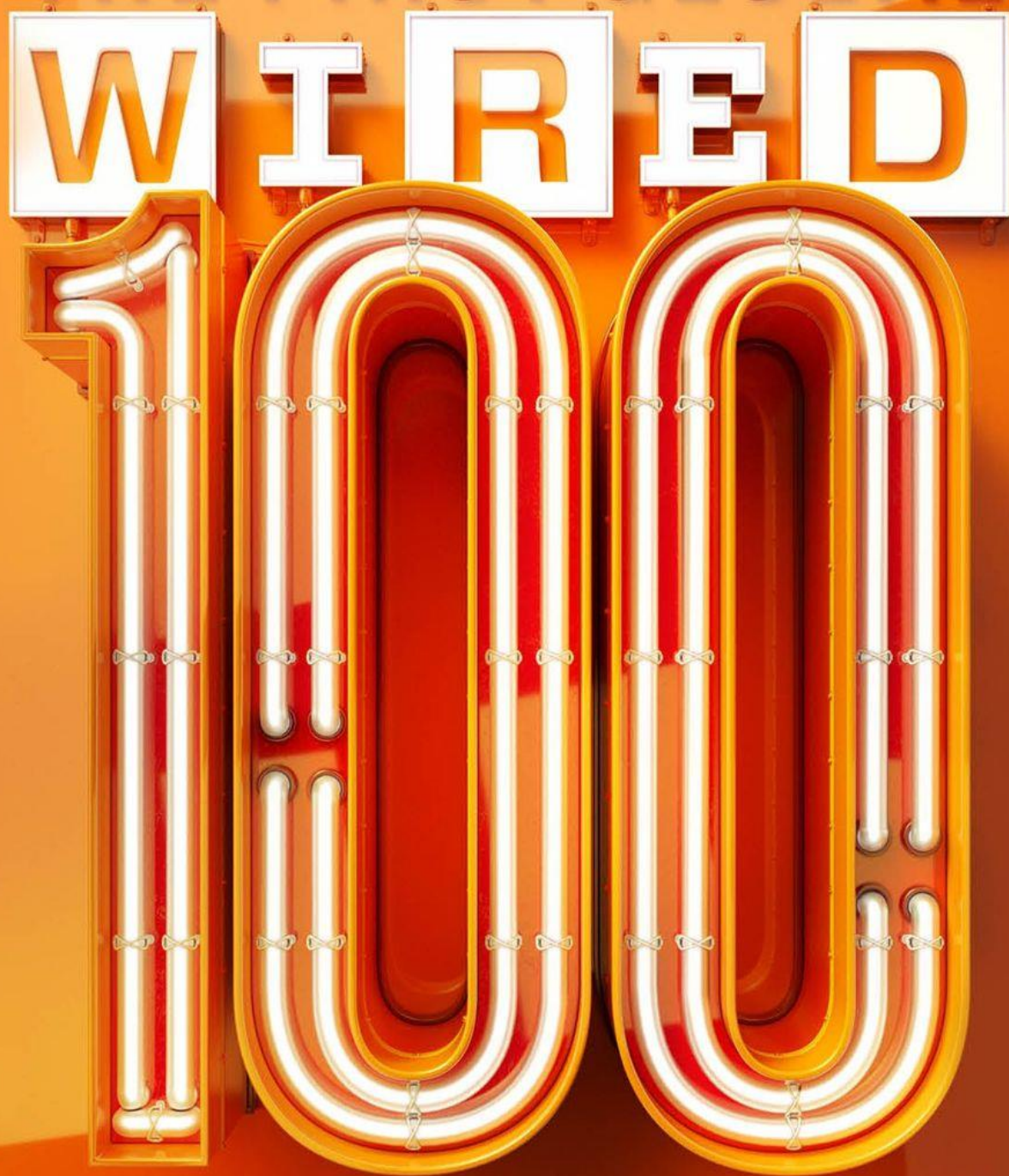
Needs understanding of possible failures.

+

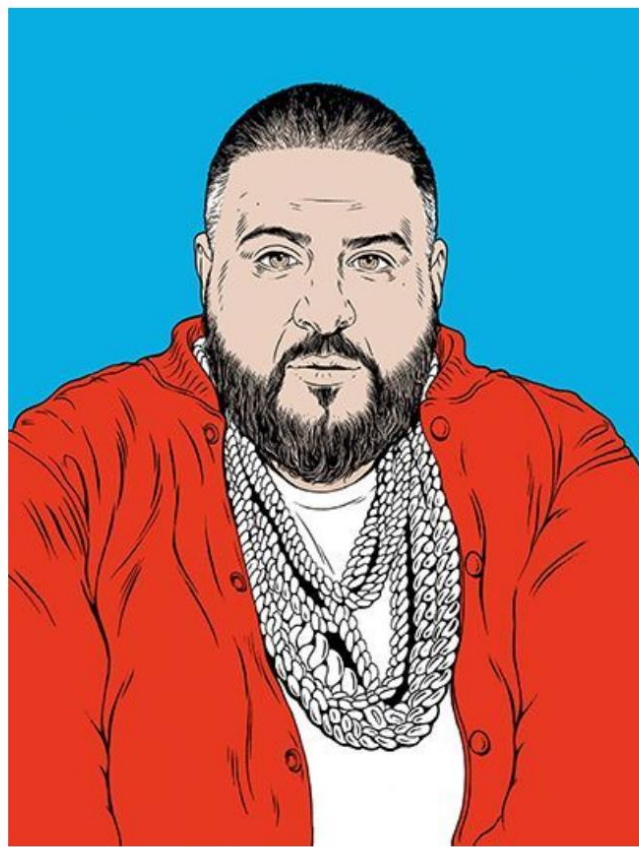
Responsible approach.

