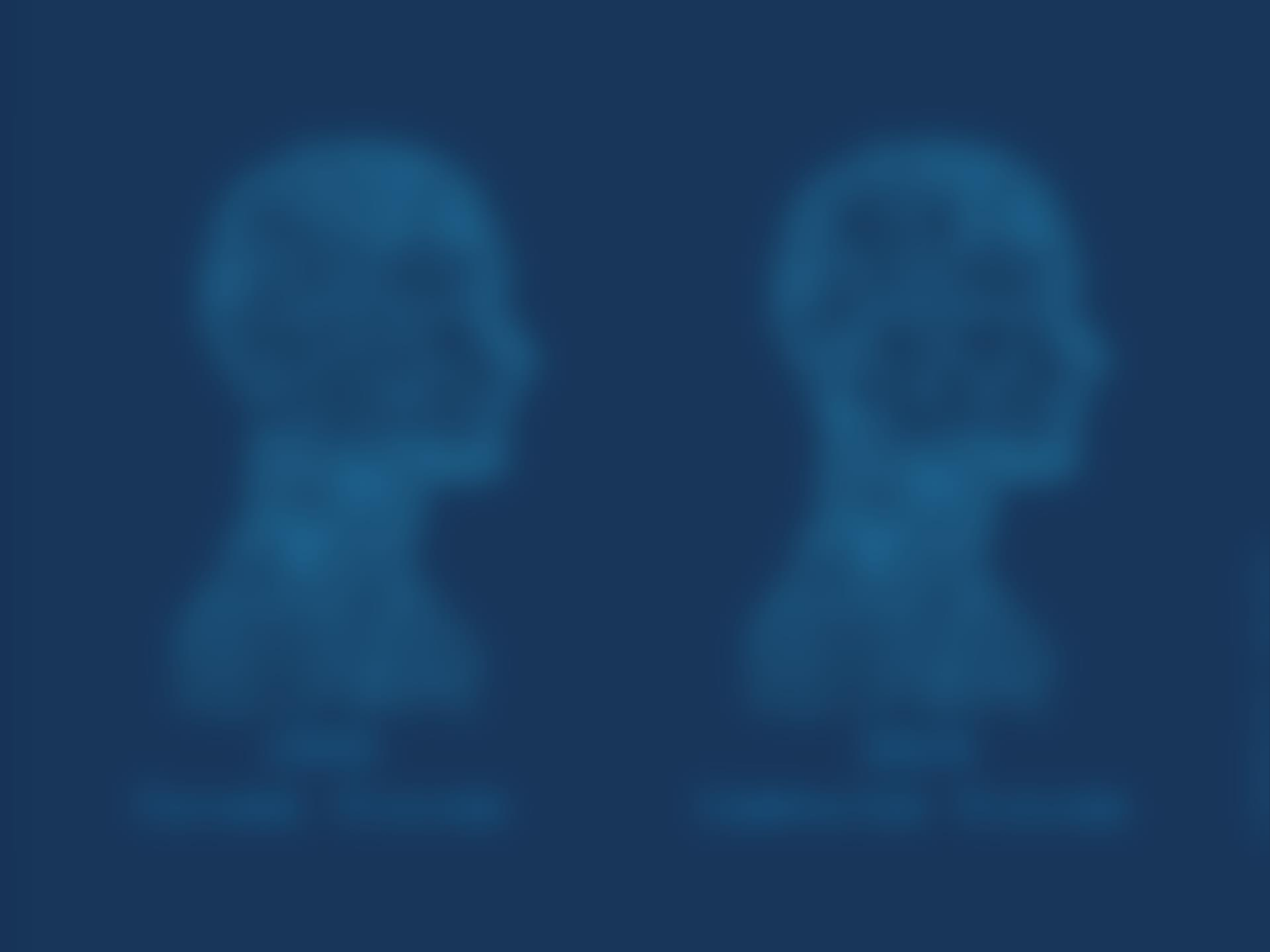
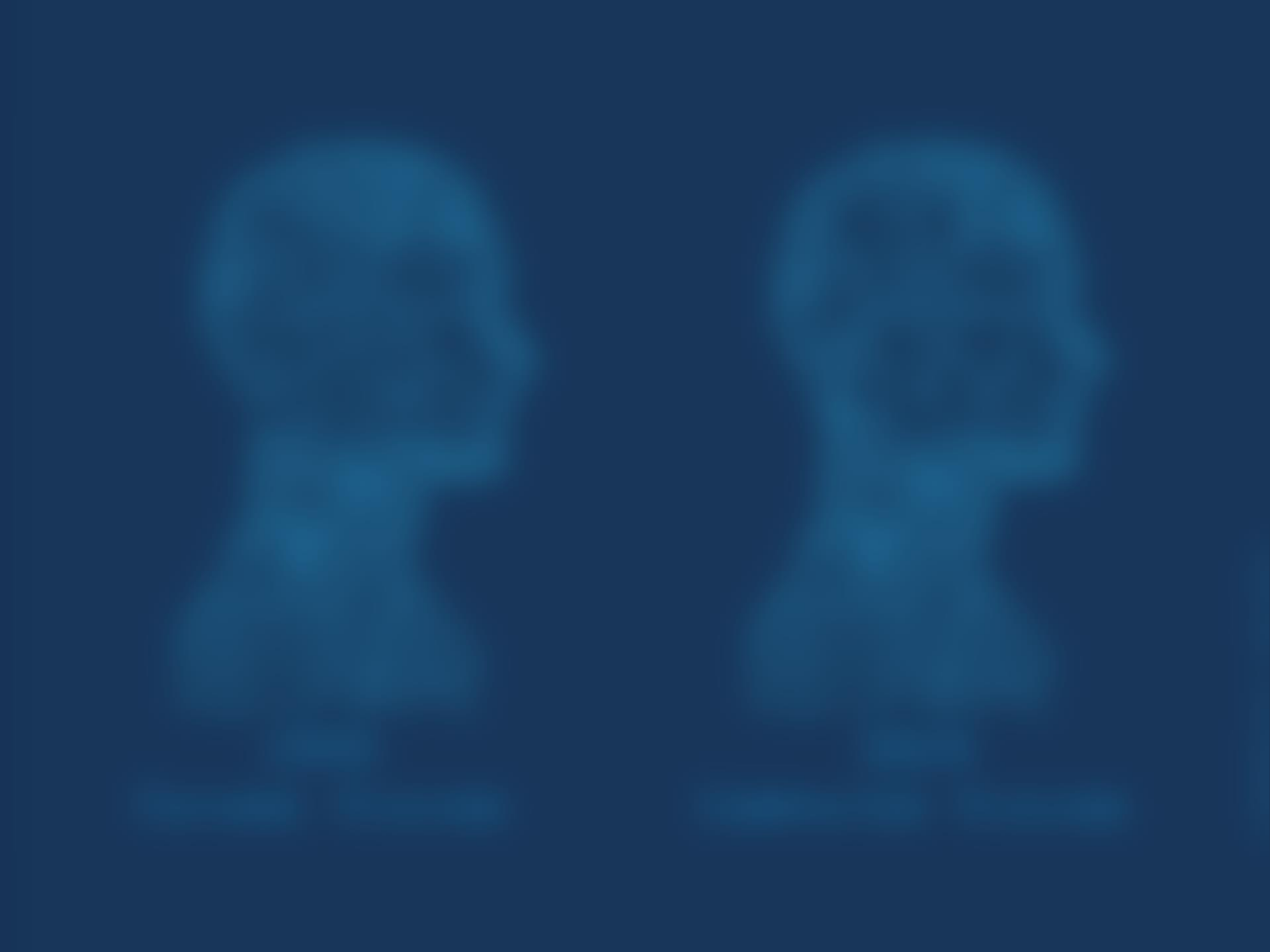
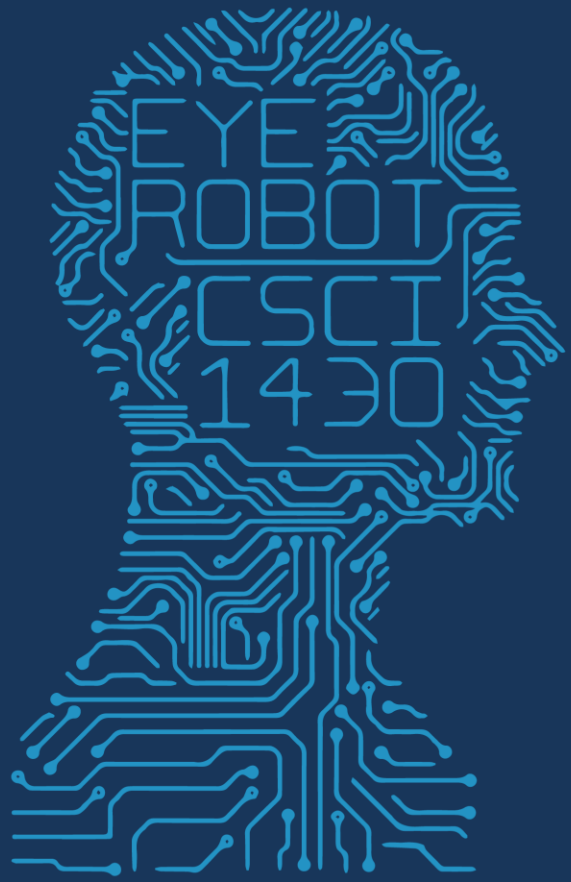




1950

Future Vision



2020

Computer Vision

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?

Think-Pair-Share

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

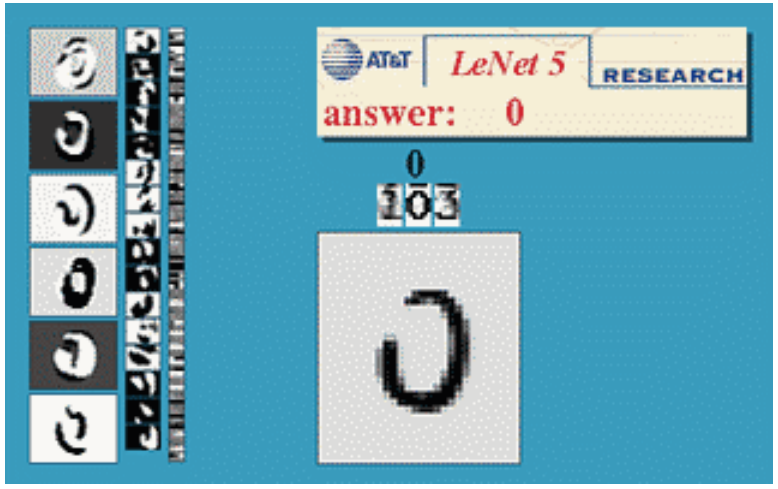
Transportation: Assisted driving (everything), face tracking/iris dilation for drunkeness, drowsiness, automated distribution (all modes)

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/

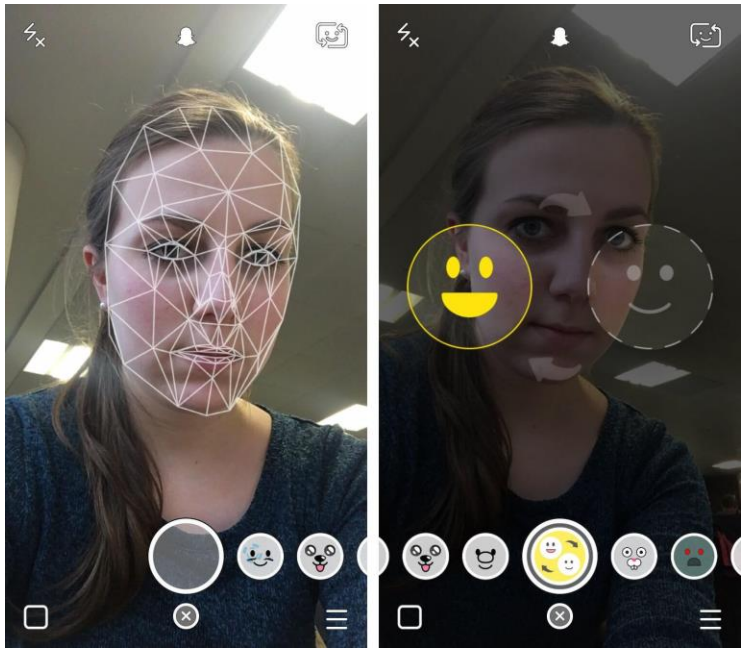
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

