

Introduction and Overview

Computer Vision I

CSE 252A

Lecture 1

- We'll begin with some introductory material ...
- ... and end with
 - Syllabus
 - Organizational materials
 - Wait list

What is computer vision?



Add camera as input device to computer.

Done?

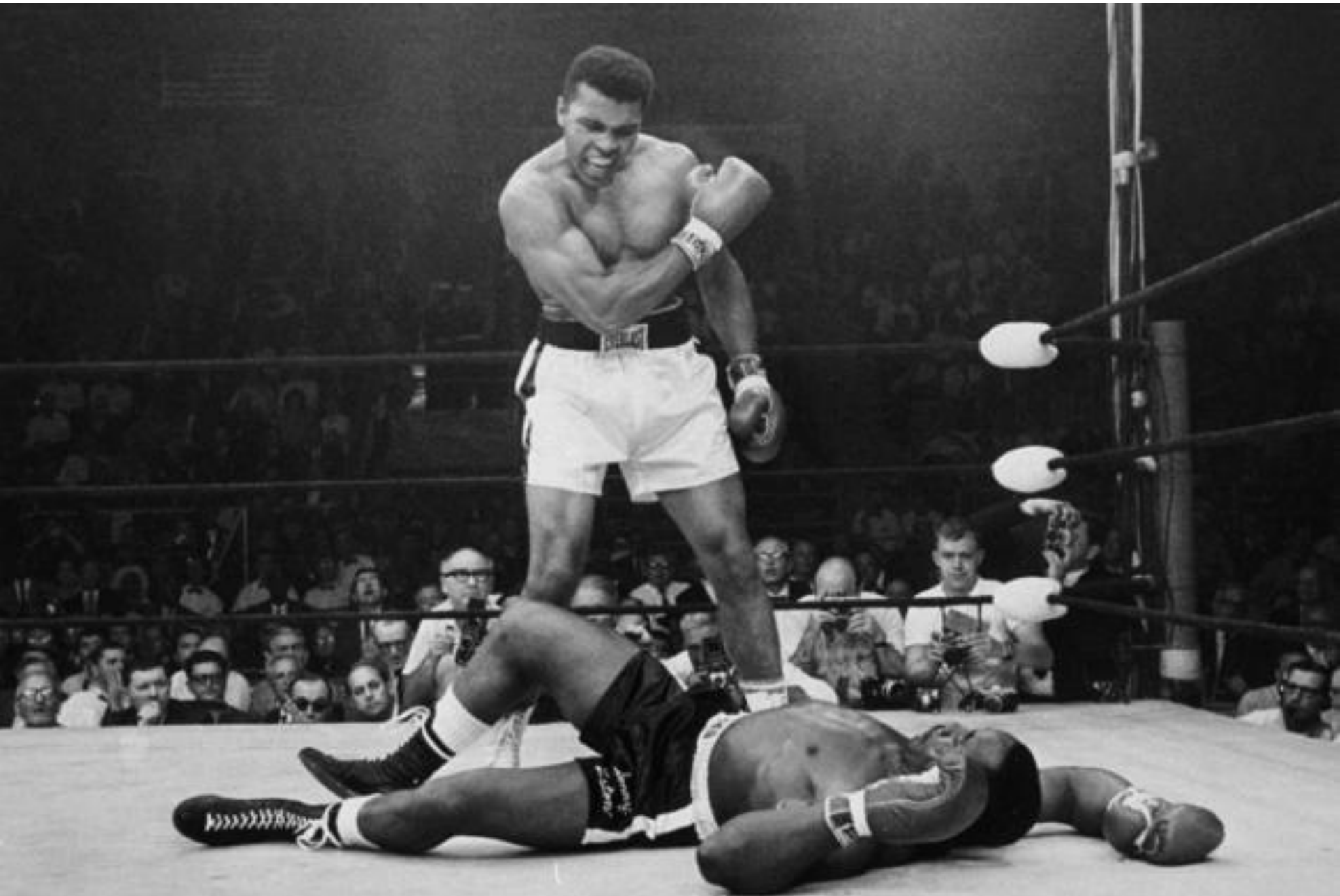
Computer Vision

- An interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos
- Other common definitions:
 - In computer vision, we are trying to...describe the world that we see in one or more images and to reconstruct its properties (Szeliski)
 - Computing properties of the 3-D world from one or more digital images (Trucco and Verri)
 - The construction of explicit, meaningful description of physical objects from images (Ballard and Brown)
 - Extracting descriptions of the world from pictures or sequences of pictures (Forsyth and Ponce)
 - To make useful decisions about real physical objects and scenes based on sensed images (Stockman and Shapiro)

Computer Vision

- An interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos
- Engineering perspective
 - Computer vision seeks to automate tasks that the human visual system can do

Caption this image



Why is this hard?



What is in this image?

1. A hand holding a man?
2. A hand holding a mirrored sphere?
3. An Escher drawing?
4. A 1935 self portrait of Escher

- Interpretations are ambiguous
- The forward problem (computer graphics) is well-posed
- The “inverse problem” (computer vision) is not