+

Techniques to overcome bias.



WHO'S SHAPING THE DIGITAL WORLD?



DJ Khaled

Credit **Louise Zergaeng Pomeroy**

73. DJ Khaled

Snapchat icon; DJ and producer

Louisiana-born Khaled Mohamed Khaled, aka DJ Khaled, cut his musical chops in the early 00s as a host for Miami urban music radio WEDR. He proceeded to build a solid if not dazzling career as a mixtape DJ and music producer (he founded his label We The Best Music Group in 2008, and was appointed president of Def Jam South in 2009).

69. Geoffrey Hinton

Psychologist, computer scientist; researcher, Google Toronto

British-born Hinton has been dubbed the "godfather of deep learning". The Cambridge-educated cognitive psychologist and computer scientist started being an ardent believer in the potential of neural networks and deep learning in the 80s, when those technologies enjoyed little support in the wider AI community.

But he soldiered on: in 2004, with support from the Canadian Institute for Advanced Research, he launched a University of Toronto programme in neural computation and adaptive perception, where, with a group of researchers, he carried on investigating how to create computers that could behave like brains.

Hinton's work – in particular his algorithms that train multilayered neural networks – caught the attention of tech giants in Silicon Valley, which realised how deep learning could be applied to voice recognition, predictive search and machine vision.

The spike in interest prompted him to launch a free course on neural networks on e-learning platform Coursera in 2012. Today, 68-year-old Hinton is chair of machine learning at the University of Toronto and moonlights at Google, where he has been using deep learning to help build internet tools since 2013.

66. PewDiePie

YouTuber, Brighton

In just six years, 26-year-old Swedish YouTuber Felix Kjellberg has risen from dropout who funded his videos by working at a hot-dog stand to a \$12 million global internet sensation. Kjellberg – aka [PewDiePie](#) – has built a quasi-cult fandom by uploading longish videos (Is This Game Too Sexual?!; Worst Game on the Planet!; What Happened To Resident Evil 7??) where he fools around and cracks jokes while playing video games.

In May 2016, Kjellberg's YouTube channel passed 45 million subscribers and, with 12 billion views, it's now the most-watched channel in the history of the video-sharing website. In the gaming industry, it is thought that he can single-handedly determine the success of a game by simply mentioning it in one of his videos.

Kjellberg – who lives in Brighton with his girlfriend and fellow YouTuber Marzia Bisognin, aka CutiePieMarzia – has tried to use his online clout for good, asking his fanbase (dubbed the "Bro Army") to raise money for charities on several occasions.

Recently, he teamed up with The Walking Dead creator Robert Kirkman to produce his first web series – Scare PewDiePie. The series, which features the YouTuber cavorting in horror-game-like situations, debuted on YouTube Red in February 2016.

63. Yann Lecun

Director of AI research, Facebook, Menlo Park

LeCun is a leading expert in deep learning and heads up what, for Facebook, could be a hugely significant source of revenue: understanding its user's intentions.

63. Yann Lecun

Director of AI research, Facebook, Menlo Park

LeCun is a leading expert in deep learning and heads up what, for Facebook, could be a hugely significant source of revenue: understanding its user's intentions.

61. Taylor Swift

Entertainer, Los Angeles





Credit **Google DeepMind**



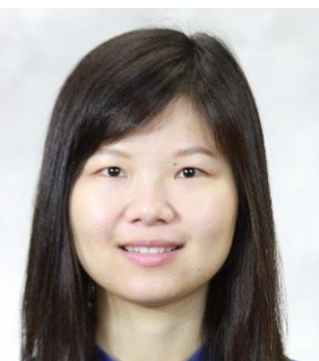
Google-backed startup **DeepMind Technologies** has built an **artificial intelligence** agent that can learn to successfully play 49 classic Atari games by itself, with minimal input.

8. Demis Hassabis

Co-founder and CEO, [DeepMind](#), London

In March 2016, DeepMind's AI [AlphaGo](#) beat the Go world champion [Lee Se-dol](#). The Google-owned startup is moving machine learning forward at a pace that could affect every industry, from healthcare to commerce.