WAYGO

Live
Camera
Translation

JH

Face detection

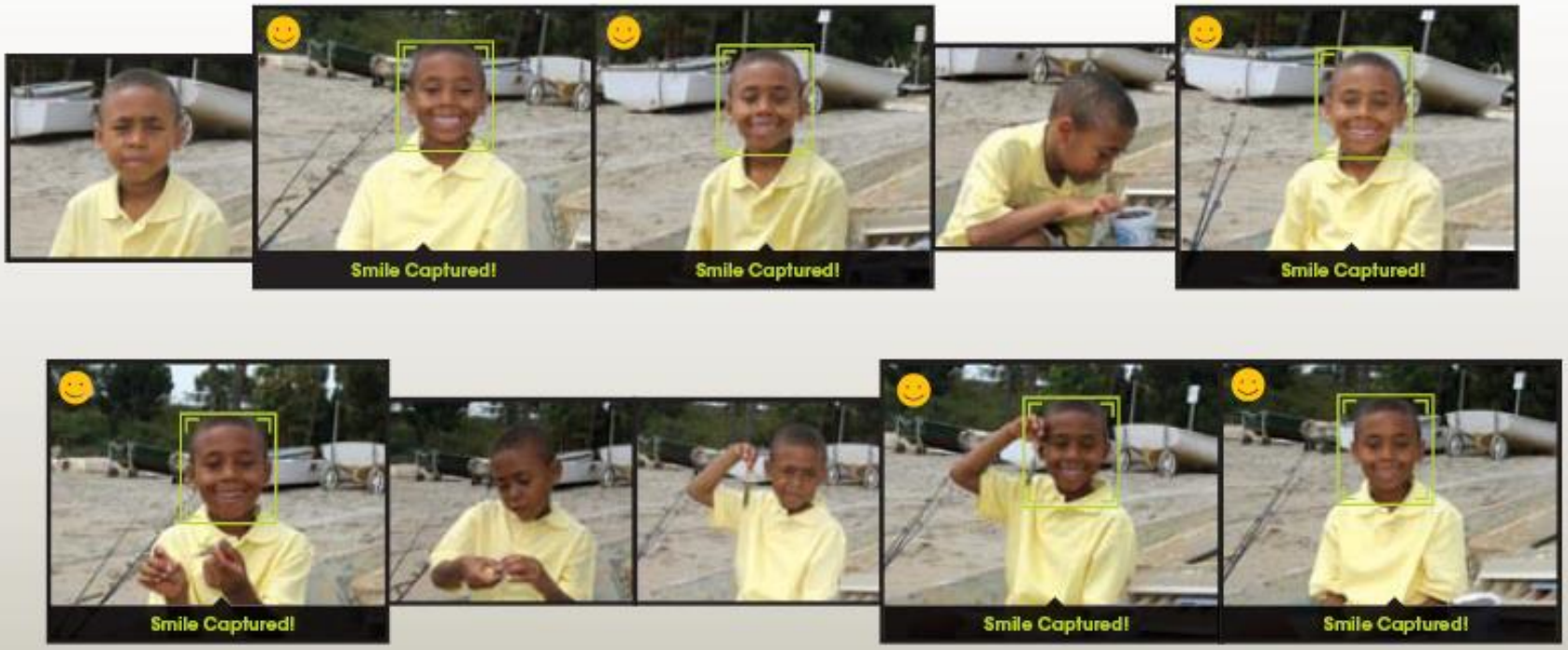


- Almost all digital cameras detect faces
- Snapchat face filters

Smile detection

The Smile Shutter flow

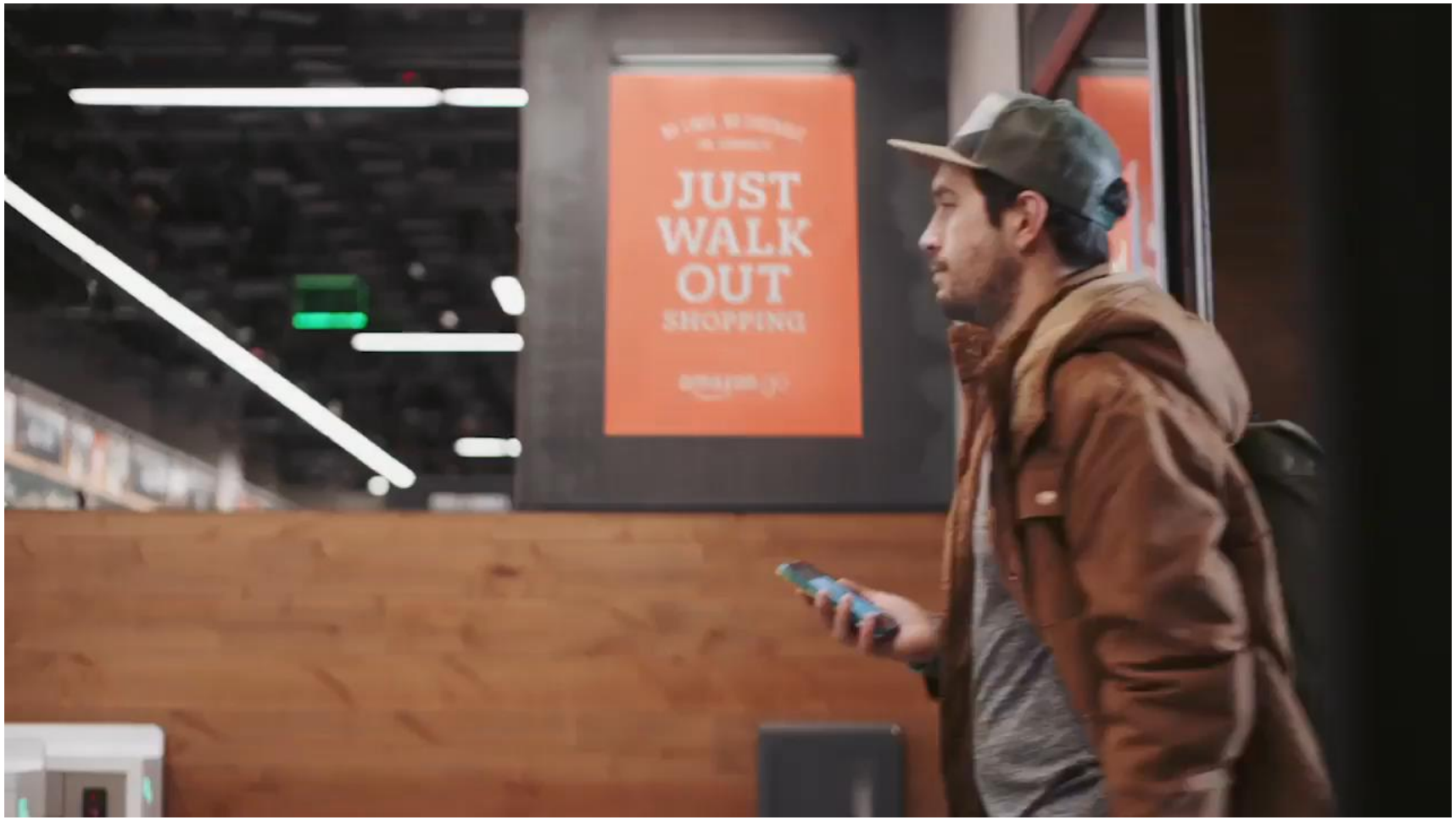
Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



Sony Cyber-shot® T70 Digital Still Camera

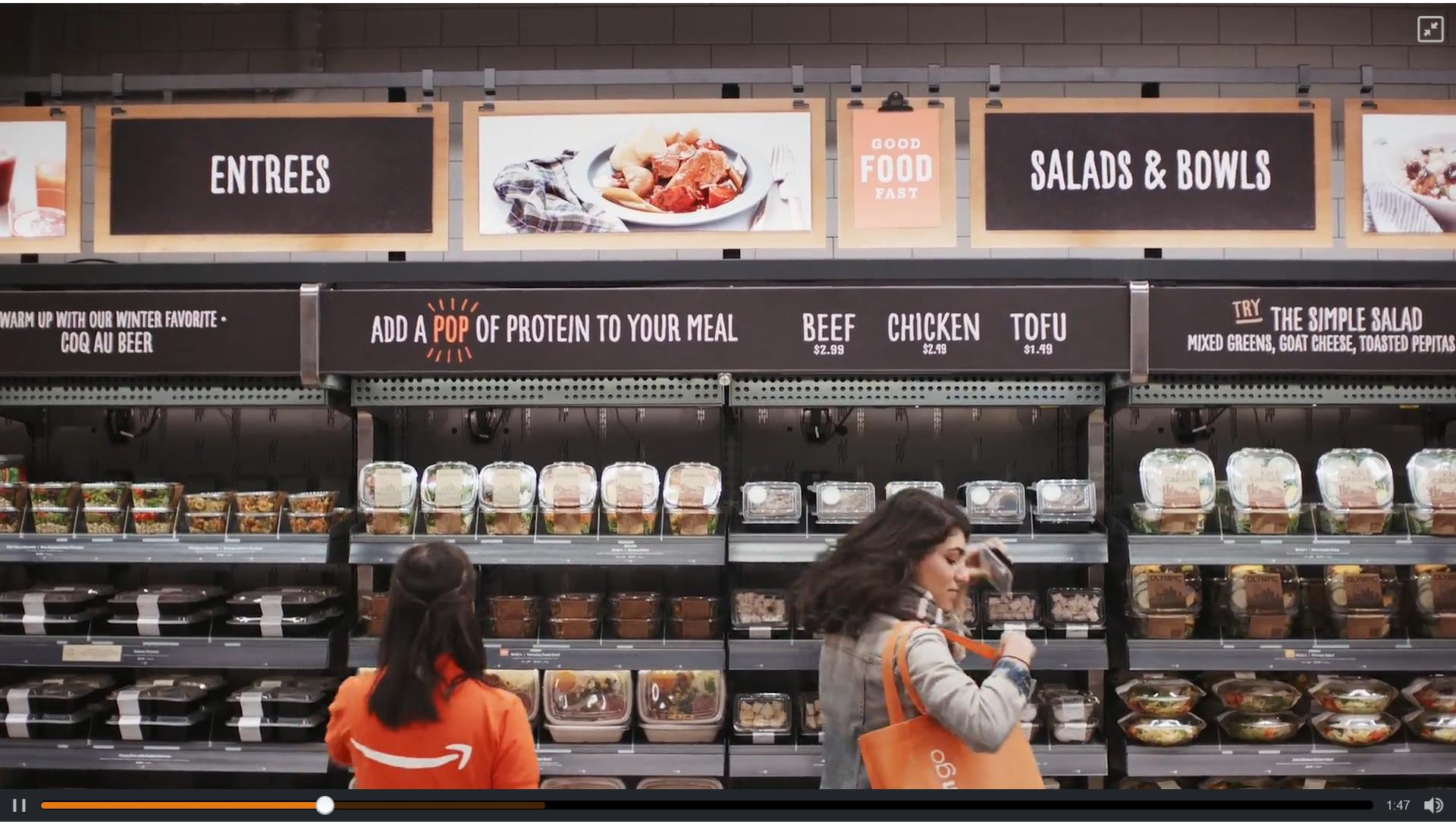
JH

Object recognition (in supermarkets)

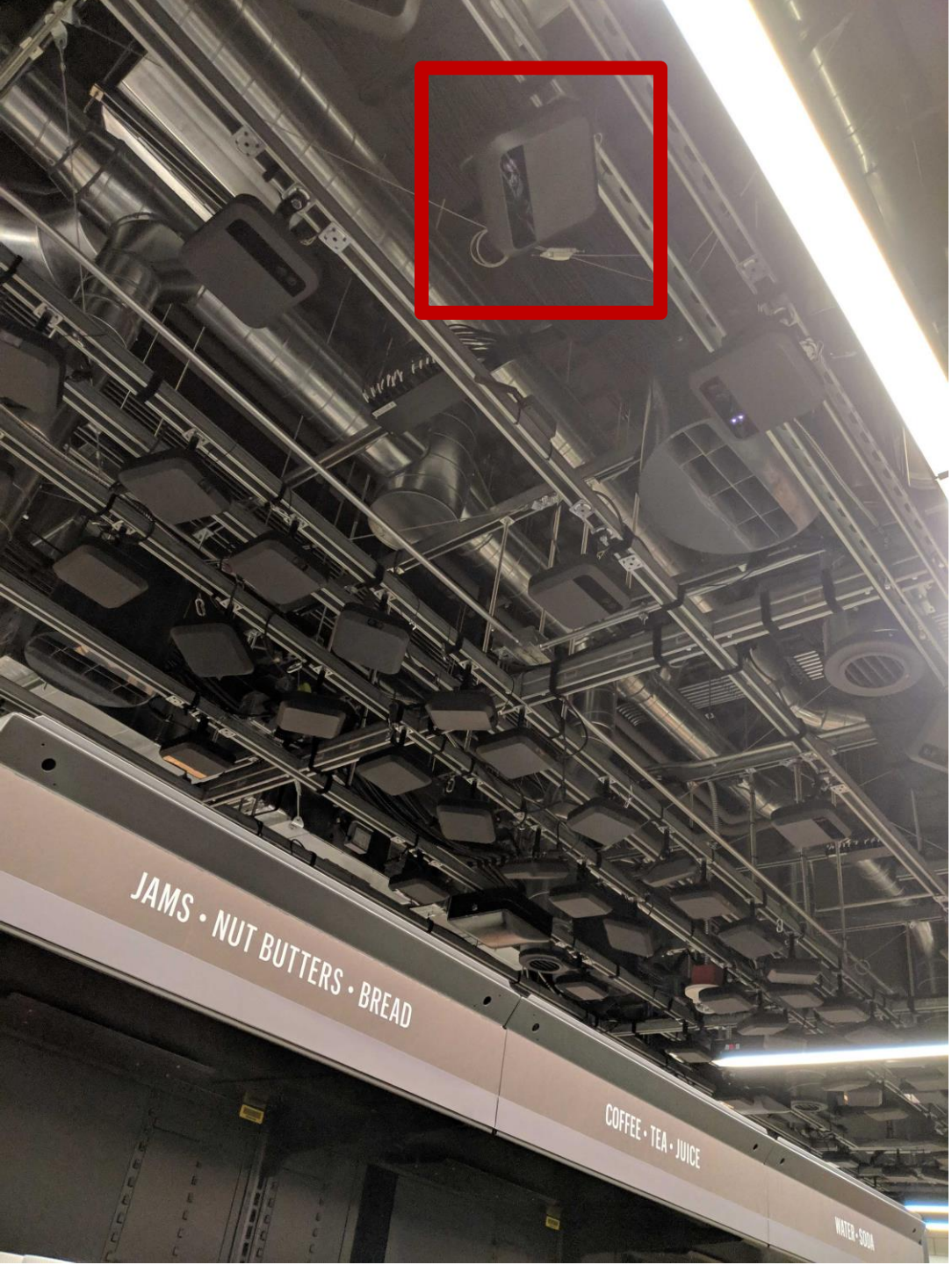


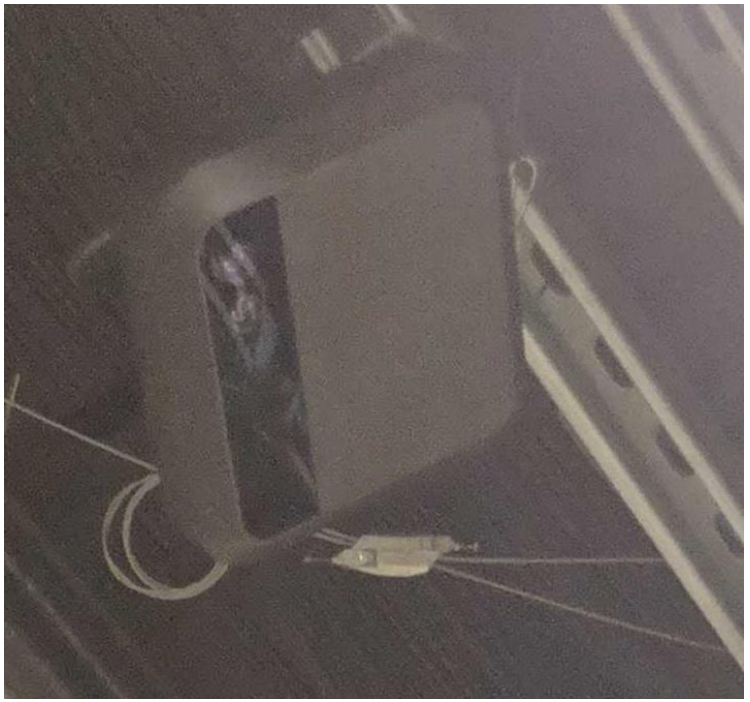
How does it work?

Think-Pair-Share



How does it work?

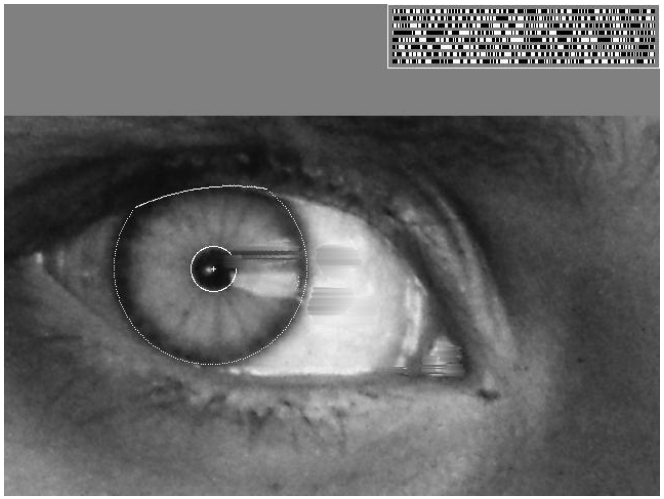


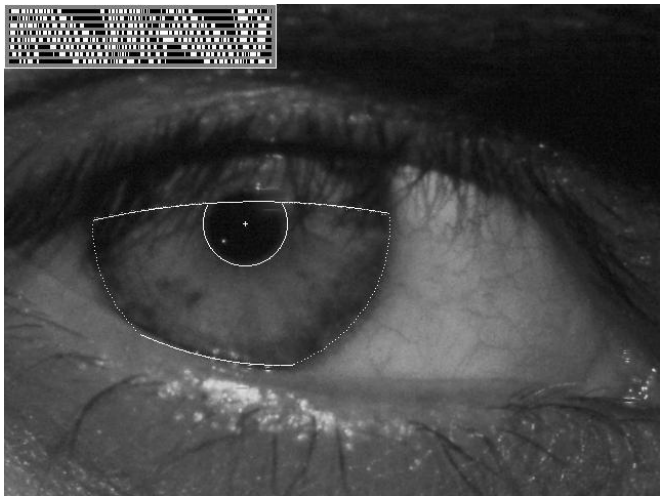


Thanks to Vivek Ramanujan

Vision-based biometrics

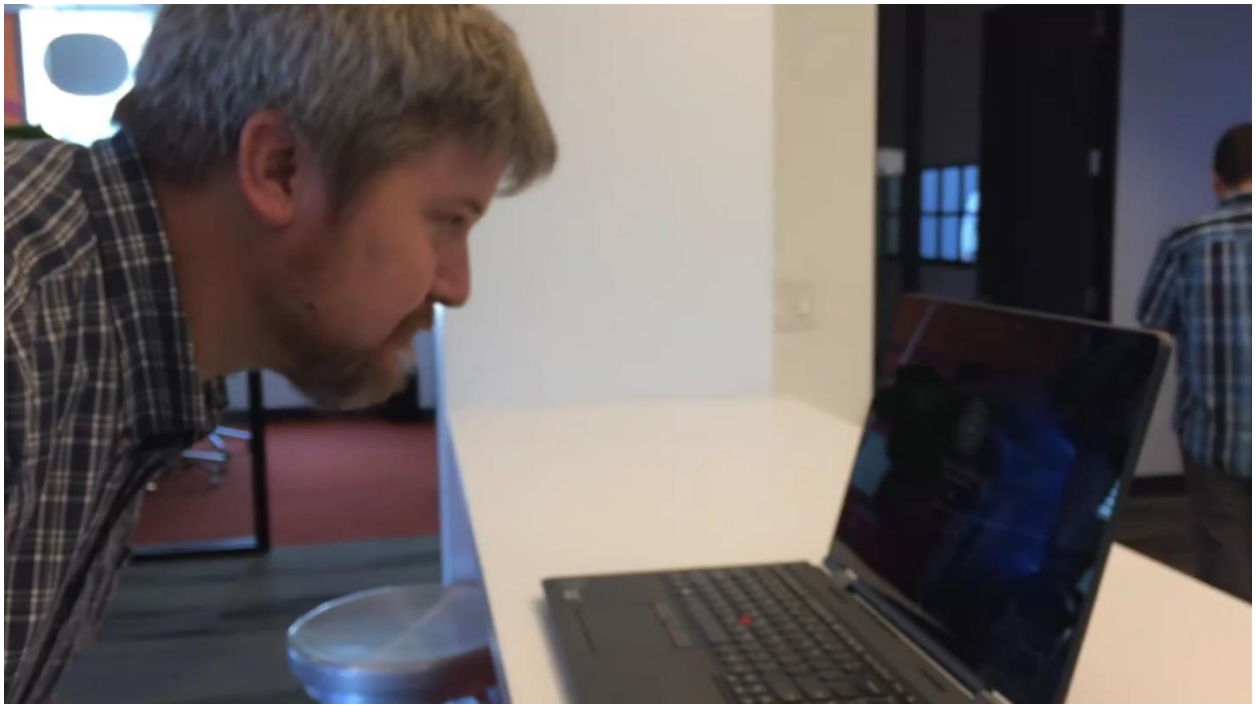
"*How the Afghan Girl was Identified by Her Iris Patterns*"
Read the story (Wikipedia)