Underestimates

- “640K ought to be enough for anybody.”
 - Bill Gates, 1981
- “... in three to eight years we will have a machine with the general intelligence of an average human being ... The machine will begin to educate itself with fantastic speed. In a few months it will be at genius level and a few months after that its powers will be incalculable ...”
 - Marvin Minsky, LIFE Magazine, 1970



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

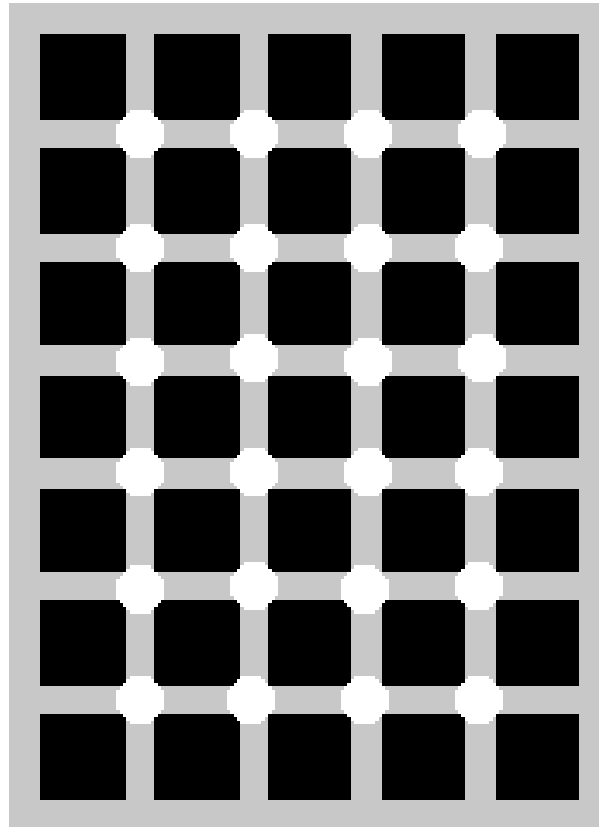
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Should Computer Vision follow from our understanding of Human Vision?

Yes & No

1. Who would ever be crazy enough to even try creating machine vision?
 2. Human vision “works”, and copying is easier than creating.
 3. Secondary benefit – in trying to mimic human vision, we learn about it.
-
1. Why limit oneself to human vision when there is even greater diversity in biological vision
 2. Why limit oneself to biological vision when there may be greater diversity in sensing mechanism?
 3. Biological vision systems evolved to provide functions for “specific” tasks and “specific” environments. These may differ for machine systems
 4. Implementation – hardware is different, and synthetic vision systems may use different techniques/methodologies that are more appropriate to computational mechanisms

Hermann Grid



Scan your eyes over the figure. Do you see the gray spots at the intersections? Stare at one of them and it will disappear.