COURSE DETAILS



Yuanning
Hu (HTA)



Ruizhao
Zhu (HTA)



Zhoutao Lu



Josh Roy



Eleanor
Tursman



Brendan
Walsh



Joshua
Chipman



Purvi Goel



Yan Huang



Seungchan
Kim



Jamie
DeMaria



Anna Sabel



Michael
Chen



- + Wentao Hao
- + Jiawei Zhang
- + James Tompkin

We are here to help you!

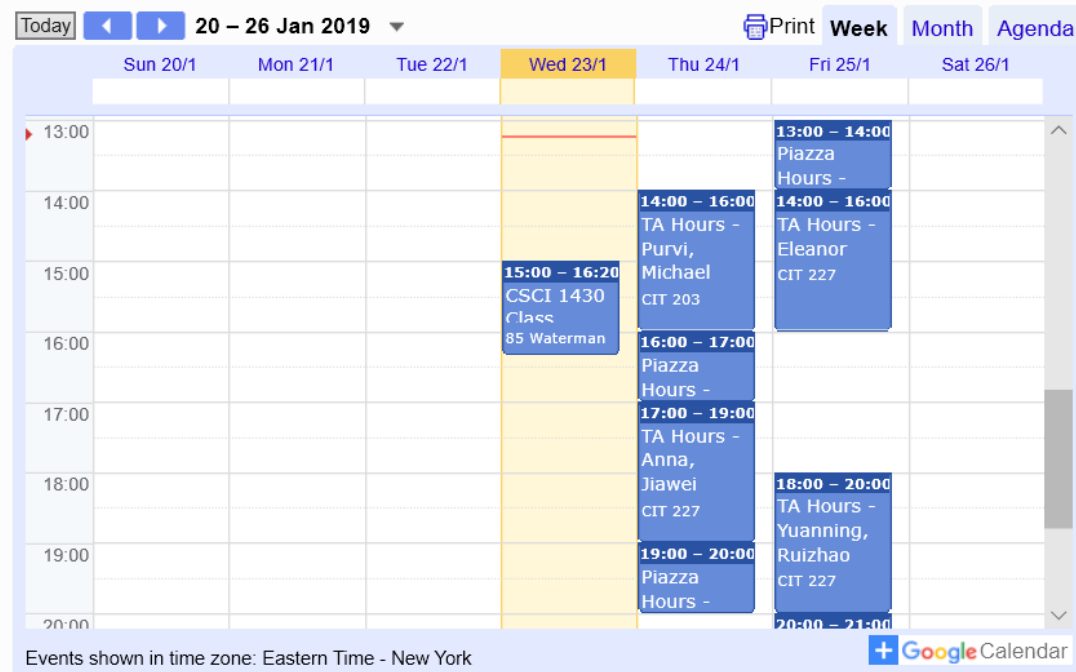
Contact

- **Piazza**—your first stop for questions and clarifications. Piazza will be staffed at specific times, when a member of the team will be answering questions (existing and new). At other times, please pull together as a class and help each other, and we'll help soon.
- cs143otas@lists.brown.edu—your second stop for less typical questions.

Office/Piazza Hours

TA Hours

Course webpage



Course Description

This course provides an introduction to computer vision, including fundamentals of image formation, camera imaging geometry, feature detection and matching, stereo, motion estimation and tracking, image classification, scene understanding, and deep learning with neural networks. We will develop basic meth-

James Tompkin



Assistant Prof. @ Brown CS



Contact

Office hours: Tues 1–2pm
[Book appointment](#)

Brown GCal: Use 'Find a Time'
[Instructions](#)

james_tompkin@brown.edu

[Follow @jamestompkin](#)

[Follow @jtompkin](#)

[YouTube](#)

Thomas J. Watson Sr. Center for
Information Technology
Room 547
115 Waterman Street
Providence, RI, 02912

Research Overview



[MP4 download \(200MB\)](#)

How can we make video a creative medium for EVERYONE?
How can computation *remove barriers* from interaction?
How can image understanding help us *explore* media?

To help answer these questions, I create graphics, vision, and interaction techniques to
prove our understanding of the connections *within* media.

Academic lineage

- Post-doc with [Prof. Hanspeter Pfister](#) at the [Harvard Paulson School of Engineering and Applied Sciences](#)
- Post-doc with [Prof. Christian Theobalt](#) at the [Max-Planck-Institute for Informatics and the Interdisciplinary Center for Computer Graphics and Visualization](#)
- Research intern with [Prof. Wojciech Matusik](#) at [Disney Research Cambridge](#).

My Office Hours

jamestompkin.com

CSCI 1430 – James Hays

- Continuing his course – many materials & projects from him + previous staff – serious thanks!
- If you see a little 'JH' in the slide corner, then it's his.