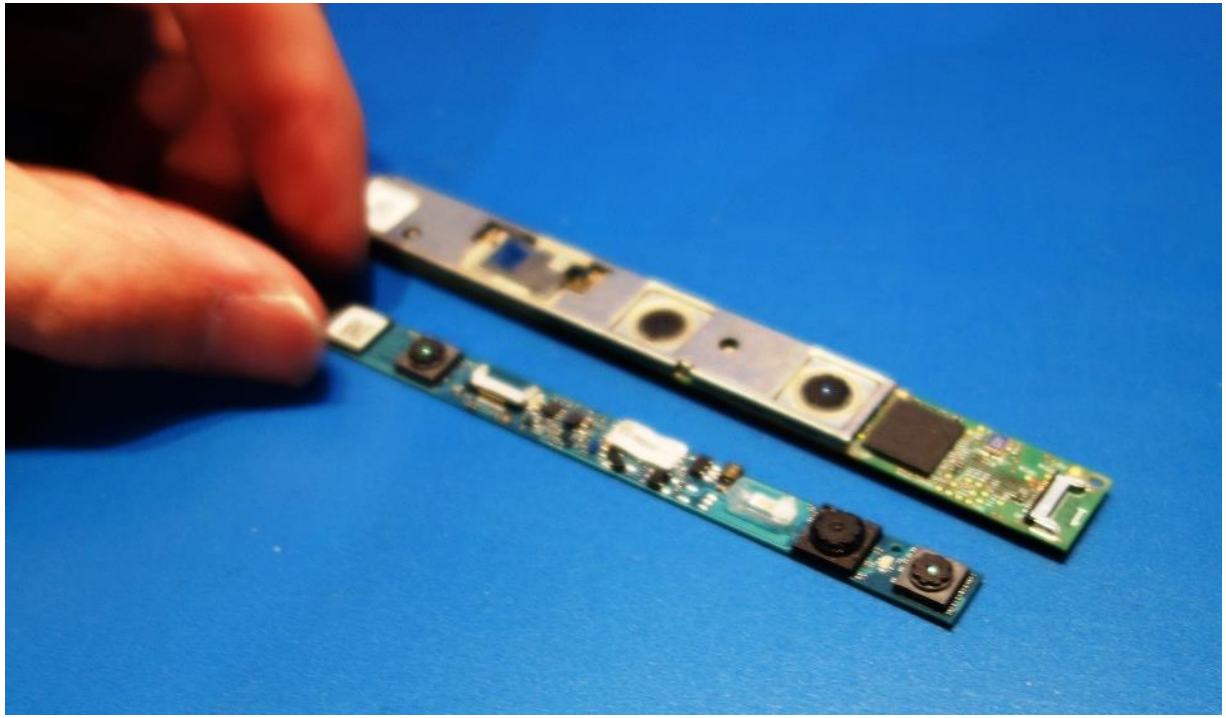
JH

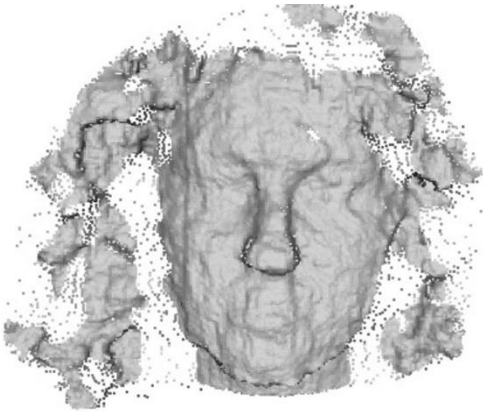
Facial login without a password…



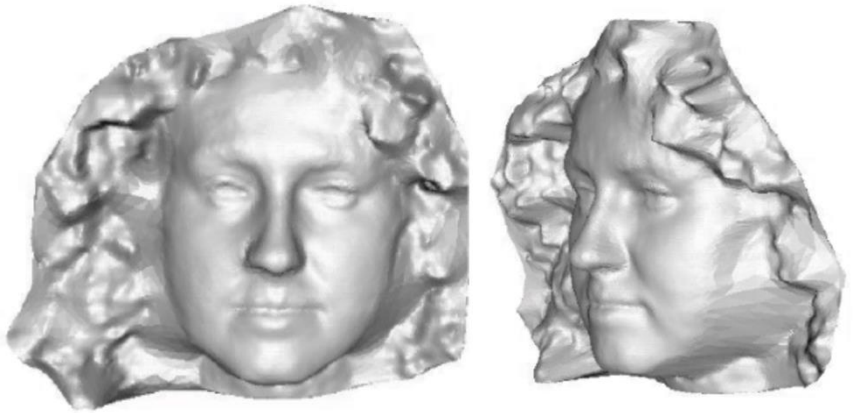
Facial login without a password…

Facial login without a password…





Single depth frame



Reconstructed 3D mesh

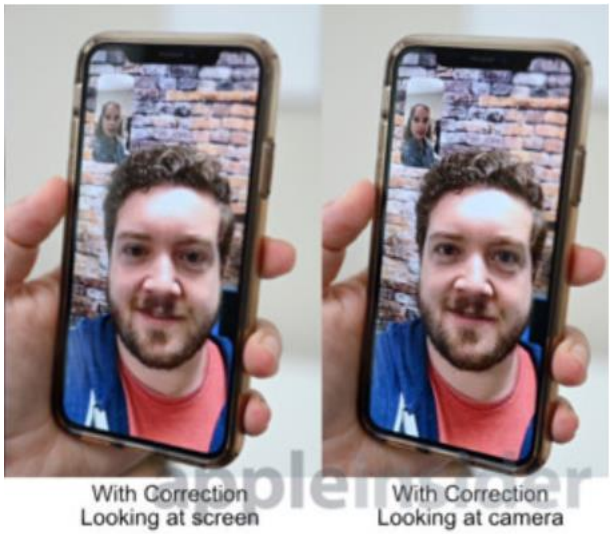
Liang et al. 2014

Video call eye gaze correction

Kuster et al., SIGGRAPH Asia 2012

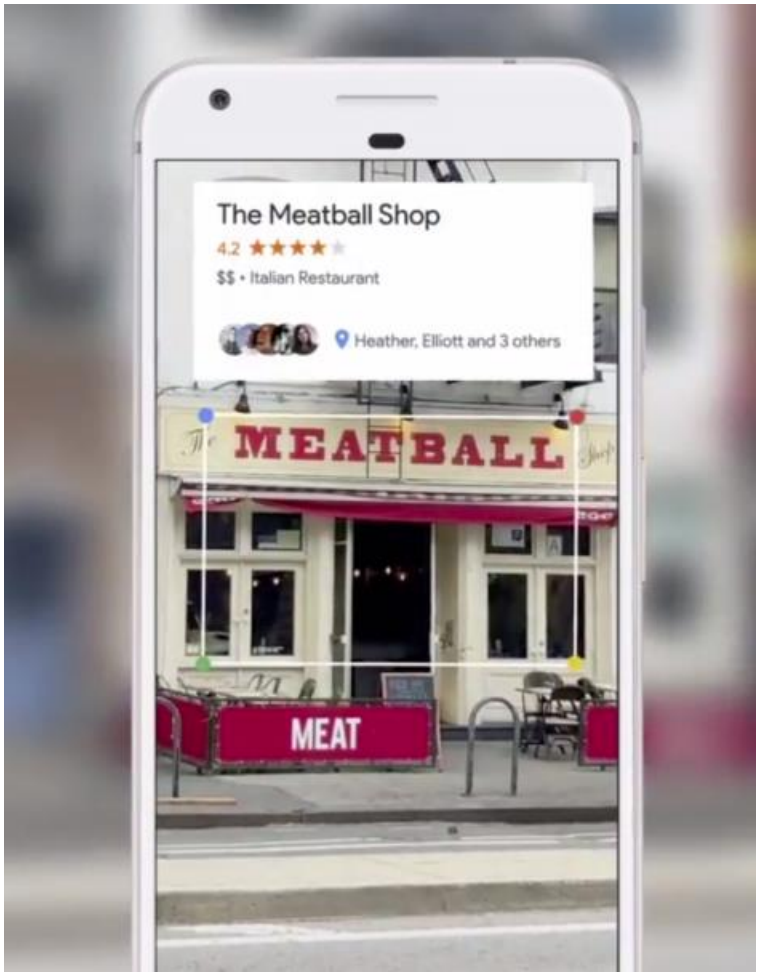
- https://cgl.ethz.ch/publications/papers/paperKus12.php

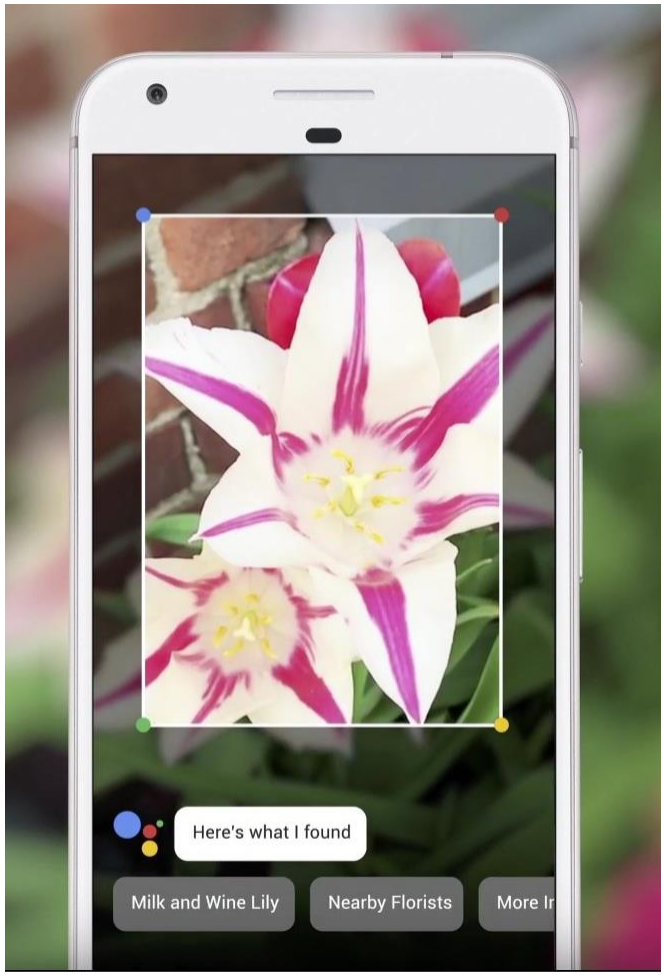
Apple FaceTime
Attention Correction



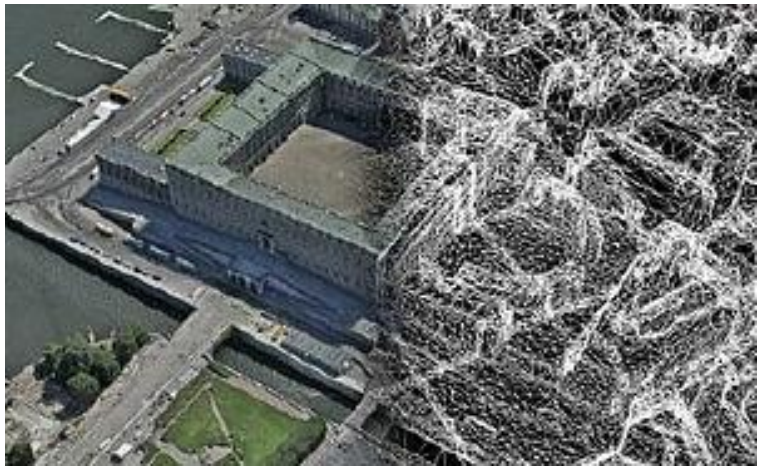
Object recognition (in mobile phones)