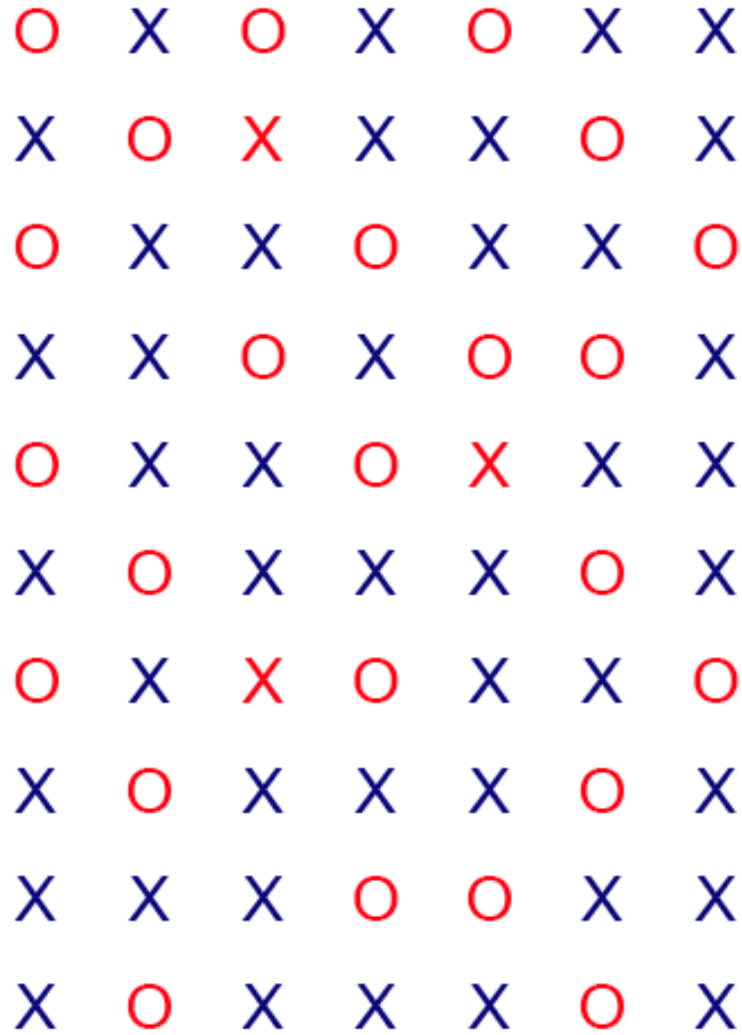
How many red X's are there?

Raise your hand when you know
the answer



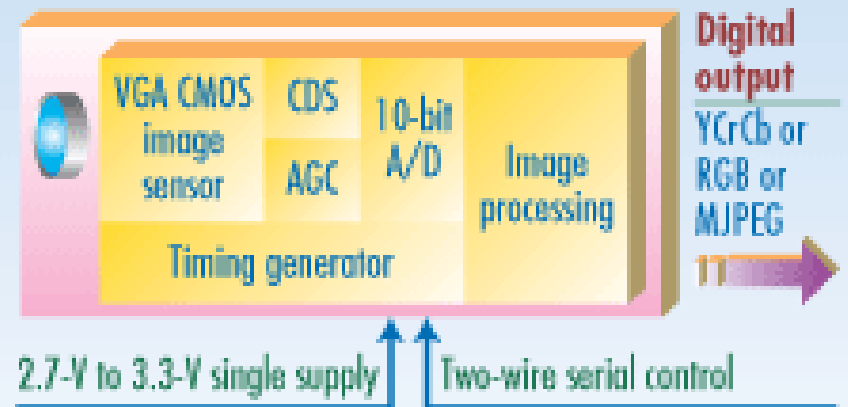
How many red X's are there?

Raise your hand when you know
the answer



The Near Future: Ubiquitous Vision

- Digital video has become very inexpensive.
- It's widely embedded in cell phones, cars, games, etc.
- 99.9% of digitized video isn't seen by a person.
- That doesn't mean that only 0.1% is important!
- And there's an enormous amount of image and video content on the internet...