Prerequisites

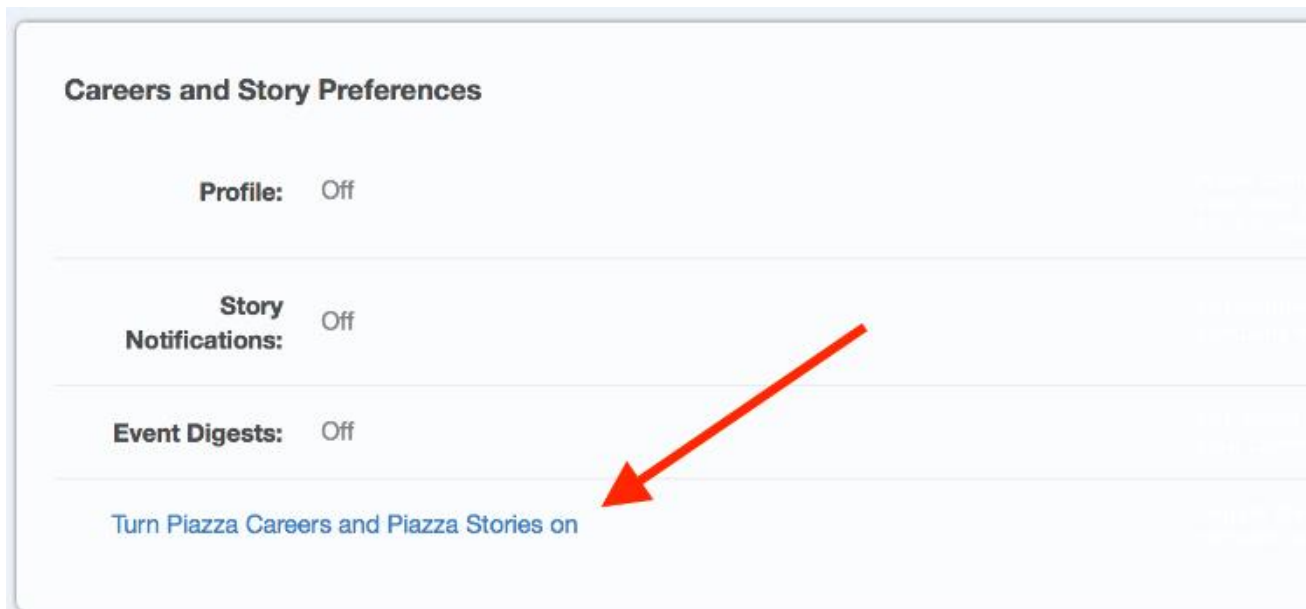
- **Linear algebra**, basic calculus and probability.
- Programming, data structures.

Contact

- Course runs *quiet hours* – 9pm to 9am.
 - We will ignore you (temporarily).
- Piazza first
 - TAs have set Piazza hours.
- cs1430tas@lists.brown.edu second

Piazza and third-party data

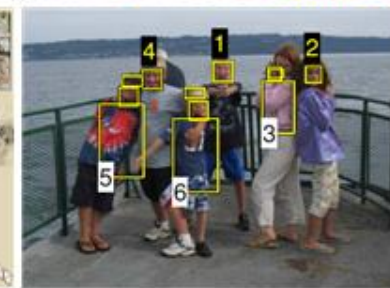
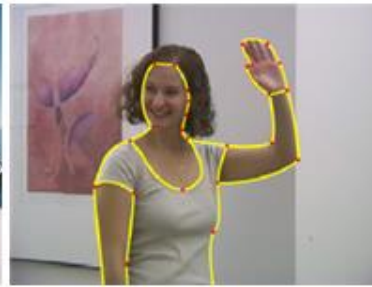
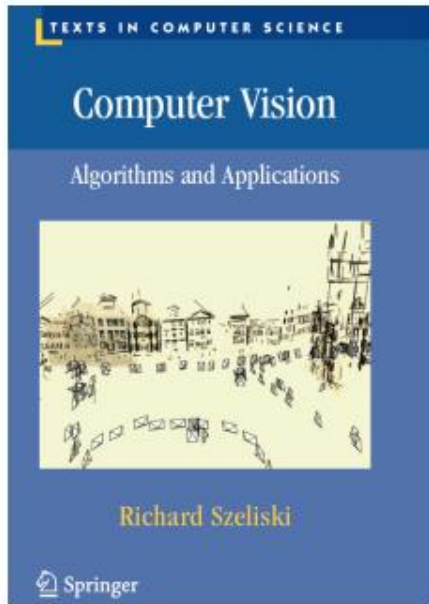
- Piazza makes money by selling your data.
- By default (at least on new accounts), this option is *enabled*.
- You can opt out.



Textbook

Computer Vision: Algorithms and Applications

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



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

Textbook

Deep Learning

An MIT Press book

Ian Goodfellow and Yoshua Bengio and Aaron Courville

- Can I get a PDF of this book?

No, our contract with MIT Press forbids distribution of too easily copied electronic formats of the book.

- Why are you using HTML format for the web version of the book?

This format is a sort of weak DRM required by our contract with MIT Press. It's intended to discourage unauthorized copying/editing of the book.

- What is the best way to print the HTML format?

Printing seems to work best printing directly from the browser, using Chrome. Other browsers do not work as well.