e.g., Google Lens





3D from images

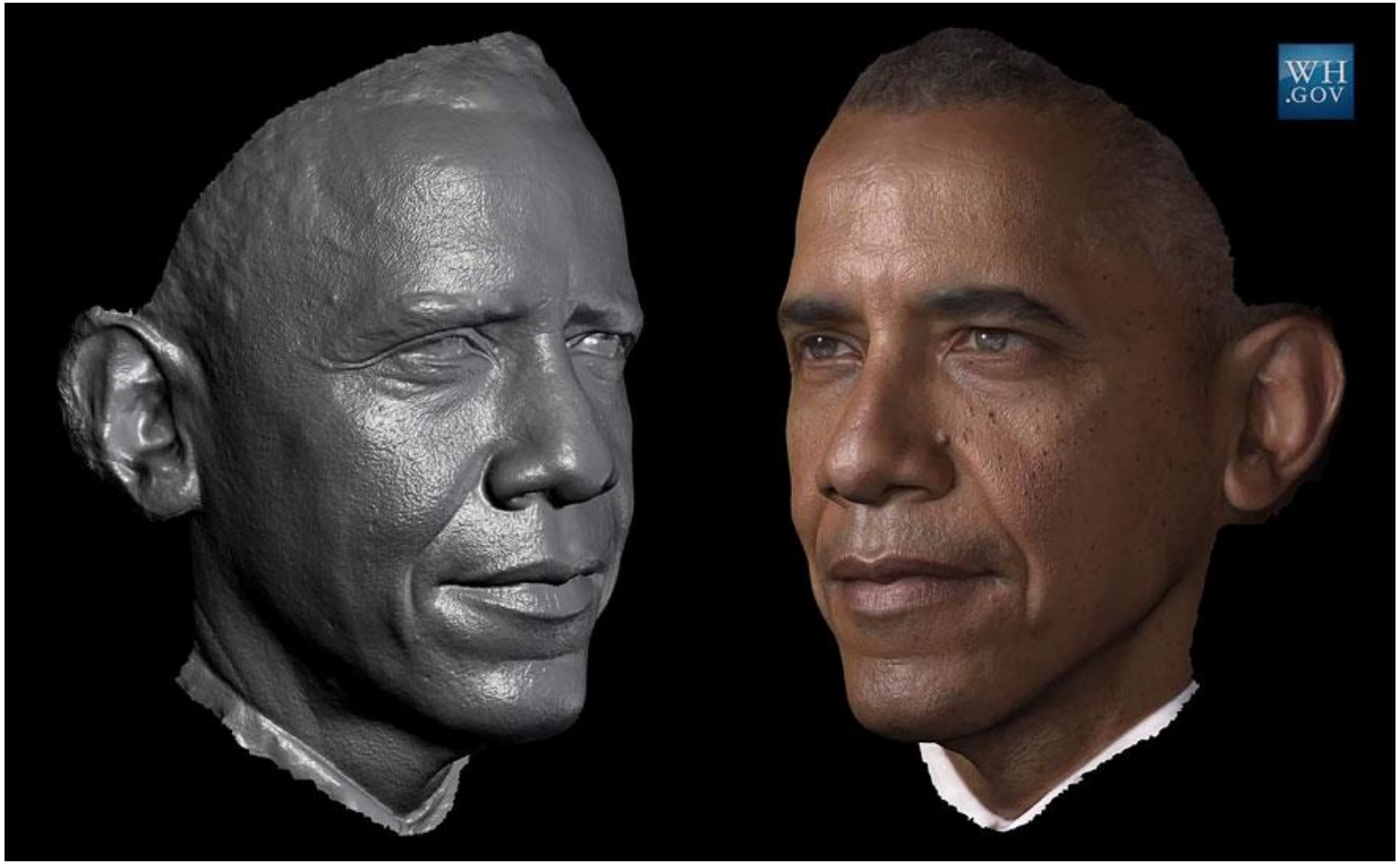


Building Rome in a Day: Agarwal et al. 2009

Human shape capture

Human shape capture

Human shape capture

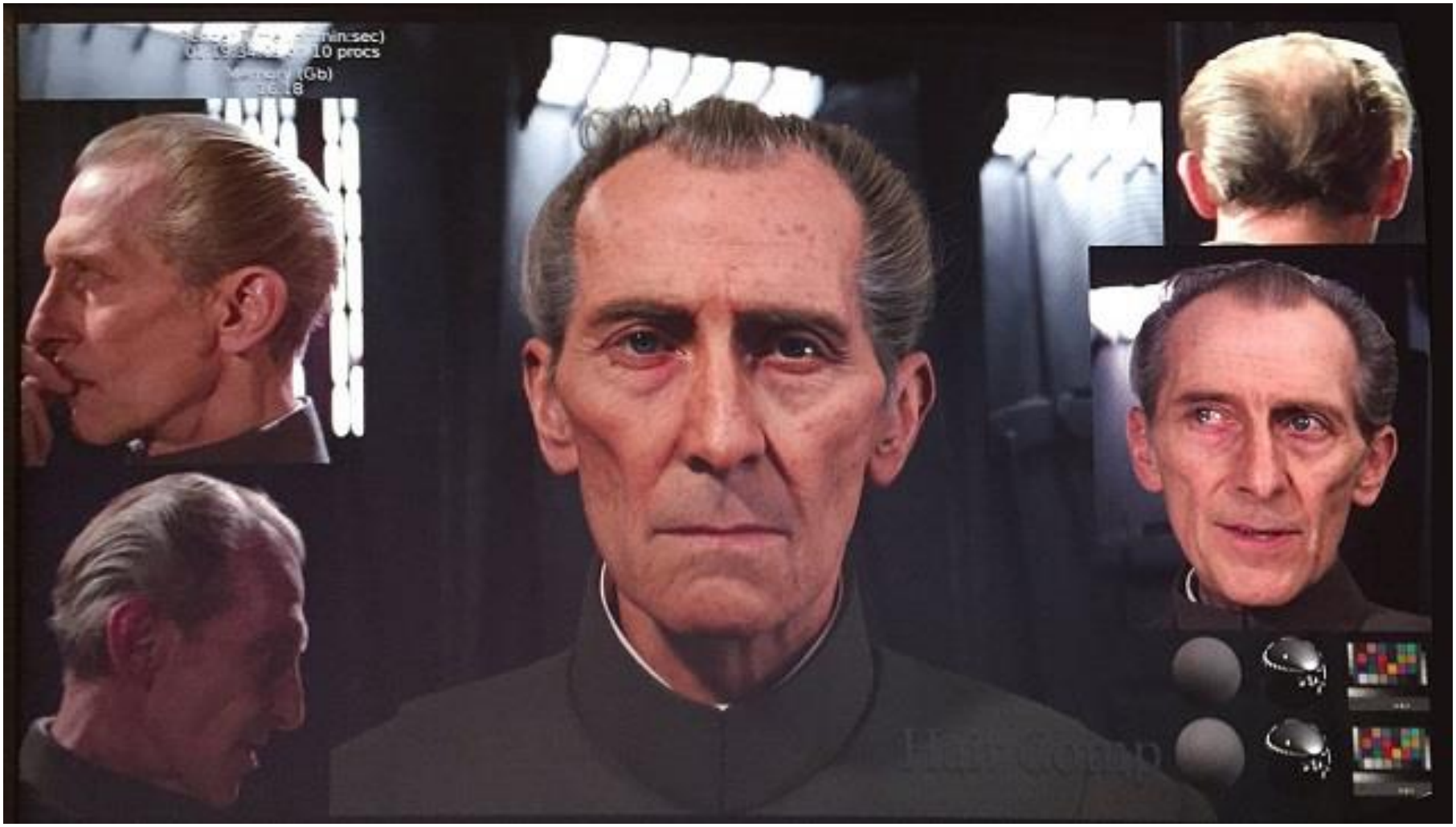


Human shape capture

Special effects: shape capture

Star Wars: Rogue One – Peter Cushing / Admiral Tarkin

Special effects: shape capture



Special effects: motion capture

Interactive Games

Object Recognition:
http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o

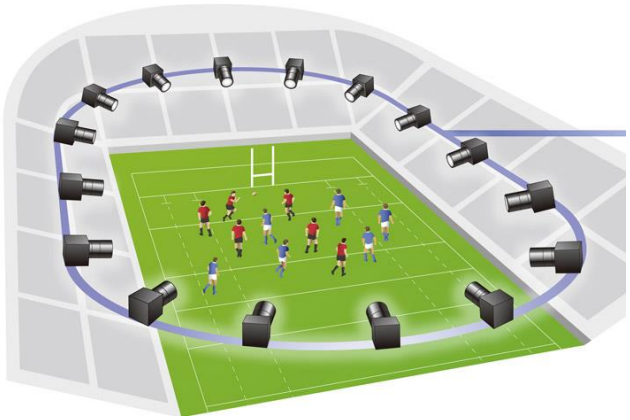
Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg

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

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

JH

Sports



Virtual pitch markings

Sportvision first down line

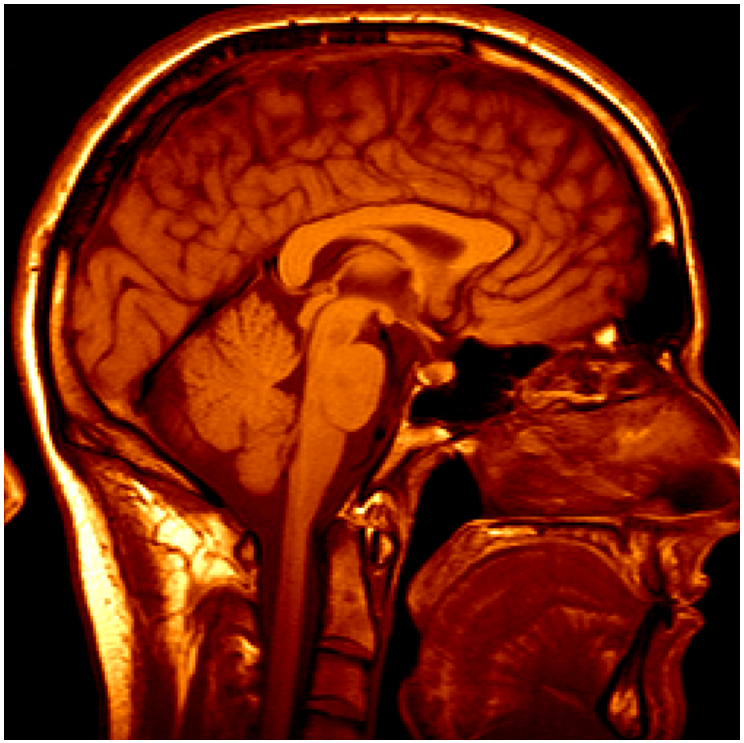
Nice [explanation](#) on [www.howstuffworks.com](#)

Free viewpoint video

[Canon 2017]

JH

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

AutoCars - Uber bought CMU's lab

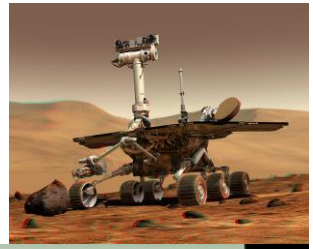

Google Street View
Explore the world at google.com/streetview

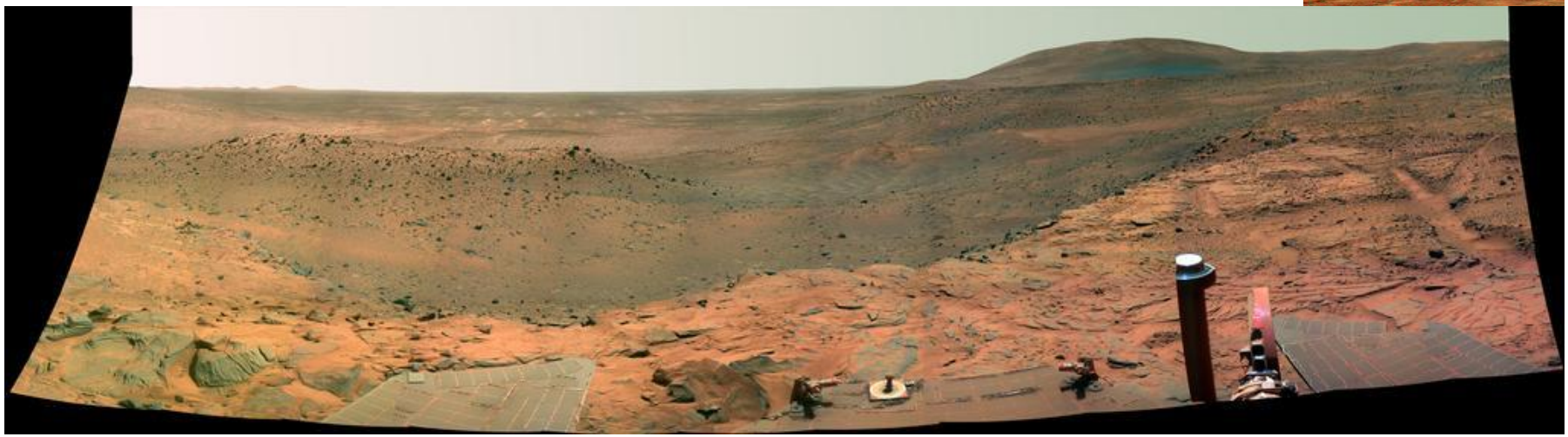
Industrial robots

Vision-guided robots position nut runners on wheels

JH

Vision in spaaaaace





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

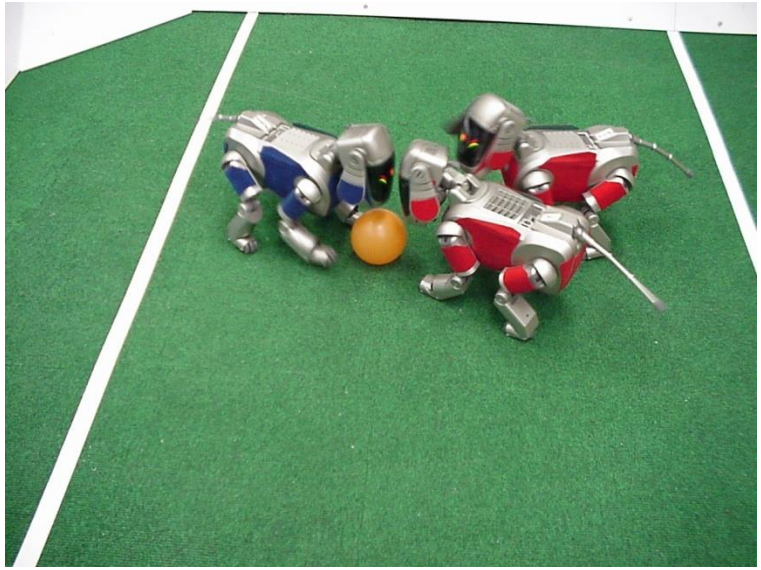
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

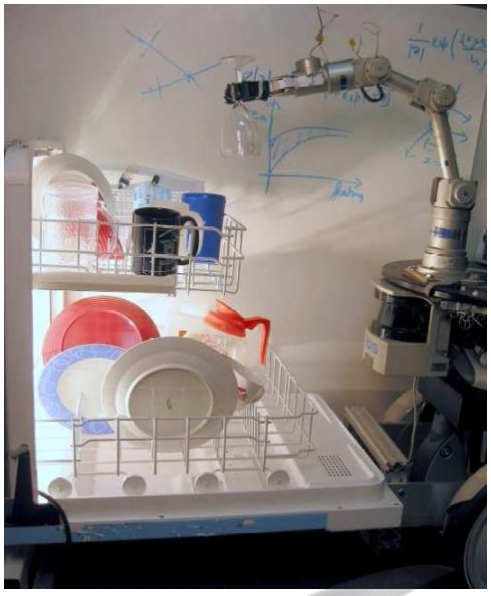
JH

2019 Chang-e 4 Moon Lander

Mobile robots

http://www.robocup.org/





Saxena et al. 2008
STAIR at Stanford

amazon Prime Air

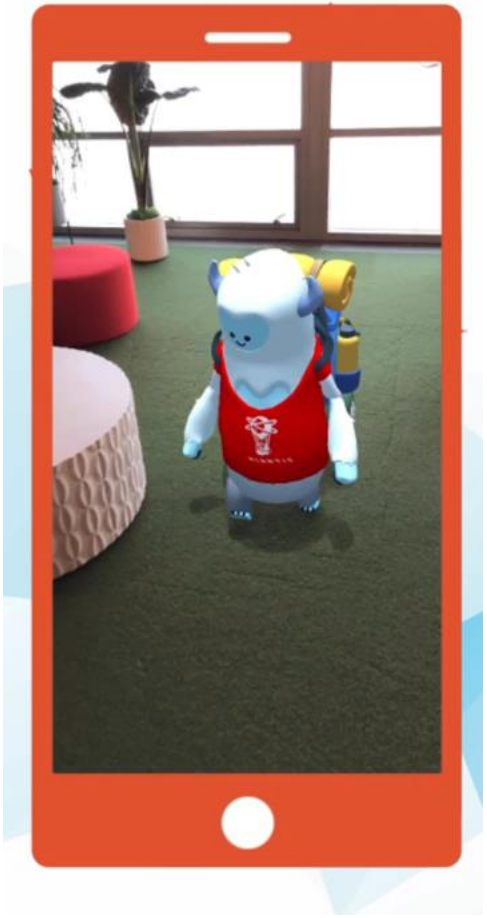
Skydio 2 drone
6x fisheye cameras for obstacle avoidance
Onboard NVIDIA GPU

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



Niantic

Real-time 3D hand pose estimation

Oculus (Quest)

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].”

Computer Vision and Nearby Fields

Derogatory summary of computer vision:

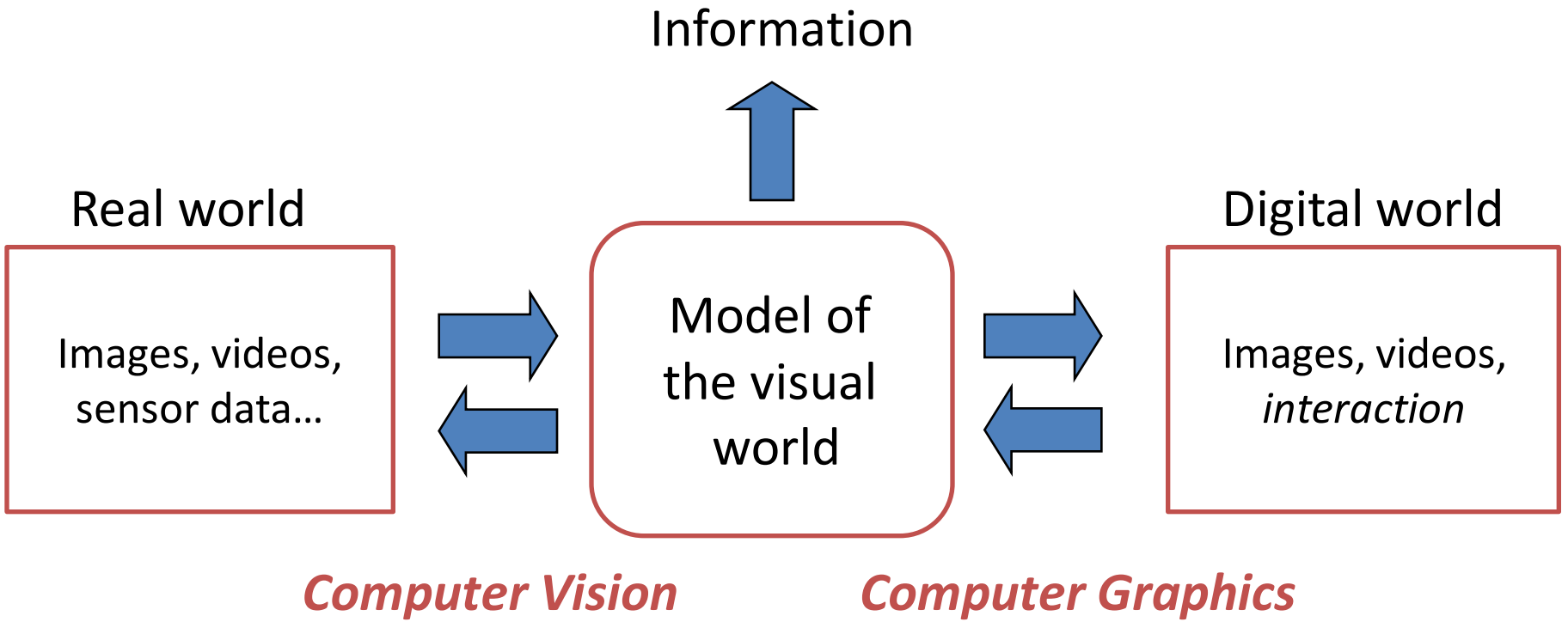
"Machine learning applied to visual data."