CMOS module operates at 30 frames/s.

SOURCE: PHOTON

Applications: touching your life

- Optical Character Recognition
- Football
- Movies
- Surveillance
- HCI – hand gestures
- Aids to the blind
- Face recognition & biometrics
- Road monitoring
- Industrial inspection
- Virtual Earth; street view
- Robotic control
- Autonomous driving
- Space: planetary exploration, docking
- Medicine – pathology, surgery, diagnosis
- Microscopy
- Military
- Remote Sensing
- Digital photography
- Google Goggles
- Video games

Vision to explore the world

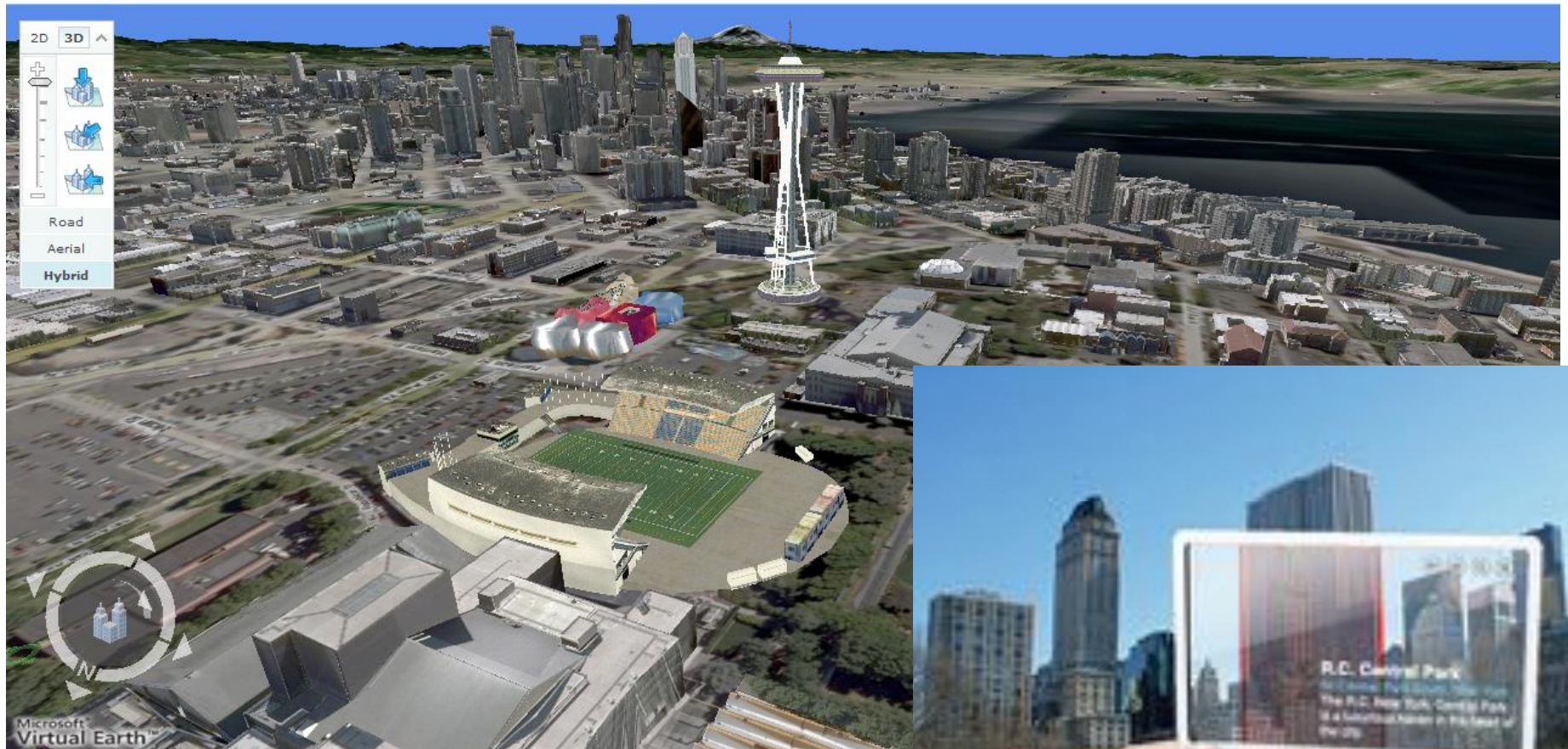


Image from Microsoft Virtual Earth
(see also: Google Earth)