Projects / Grading

- 100% projects (7 total)
- Project 0: Intro / setup
- Projects 1-5: Structured conceptual / code
- Project 6: Final group project

Project	Percent
0	2%
1-5	~14.6%
6	~25%

Submission system: Gradescope

 gradescope <≡

Brown CSCI 1430

Introduction to Computer Vision


 Dashboard

 Assignments

 Roster

 Course Settings

INSTRUCTOR

 James Tompkin

Brown CSCI 1430 | Spring 2019

DESCRIPTION

<https://cs.brown.edu/courses/csci1430/>

◆ ACTIVE ASSIGNMENTS

RELEASED

DUE (EST) ▼

Project 1 Code

JAN 23

FEB 01 AT 9:00PM

Project 1 Written Questions

JAN 23

FEB 01 AT 9:00PM

Project 0 Written Questions

JAN 23

JAN 25 AT 9:00PM

gradescope <≡

Brown CSCI 1430 | Fall 2017

DESCRIPTION

<http://cs.brown.edu/courses/csci1430/>

INSTRUCTOR

James Tompkin

NAME	STATUS	RELEASED	DUE (EDT)
Project 1 Code	NO SUBMISSION	SEP 08	1 WEEK, 4 DAYS LEFT SEP 22 AT 9:00PM LATE DUE DATE: SEP 25 AT 9:00PM
Project 1 Written Questions	NO SUBMISSION	SEP 08	6 DAYS, 8 HOURS LEFT SEP 17 AT 9:00PM LATE DUE DATE: SEP 18 AT 9:00PM
Project 0 Written Questions	SUBMITTED	SEP 06	4 DAYS, 8 HOURS LEFT SEP 15 AT 9:00PM LATE DUE DATE: SEP 18 AT 9:00PM

James Tompkin

- Anonymous submissions please.
 - Don't put your name into Gradescope; only your Brown email.
 - Use fake name if you need to.
- Written questions:
 - It's a template.
 - Use the space provided only
 - Please don't make more/less pages
 - You can put extra pages at the end

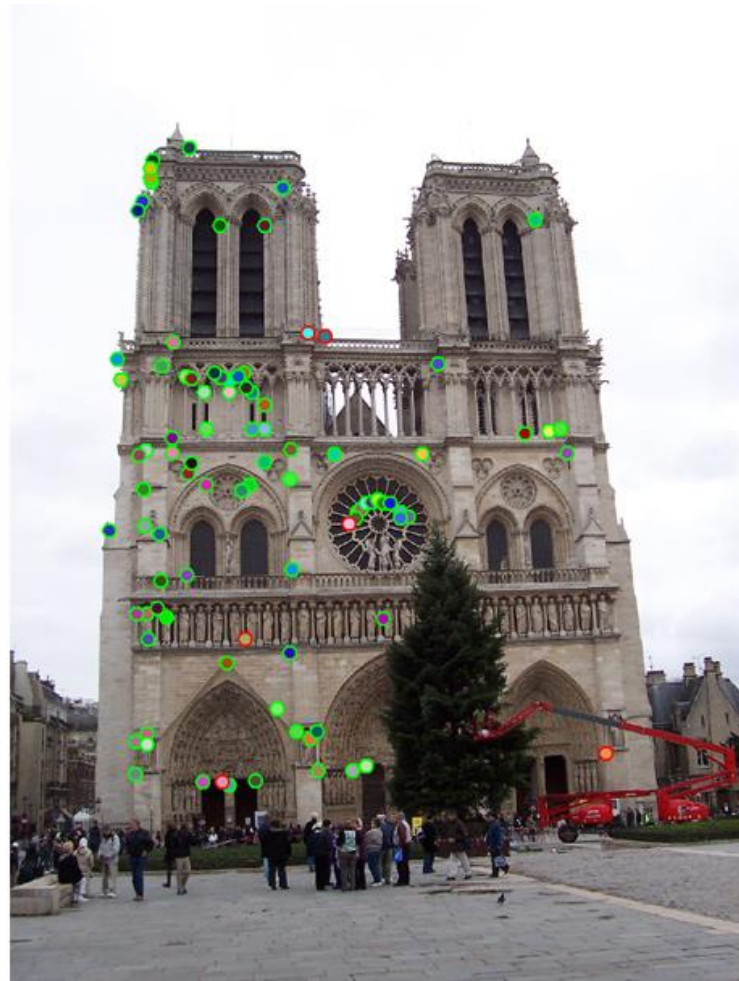
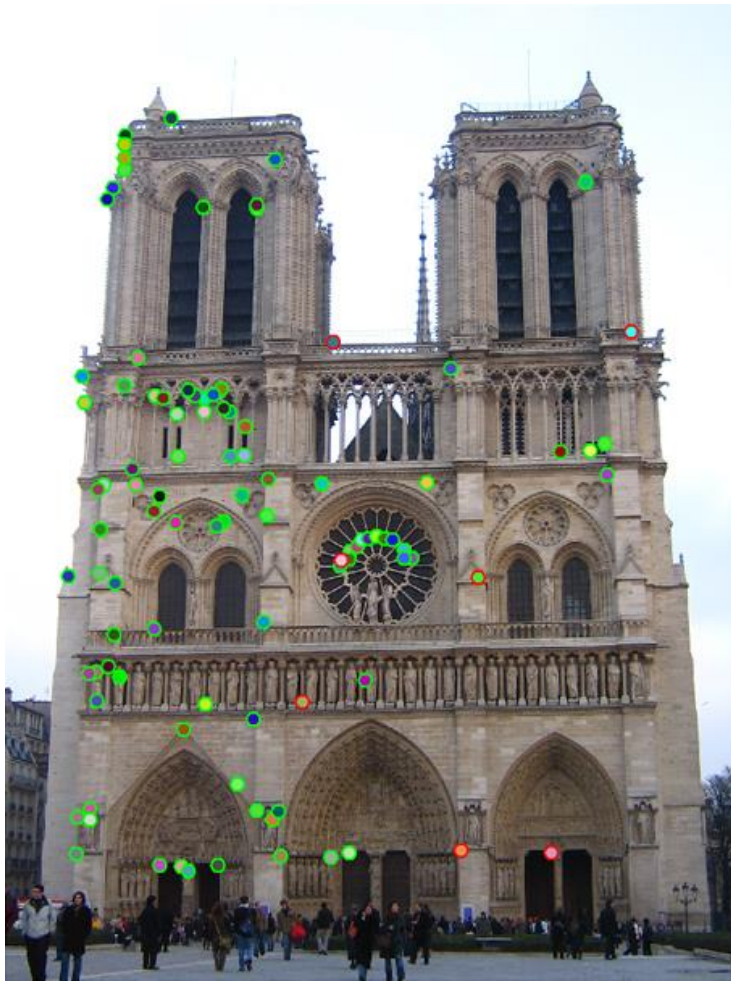
Proj 1: Image Filtering and Hybrid Images