JH

Computer Vision and Nearby Fields

Derogatory summary of computer vision:

"Machine learning applied to visual data."



JH

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.

JH

Vision and Society

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

Needs understanding of possible failures.

+

Responsible approach.

+

Techniques to overcome bias.

BBC Sign in News Sport Reel Worklife Travel Future

NEWS

Home Video World US & Canada UK Business Tech Science Stories Enter

Technology

Facial recognition: EU considers ban of up to five years

17 January 2020 Share

The European Commission has revealed it is considering a ban on the use of facial recognition in public areas for up to five years.

"Regulators want time to work out how to prevent the technology being abused.

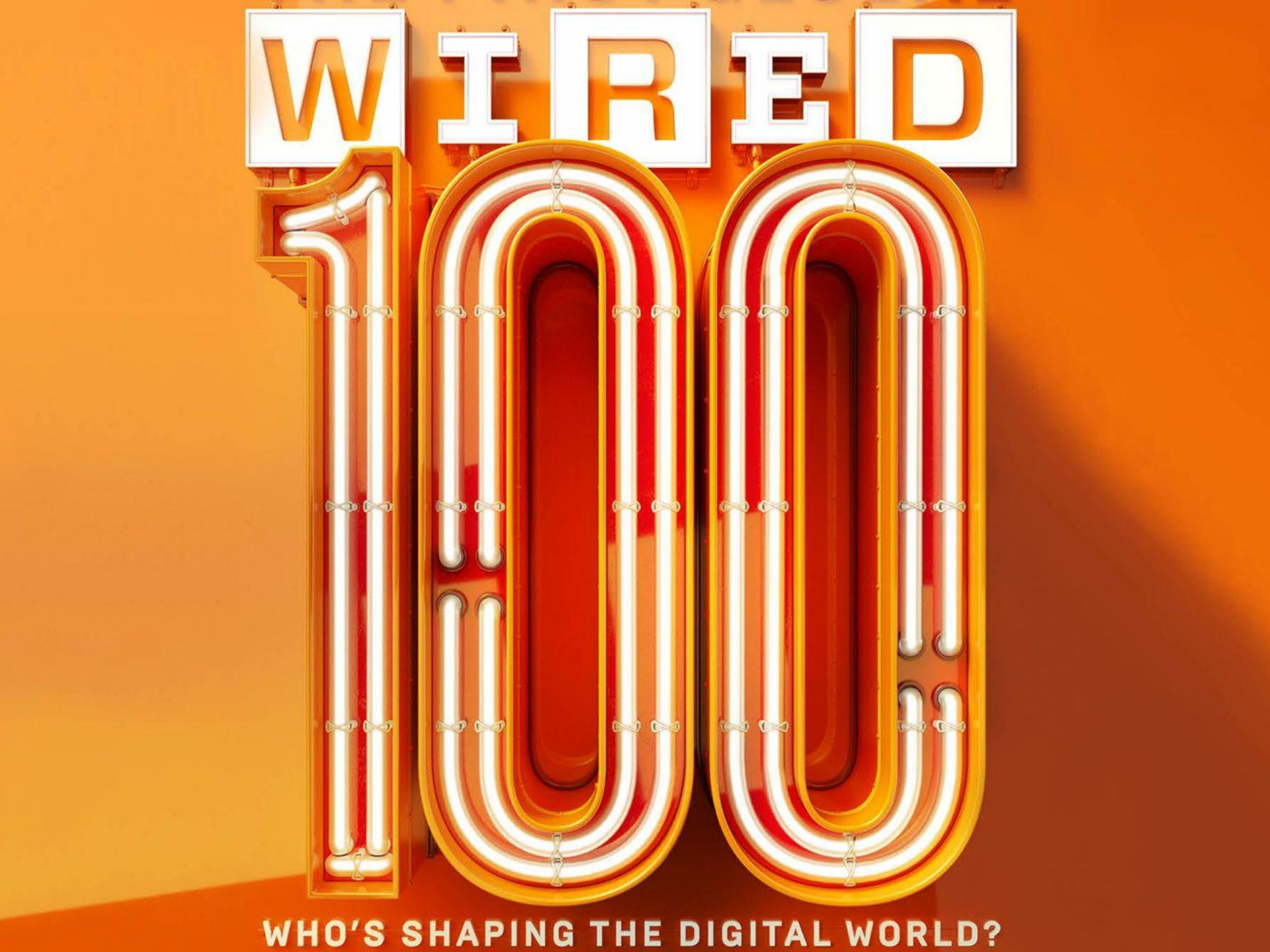
The technology allows faces captured on CCTV to be checked in real time against watch lists, often compiled by police.

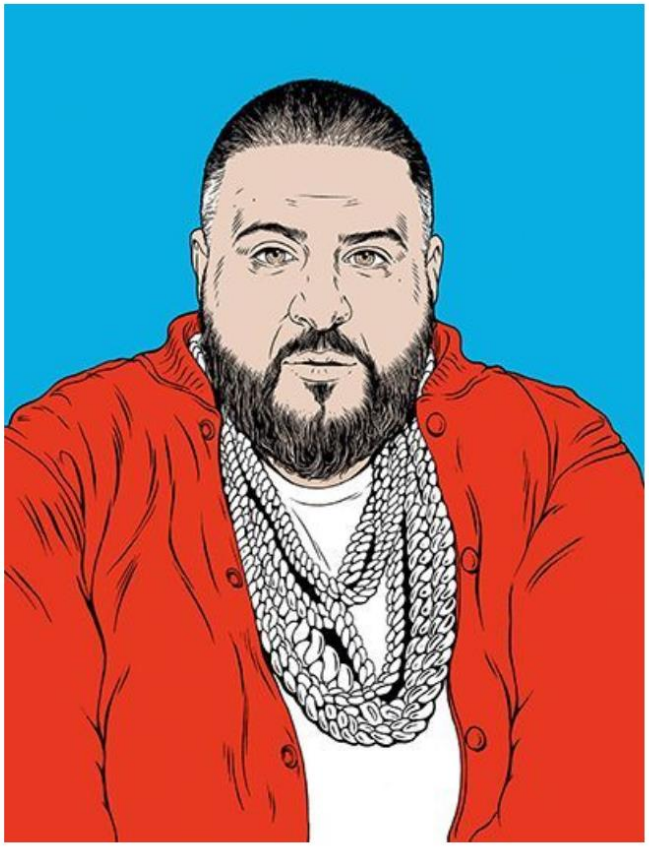
The Commission … suggests that new rules will be introduced to bolster existing regulation surrounding privacy and data rights …

… and urged EU countries to create an authority to monitor the new rules.

During the ban, which would last between three and five years, "a sound methodology for assessing the impacts of this technology and possible risk management measures could be identified and developed".

https://www.bbc.com/news/technology-51148501


WIRED
100
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.

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

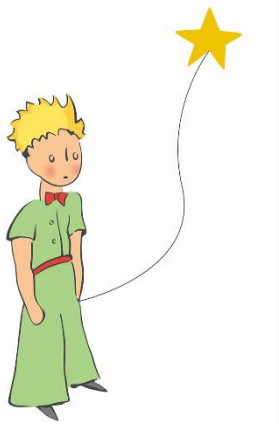
Co-founder and CEO, DeepMind, London

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.

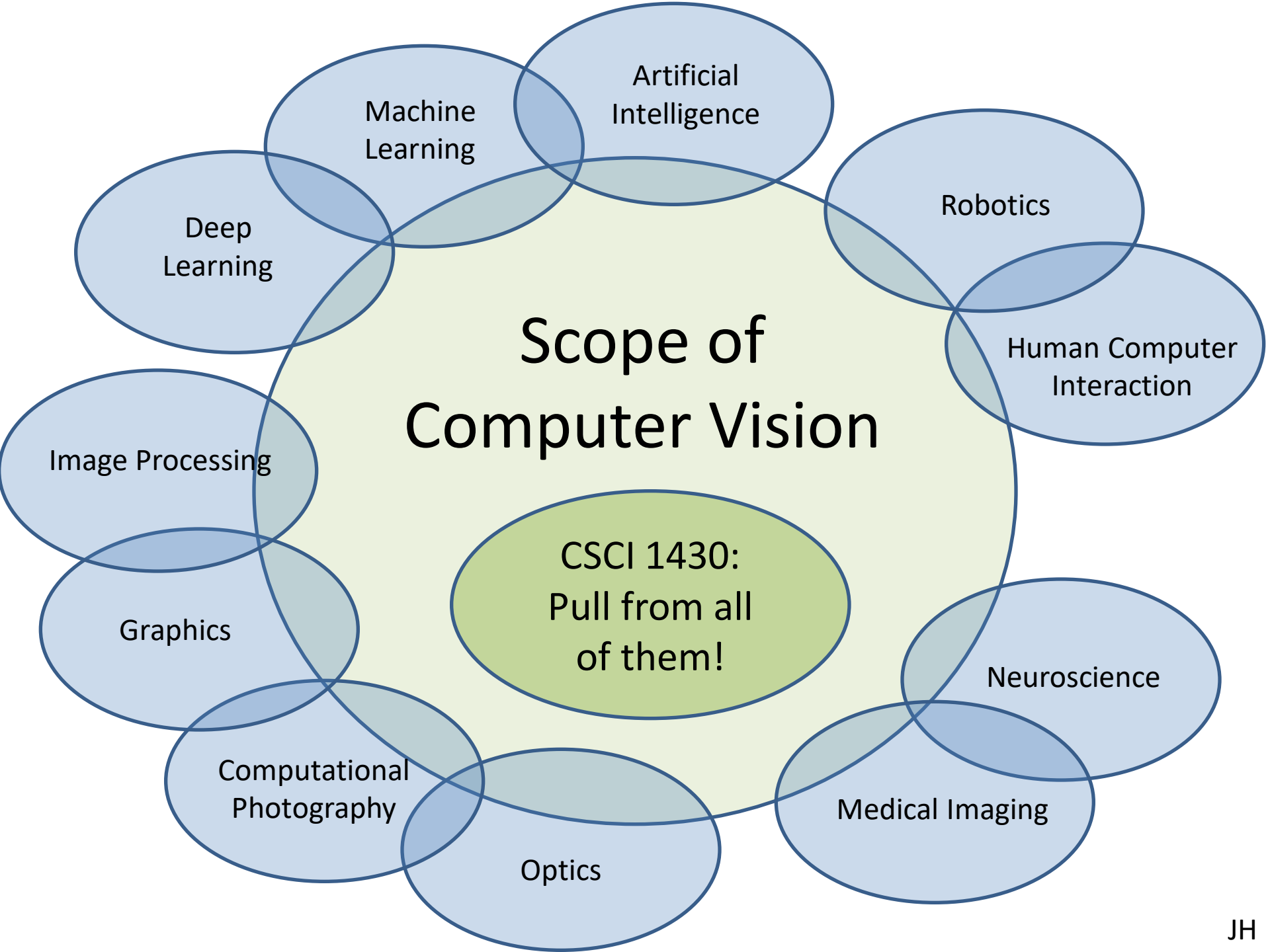
COURSE ETHOS AND SCOPE

"To create the ship is not
to weave the webs,
to forge the nails,
to read the stars,

but to give the taste of the sea."



Saint Exupéry


Scope of Computer Vision
CSCI 1430: Pull from all of them!
Artificial Intelligence
Machine Learning
Deep Learning
Robotics
Human Computer Interaction
Image Processing
Graphics
Computational Photography
Optics
Neuroscience
Medical Imaging
JH

Prerequisites

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

This is an upper-level course.

We move fast.

If you're rusty, we point you to refreshers.

HERE TO HELP!


Jason Senthil

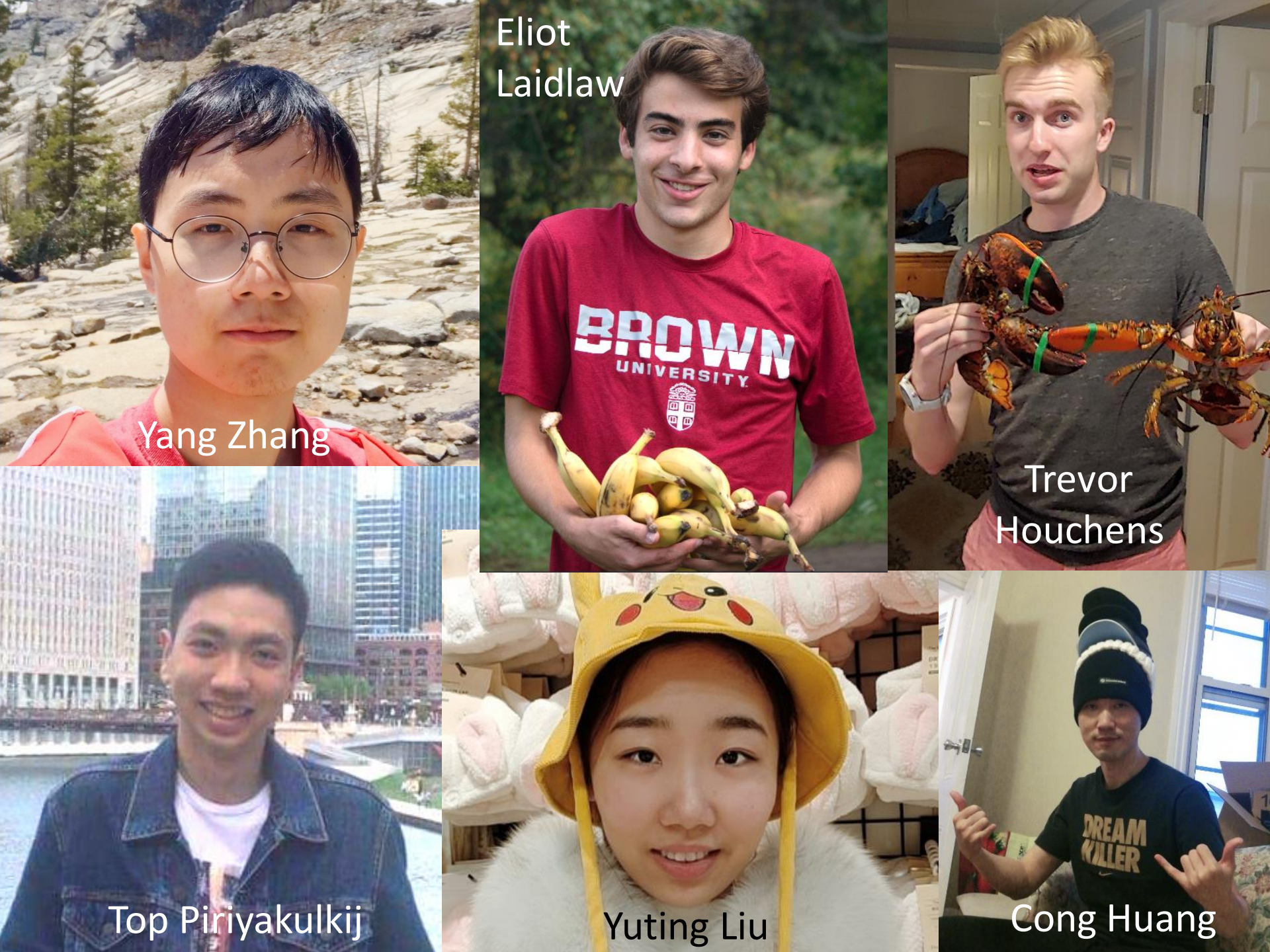

Mary
Dong


Joy Zheng
HACKERMAN

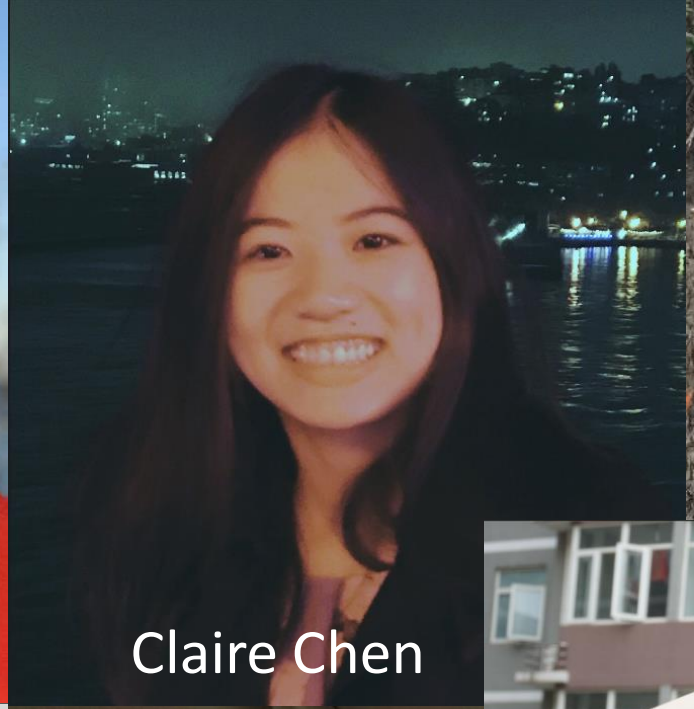

Raymond Cao


Kyle Cui


Andrew Park


Yang Zhang
Eliot Laidlaw
BROWN UNIVERSITY
Trevor Houchens
Top Piriyakulkij
Yuting Liu
DREAM KILLER
Cong Huang