Vision to explore other worlds



Image from NASA's Mars Exploration Rover Spirit

- Panorama stitching
- 3D terrain modeling
- Obstacle detection
- Position tracking

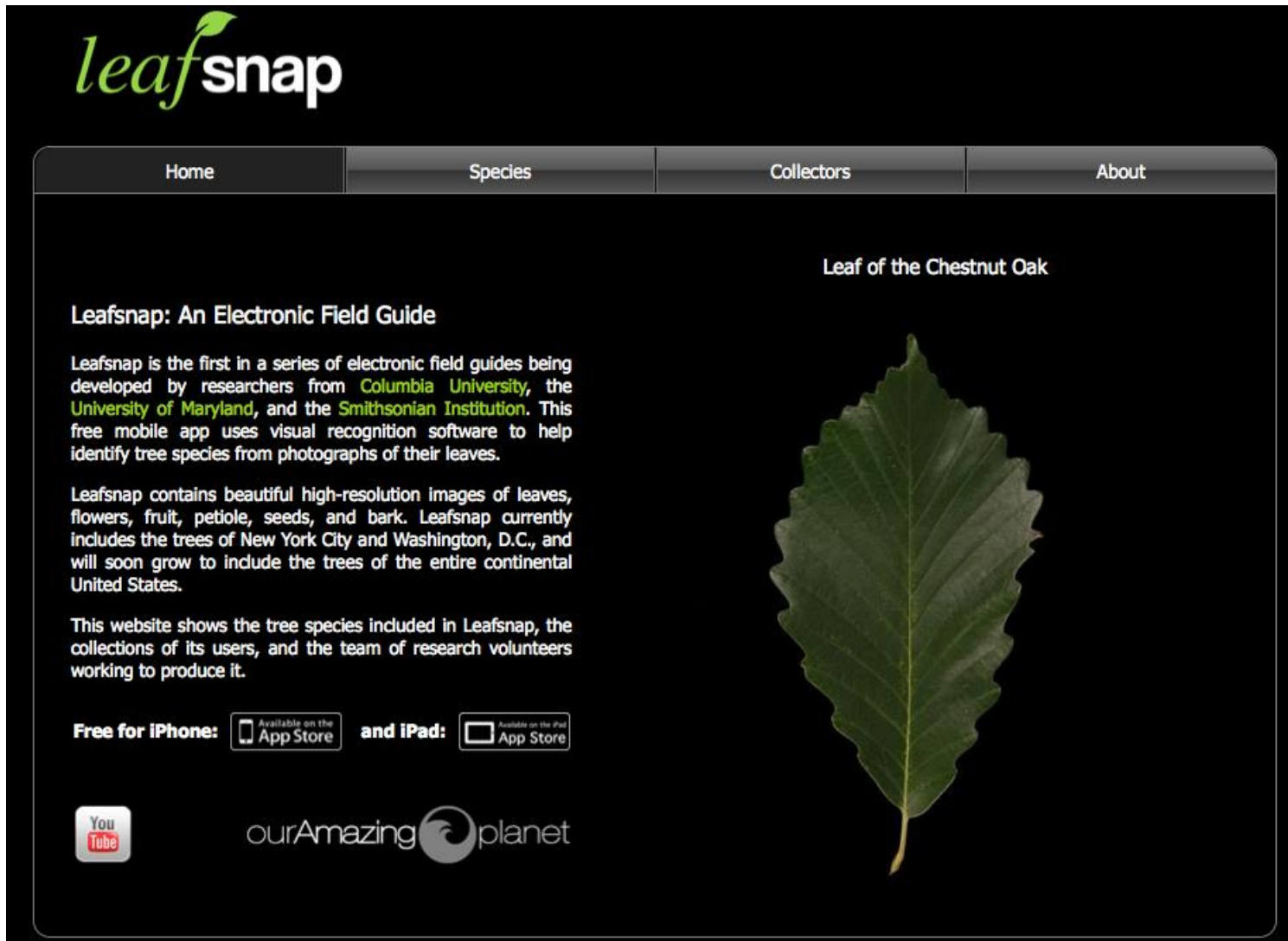
- For more, read “Computer Vision on Mars“ by Matthies et al.