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



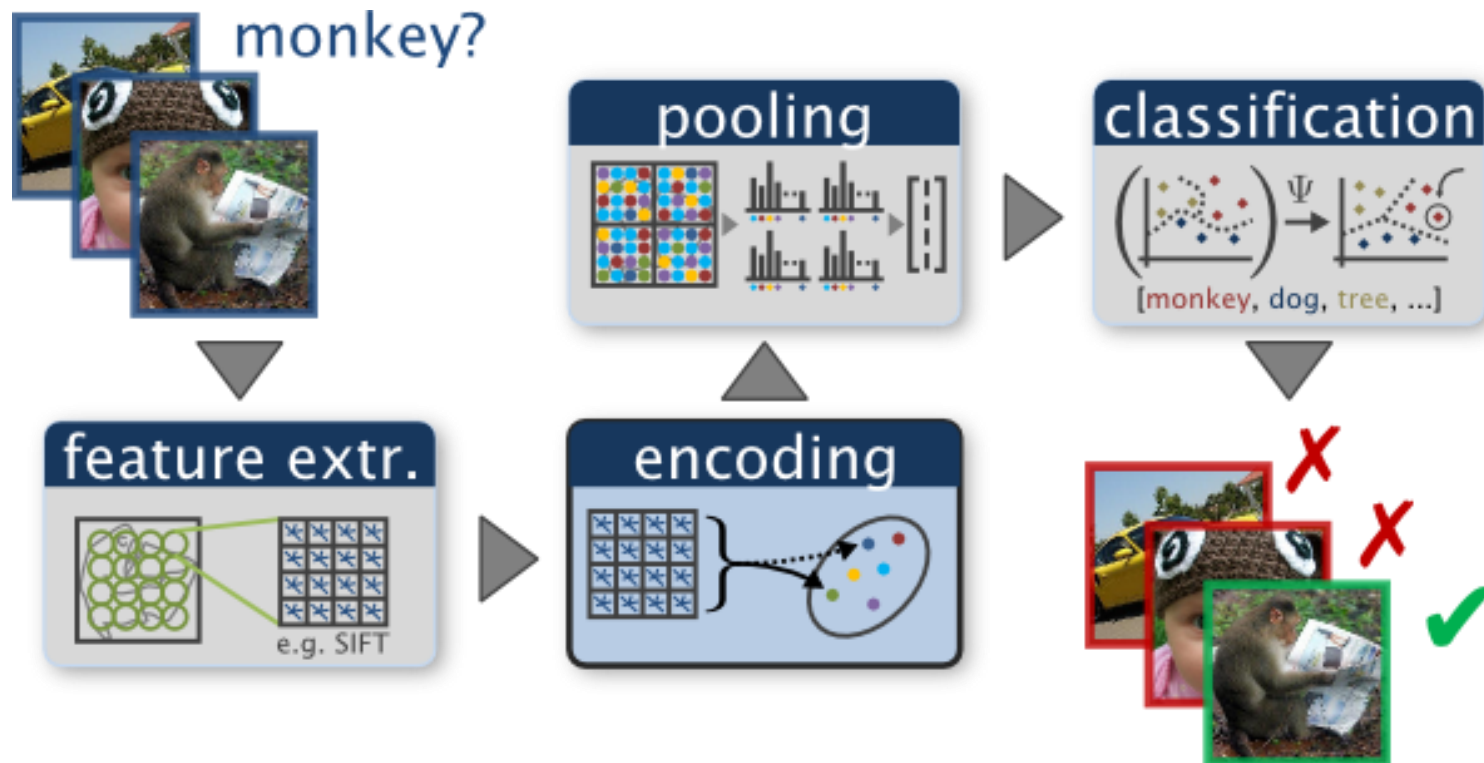
Proj 2: Local Feature Matching

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



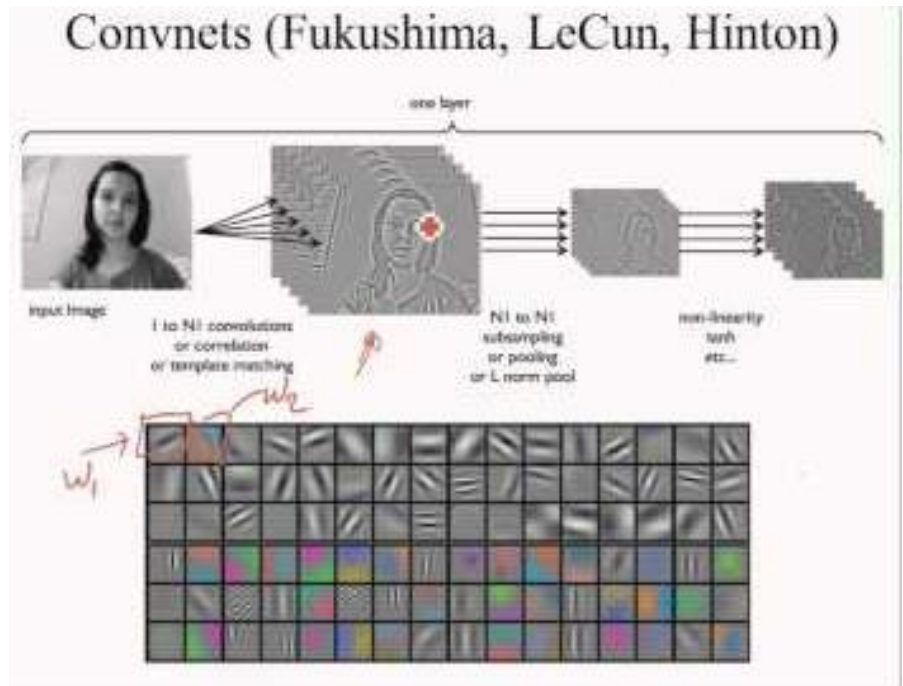