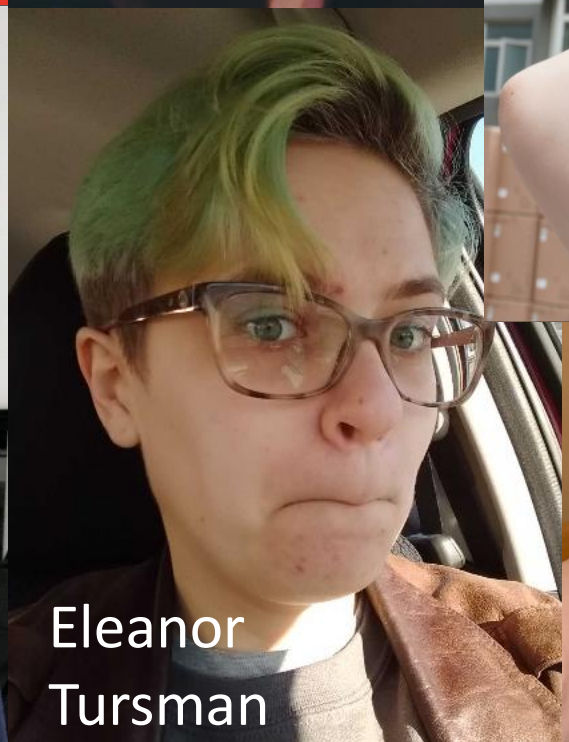

Neev Parikh


Claire Chen


Nine Prasertsup


Qiao
Jiang


Josh Roy


Eleanor
Tursman


Katie Scholl

ETAs!


Isabella
Ting


Erin Simshauser
1h ago
Katie
Friis
lift imitates art
Send a chat

Isa Milefchik (HTA)

George Lee (HTA)

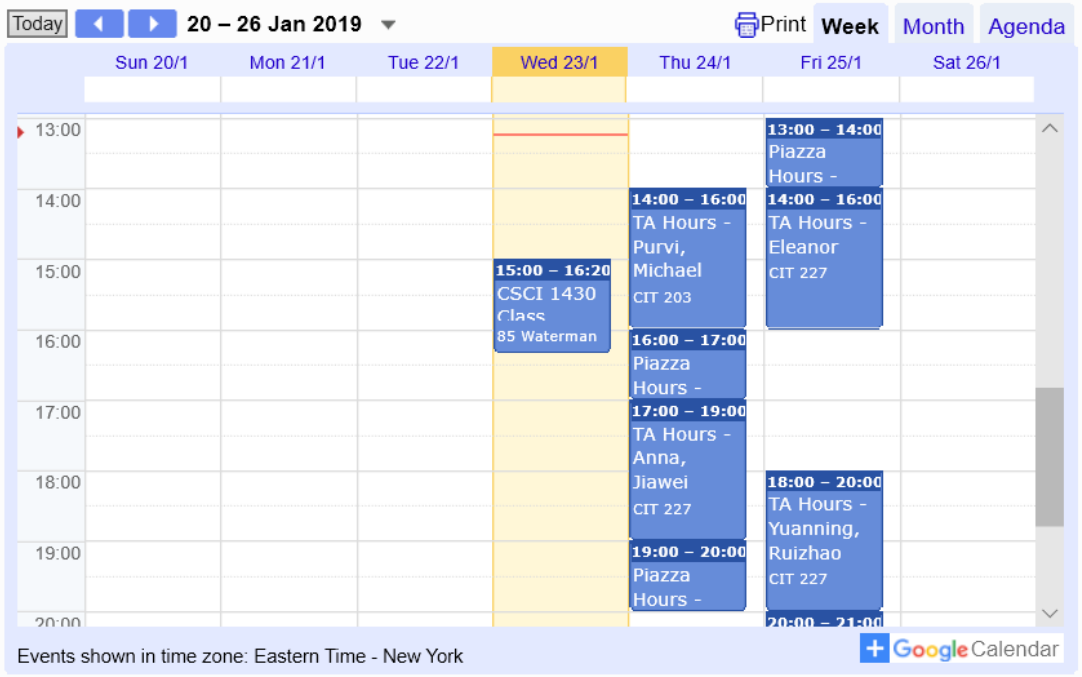
TA Hours

Course webpage

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.
- *cs1430tas@lists.brown.edu*—your second stop for less typical questions.

Office/Piazza Hours



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-

My Office Hours

jamestompkin.com

James Tompkin

Assistant Prof. @ Brown CS

BROWN Computer Science

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 EVERYON
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 technic
prove our understanding of the connections *within* media.

Academic lineage

- Post-doc with Prof. Hanspeter Pfister at the Harvard Paulson School of Engineering and Applic
- Post-doc with Prof. Christian Theobalt at the Max-Planck-Institute for Informatics and the Int
- Research intern with Prof. Wojciech Matusik at Disney Research Cambridge.

COURSE SETUP

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.

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

Waitlist / Override codes

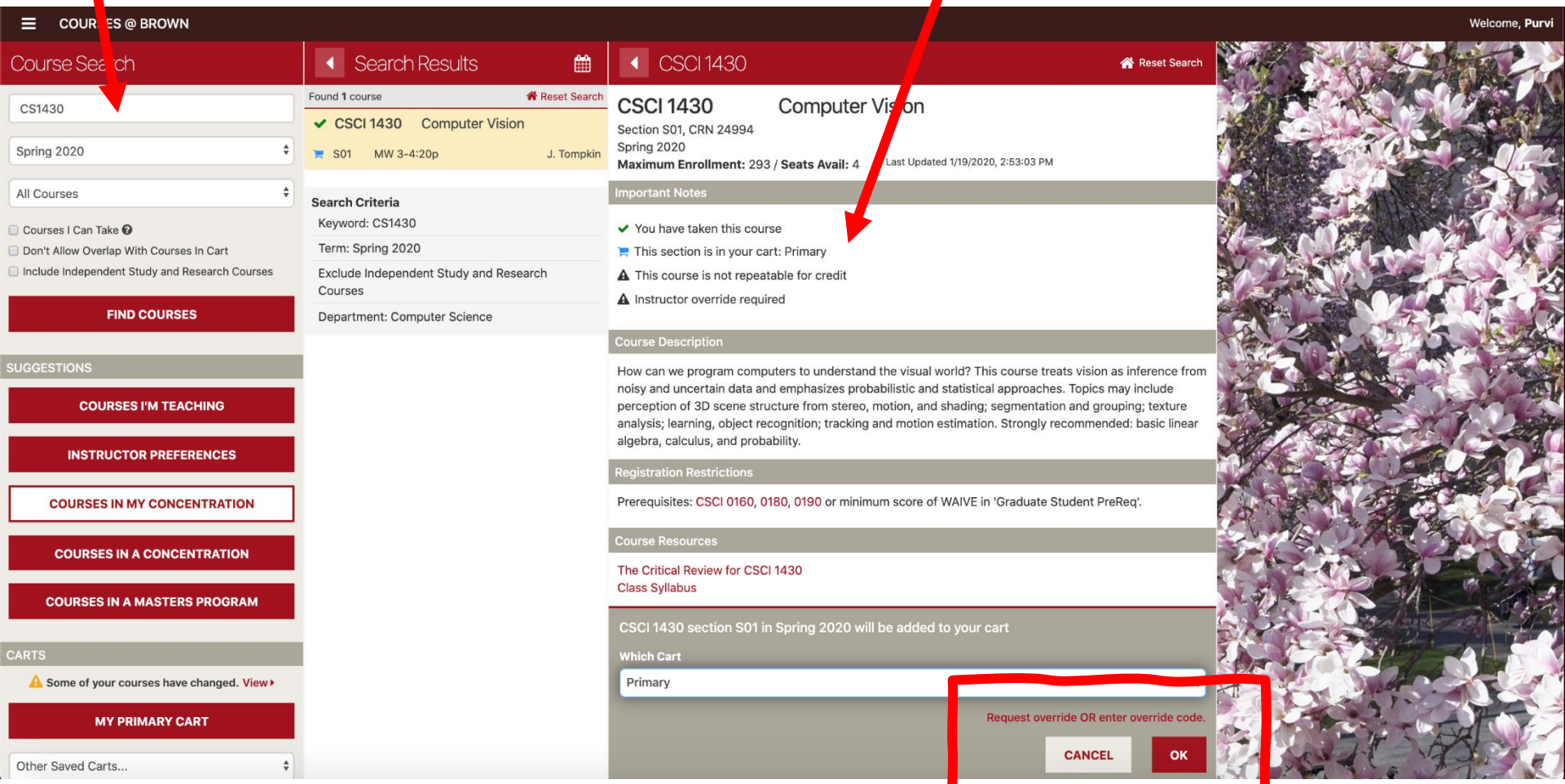
- Yes, we are oversubscribed (50+)
- Yes, we have priority
  - *seniors + grads, juniors, sophomores, fresh*

- We will authorize overrides later in shopping week once it calms down

- Request an override on cab.brown.edu

1. Search for course

2. Make sure added to cart

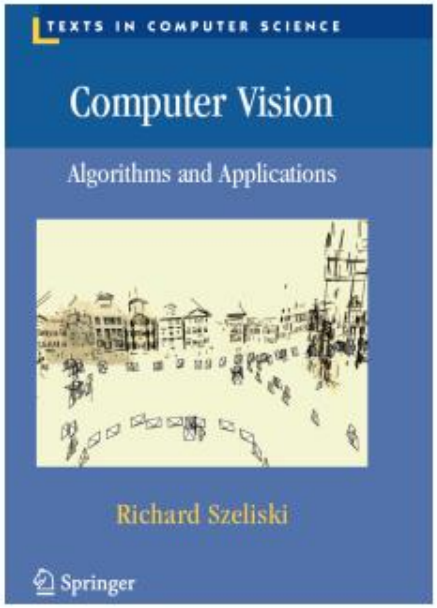


3. Request override

Textbooks

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research



http://szeliski.org/Book/

JH

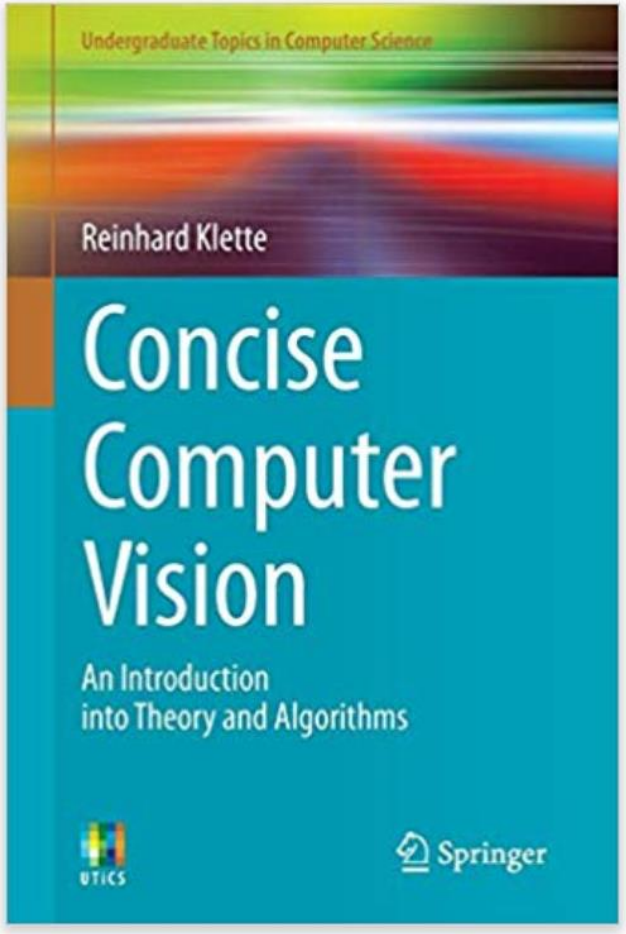
Textbooks

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

Klette
Concise Computer Vision

Digital copy at our library

More introductory than Szeliski.



Textbooks

- 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.