Vision to recognize objects



- Point & Find, Nokia
- SnapTell.com (now Amazon)
- Google Photos
- Apple Photos
- Google Image Search

Visual Field Guides



leafsnap

Home Species Collectors About



Leaf of the Chestnut Oak



Leafsnap: An Electronic Field Guide


Leafsnap is the first in a series of electronic field guides being developed by researchers from [Columbia University](#), the [University of Maryland](#), and the [Smithsonian Institution](#). This free mobile app uses visual recognition software to help identify tree species from photographs of their leaves.

Leafsnap contains beautiful high-resolution images of leaves, flowers, fruit, petiole, seeds, and bark. Leafsnap currently includes the trees of New York City and Washington, D.C., and will soon grow to include the trees of the entire continental United States.

This website shows the tree species included in Leafsnap, the collections of its users, and the team of research volunteers working to produce it.

Free for iPhone:  and iPad: 



Bird Wheel

Bird List

Bird Lab

About

Sort by **Tree of Life**

Alphabetical

Visual Recognition

Text Search

Order: **Falconiformes**
Family: **Falconidae**
Subfamily: **Caracarinae**
Genus: **Caracara**
Species: **C. cheriway**

by Andrew Spencer

- ✦ - visually similar
- ✦ - arriving
- ✦ - departing
- ✦ - migrating through



Crested Caracara

Some recordings include other species

[View original image](#)

Object recognition (in supermarkets)



LaneHawk by EvolutionRobotics (now part of iRobot)

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk, you are assured to get paid for it... “

Amazon Go



1. Turn-style entry. Consumer scans in with Amazon App on smartphone
2. Consumer goes around the store, picks up items, adds to bag, shops like normal
3. Consumer exits