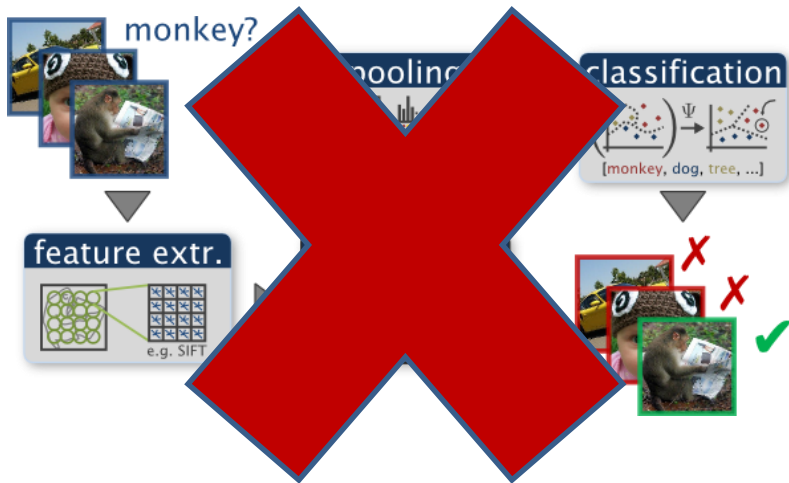
Proj 3: Scene Recognition with Bag of Words

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



Proj 4: Convolutional Neural Nets

- Proj 3 again, but state of the art.