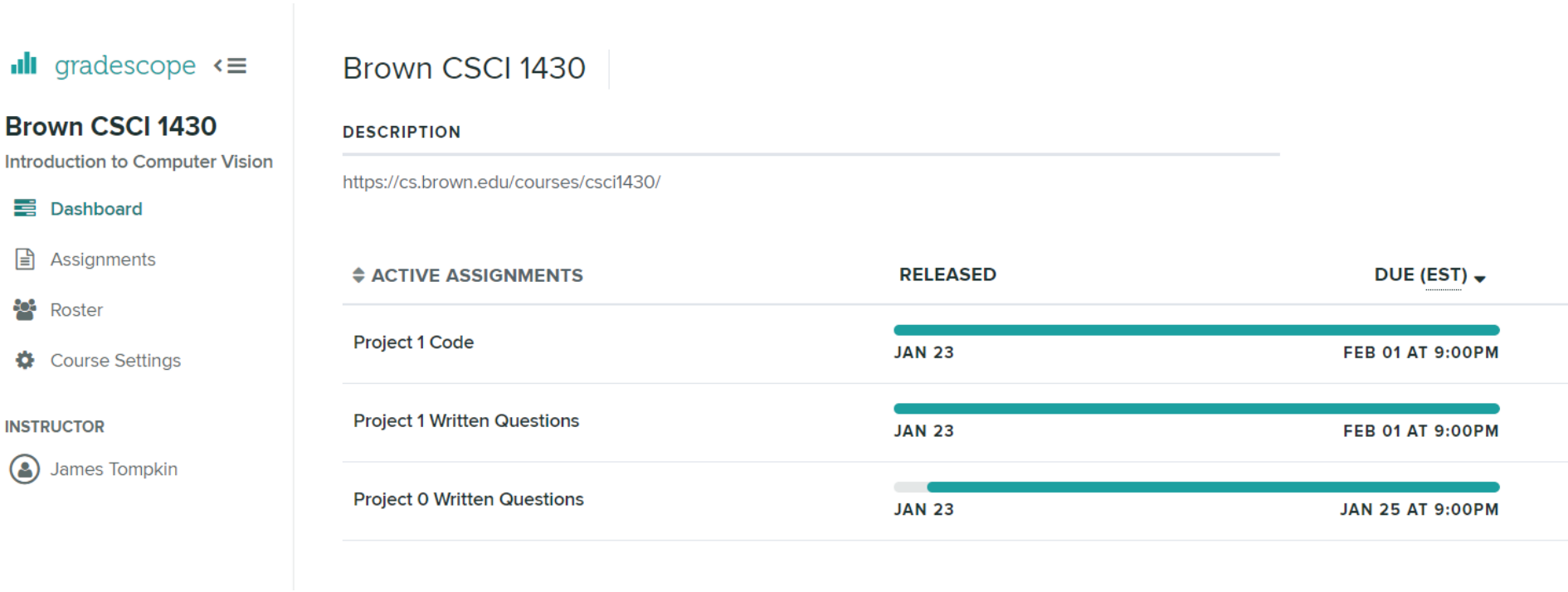
Projects / Grading

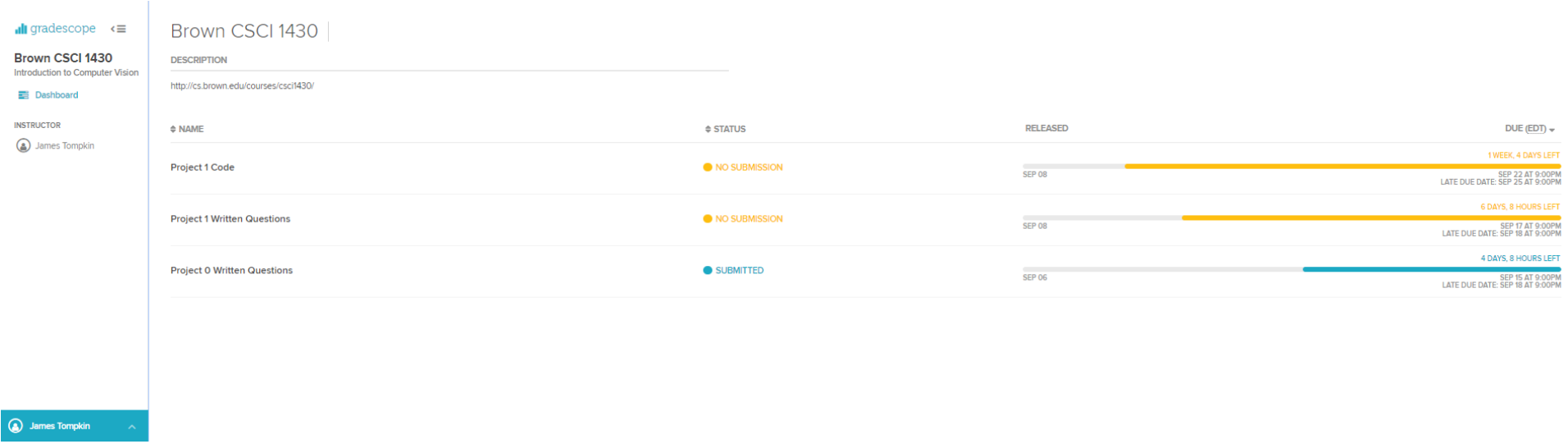
- 100% projects (7 total)

- Project 0: Setup / intro
- Projects 1-5: Structured conceptual / code
- Project 6: Final group project

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

Submission system: Gradescope





- 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 to help us grade efficiently.
  - Use _only_ the space provided
  - Please don’t make more/fewer pages
  - You can put extra pages at the end

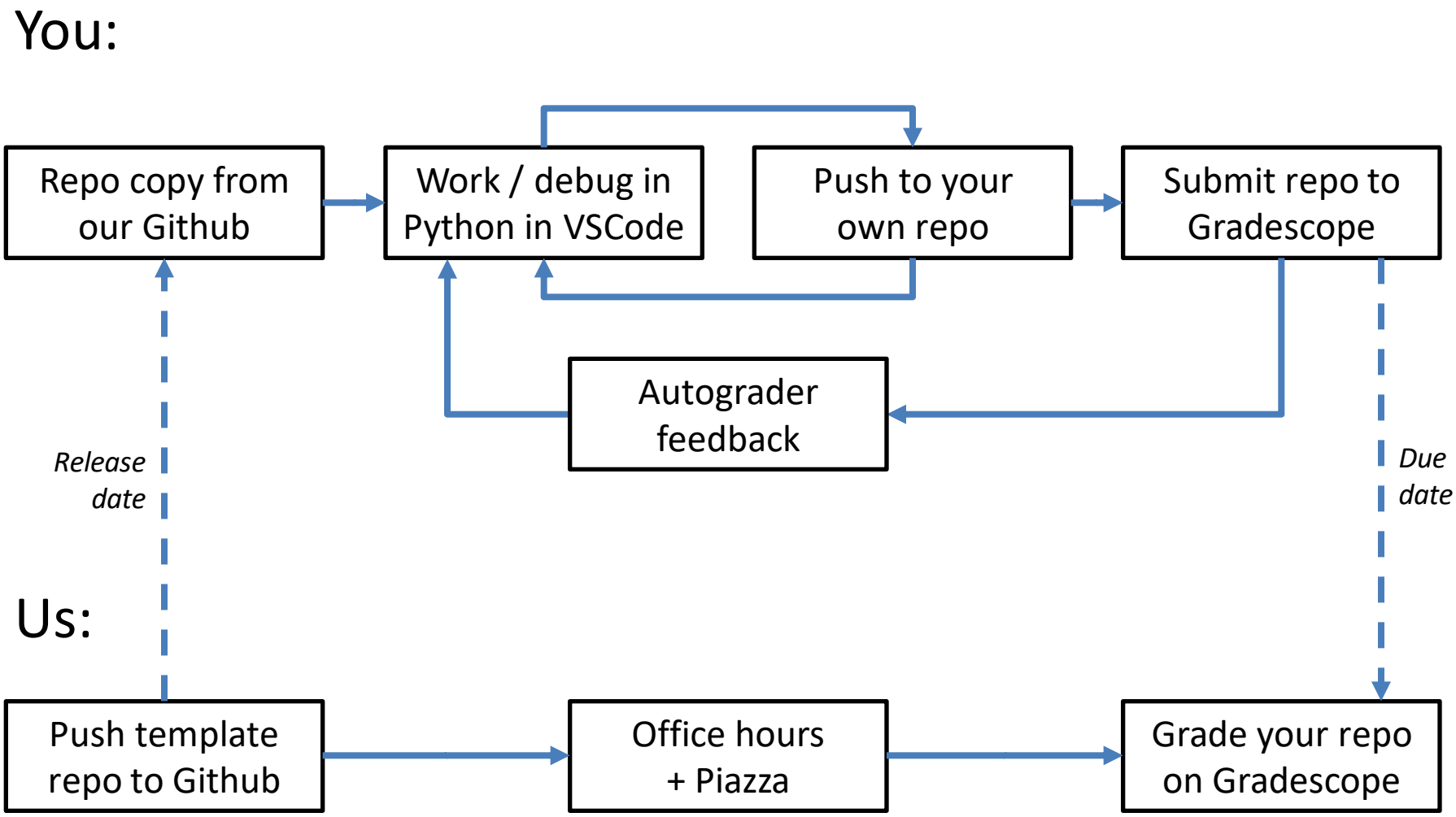
Gear-up Session (video captured):

TONIGHT 6pm Barus and Holley 168

What is Git?
What is Github?
What is Gradescope?
What is Visual Studio Code?
What is Python / Virtual Environment?
How they work together.

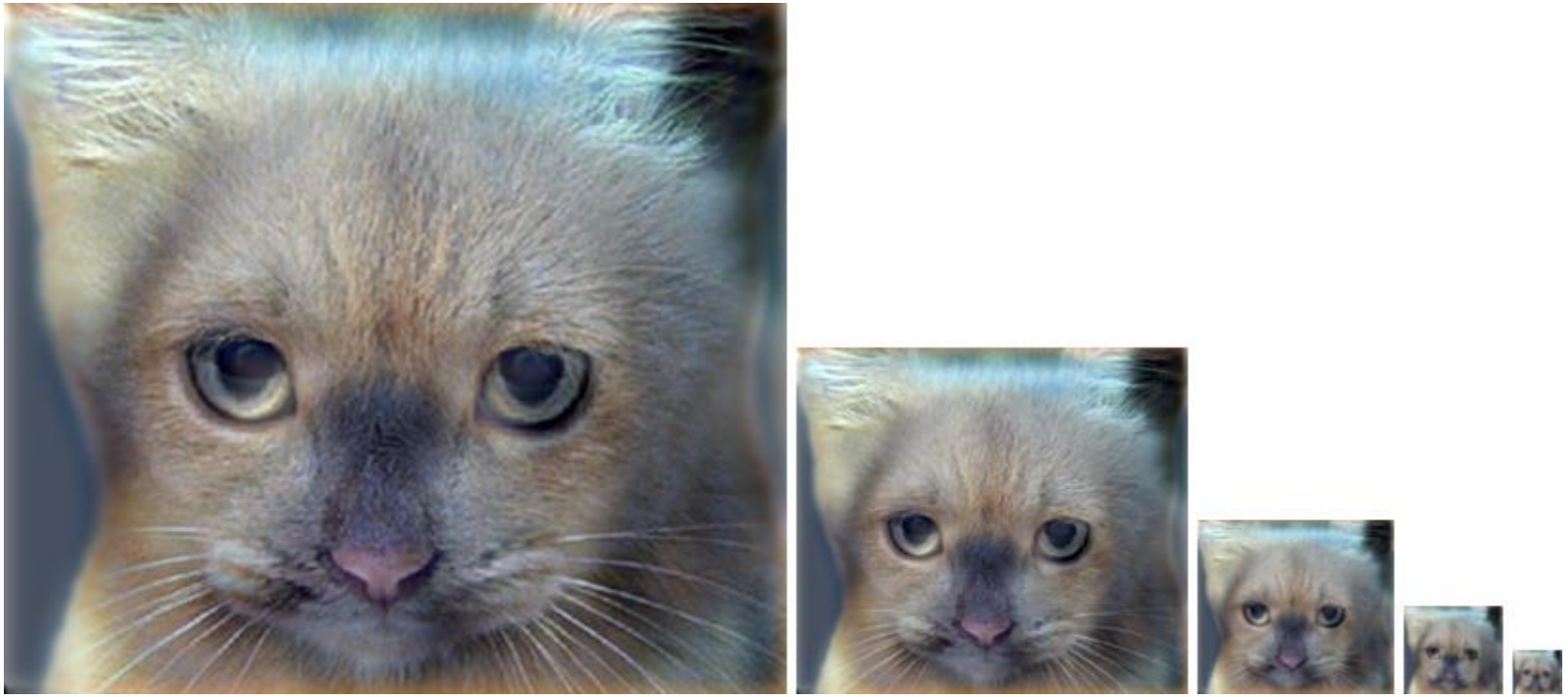
I promise it's worth it.

How it works – project example



Proj 1: Image Filtering and Hybrid Images

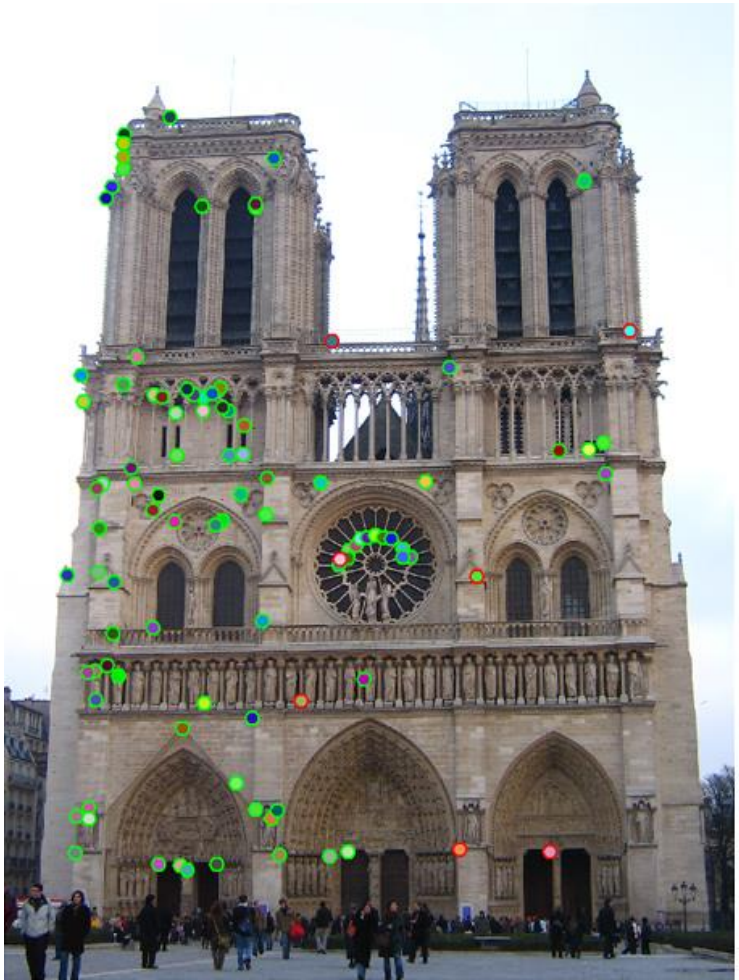
- Implement image filtering to separate high and low frequencies.
- Combine high frequencies and low frequencies from different images to create a scale-dependent image.

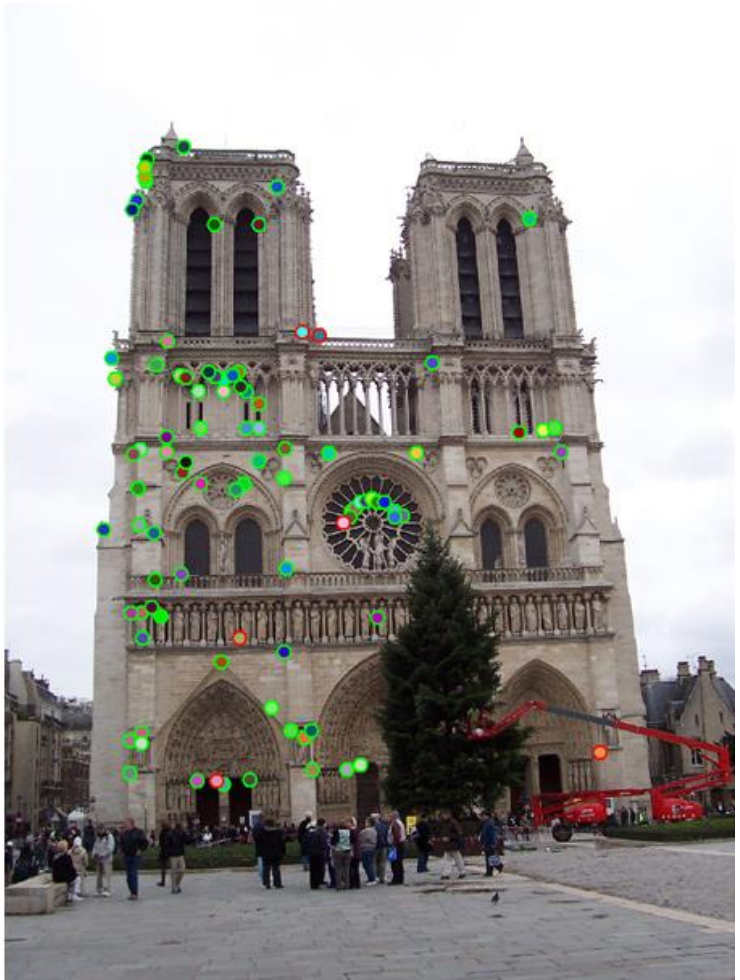


JH

Proj 2: Local Feature Matching

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

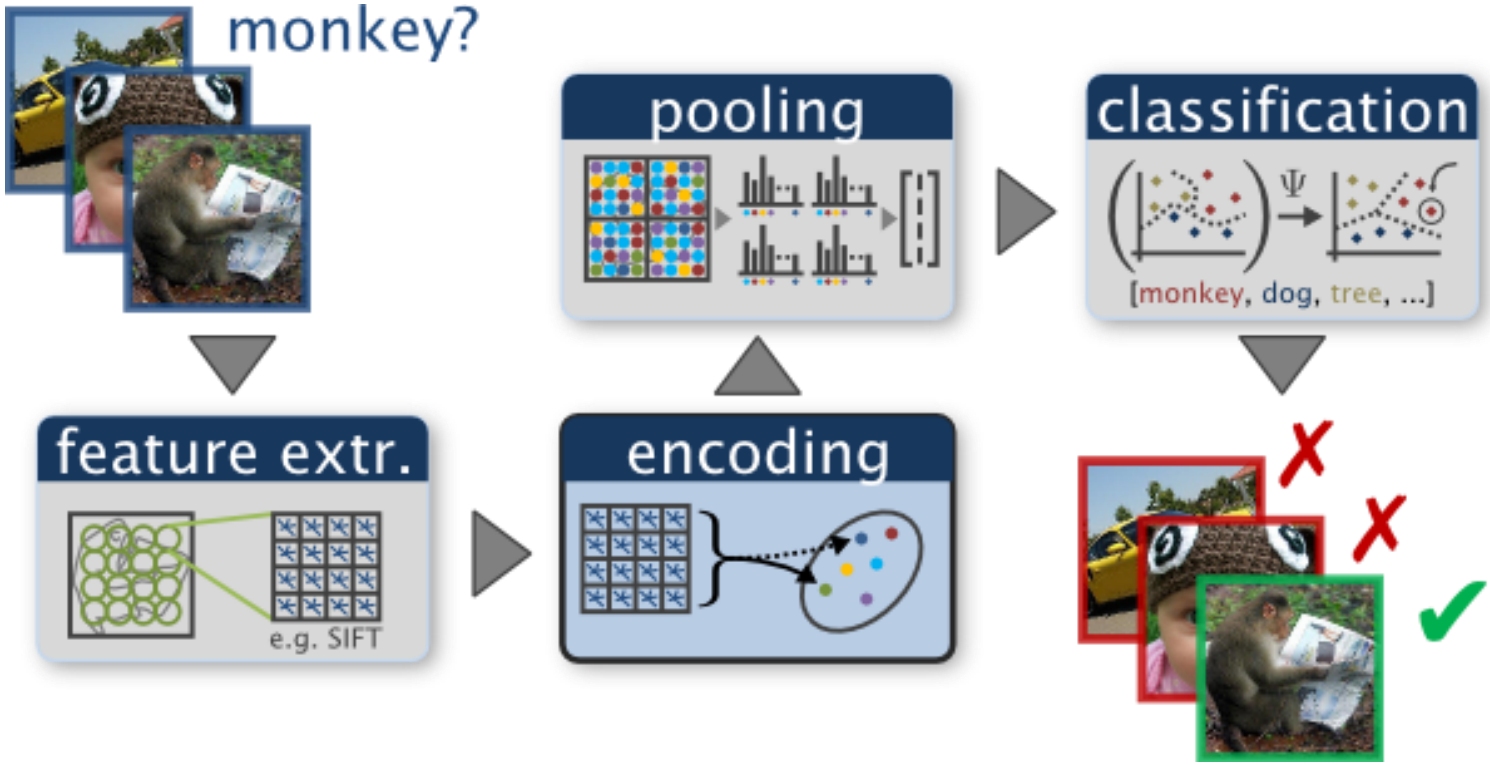




JH

Proj 3: Scene Recognition with Bag of Words

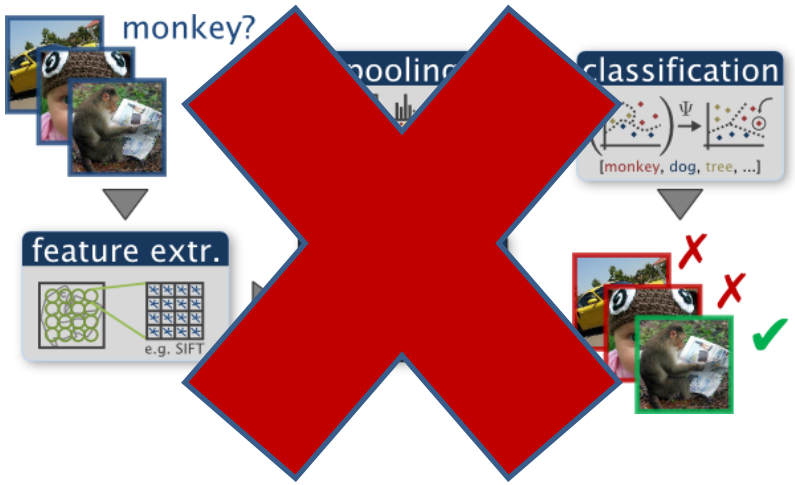
- Quantize local features into a “vocabulary”, describe images as histograms of “visual words”, train classifiers to recognize scenes based on these histograms.

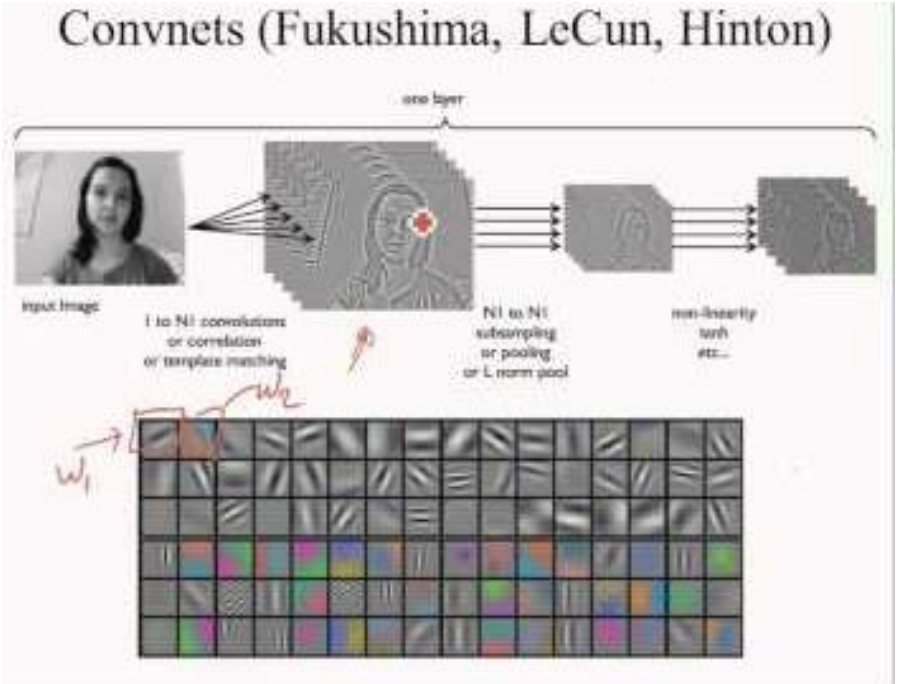


JH

Proj 4: Convolutional Neural Nets

- Proj 3 again, but state of the art.





JH

Proj 5: Multi-view Geometry

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

JH

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
- Computational photography
- Style Transfer
- Adversarial CNN hacking

Nathaniel Nguyen, Kyle Cui, Jules Becker, Leonardo Ko

Friendly neighborhood style transfer

Nathaniel Nguyen, Kyle Cui, Jules Becker, Leonardo Ko

Naive frame-by-frame rendering

Nathaniel Nguyen, Kyle Cui, Jules Becker, Leonardo Ko

Temporal Consistency Constraints

Hyperlapse Stabilization - Michael Mao, Jiaju Ma, James Li

Baseline Video

Result Video

Darius Atmar, Yueyi Sun, Zejiang Shen

CSCI1430 Spring 2019 Final Project

Any questions at this stage?!

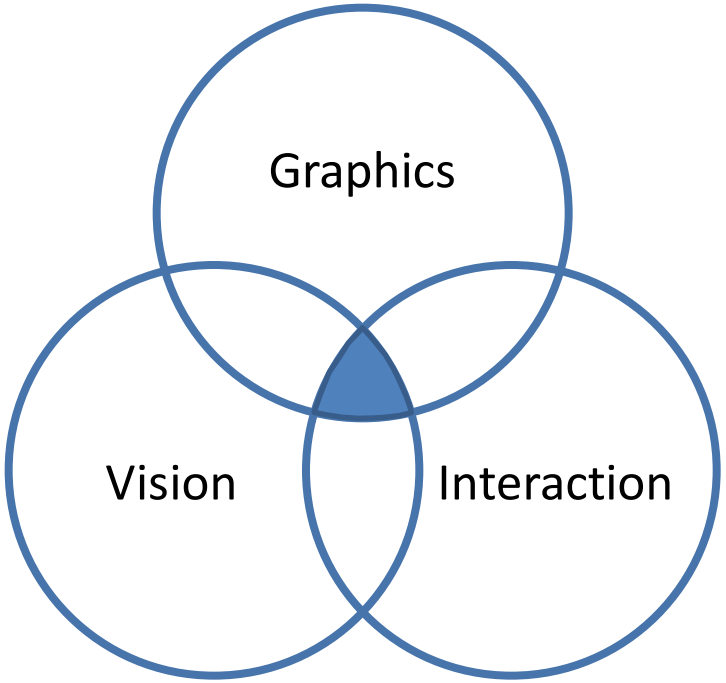
- Waitlist on cab.brown.edu – request override
  - *If you're on the waitlist, still submit project 0+1*

- Gear Up Session TONIGHT 6pm B&H 168
  - *Lecture captured if you can't make it*
- TA hours from *today*

- Project 0 due FRIDAY 9pm
- Project 1 due NEXT FRIDAY 9pm

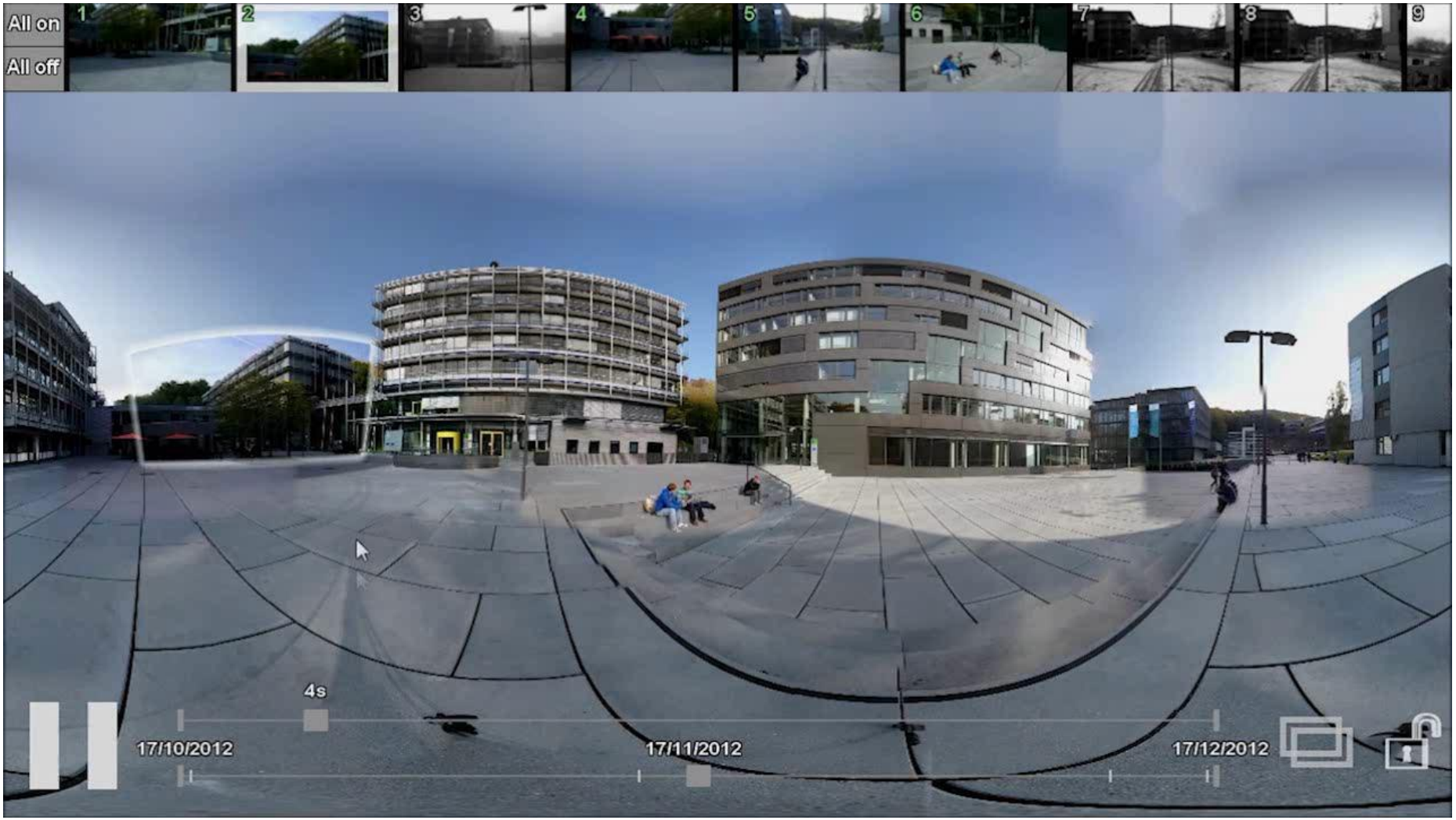
JAMES

I work in here.




Start

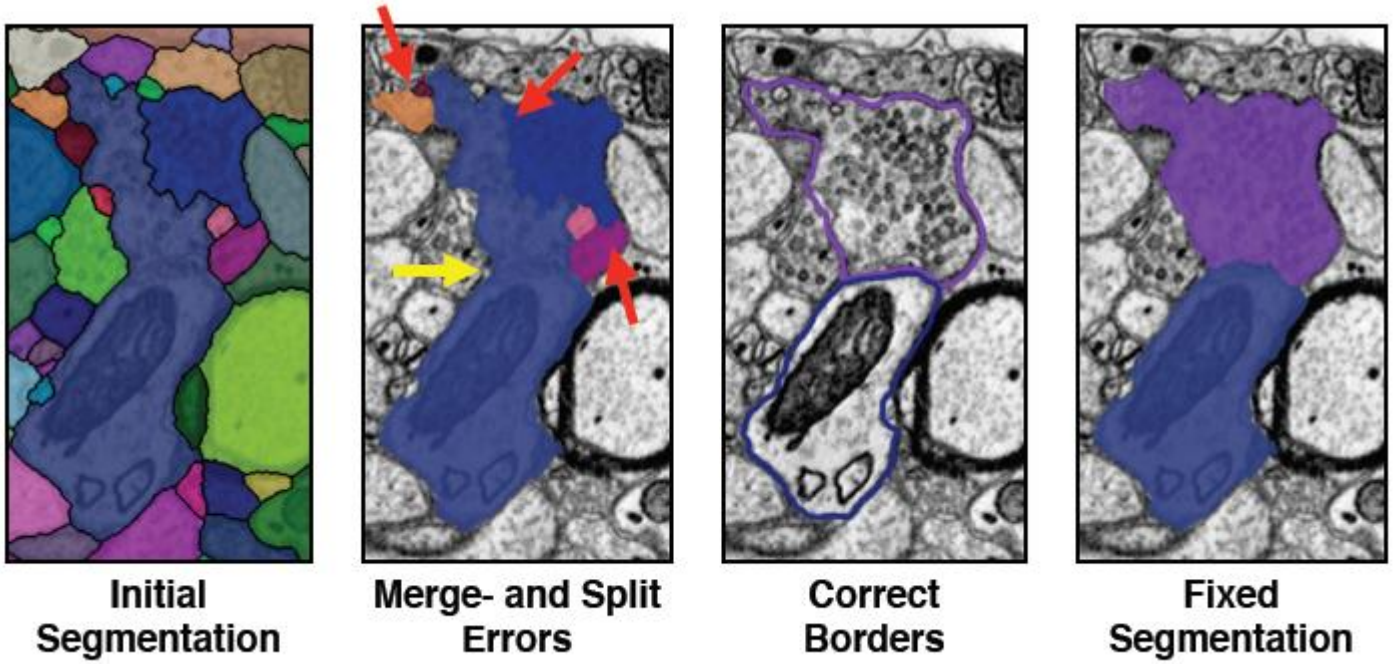

End


All on
All off
1
2
3
4
5
6
7
8
9
4s
17/10/2012
17/11/2012
17/12/2012

(a) Supervoxel Segmentation

(b) User Interactions

(c) Multi-label Segmentation



Instructor: James Tompkin

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

Max Planck Institute
Germany

University College London
UK

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

Watch my research overview video!

james_tompkin@brown.edu
Office: CIT 547