Proj 5: Multi-view Geometry

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



Proj 6: Group Final Project

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

Project examples

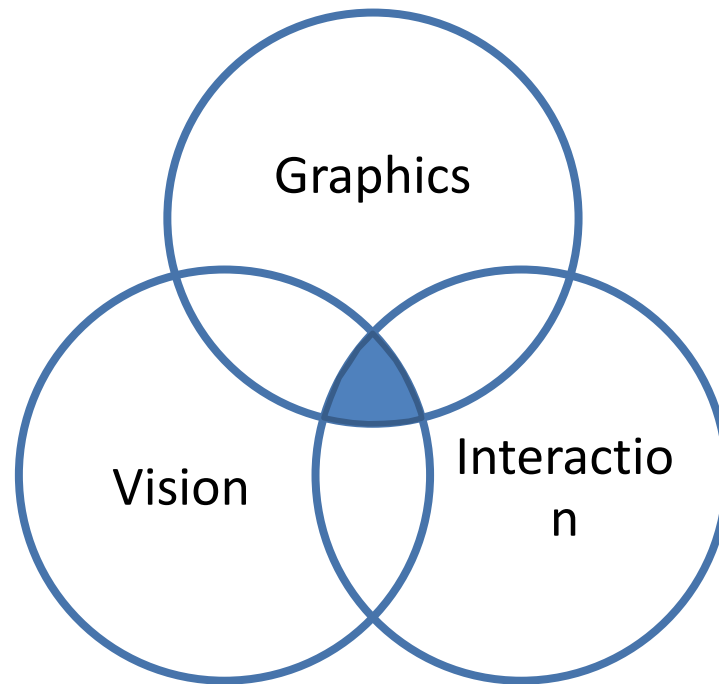
- Real-time eye tracking
- Multi-view geometry reconstruction
- Your favourite deep learning dataset
- Style Transfer / CycleGAN
- Adversarial CNN hacking

Any questions at this stage?!

- Waitlist on course webpage
- Install vscode / Python 1430 venv
- Project 0 tutorial
- Please go to TA hours from *tomorrow*
- Project 0 due FRIDAY 9pm

JAMES

I work in here.







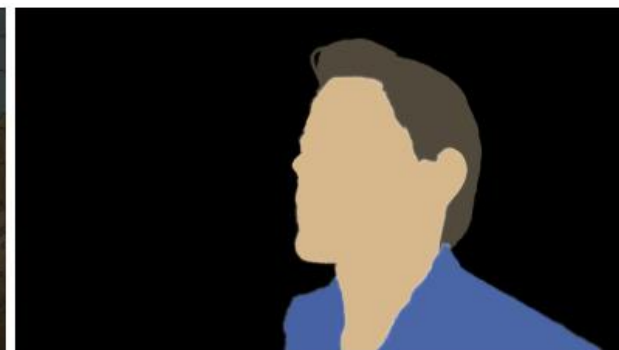




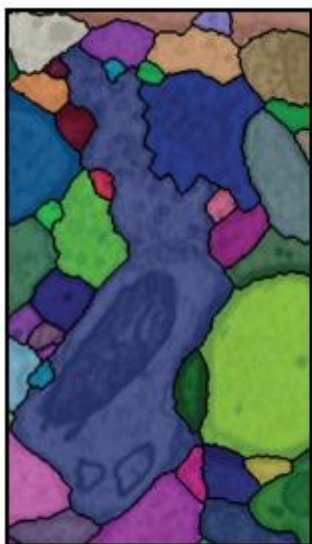
(a) Supervoxel Segmentation



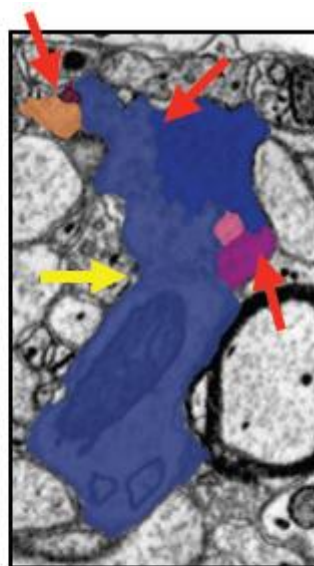
(b) User Interactions



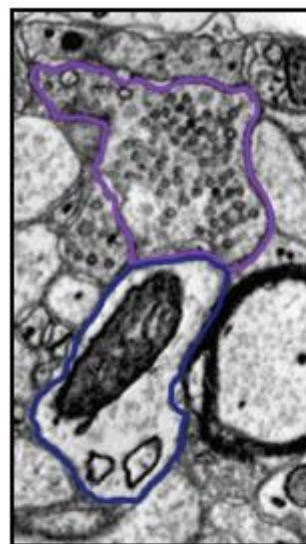
(c) Multi-label Segmentation



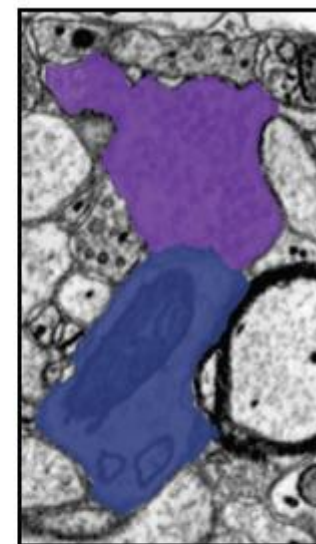
Initial Segmentation



Merge- and Split Errors



Correct Borders



Fixed Segmentation

Instructor: James Tompkin



HARVARD

John A. Paulson
School of Engineering
and Applied Sciences



mpg

Max Planck Institute
Germany



University College London
UK



Disney Research



Massachusetts
Institute of
Technology



NVIDIA

Render pixels?
Capture pixels?
Interact with pixels?
I am probably interested.

Watch my research overview video!