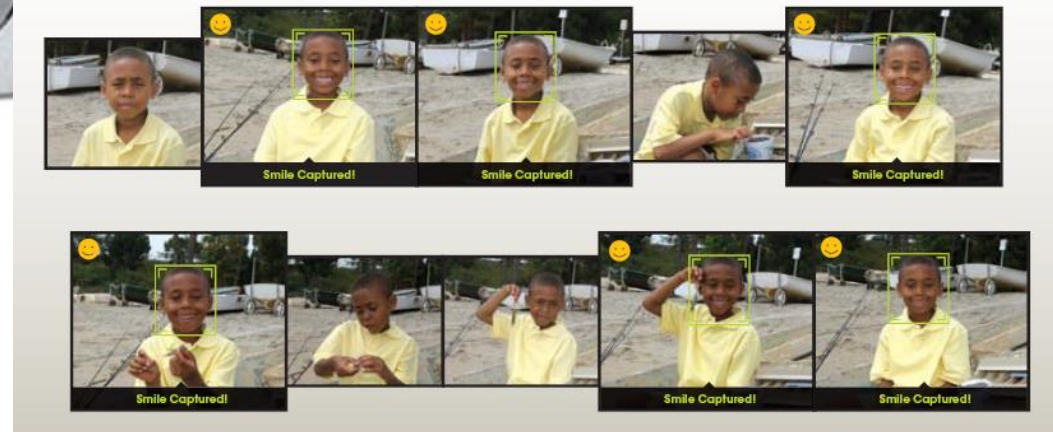


Vision to look at people



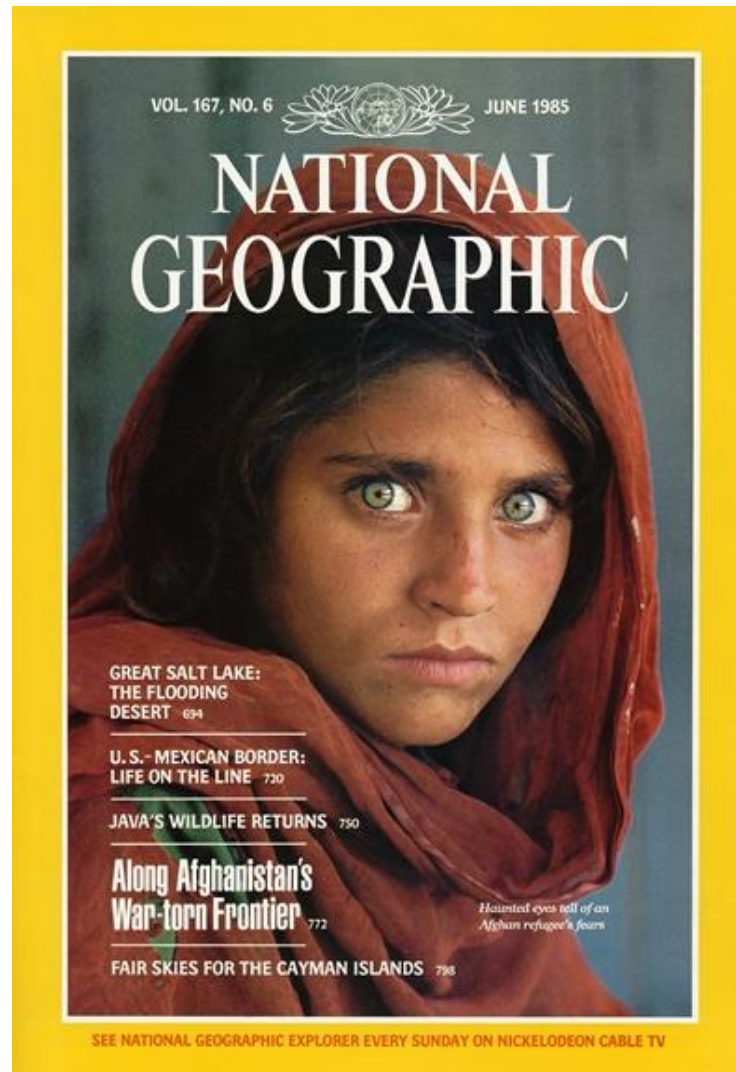
Face Detection

Smile Detection



- Digital cameras, phones, Facebook, Google Photos, Snapchat etc.

Face recognition



Who is she?

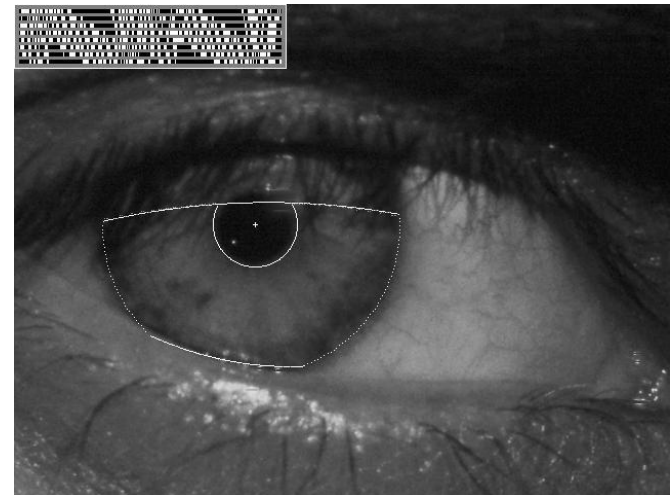
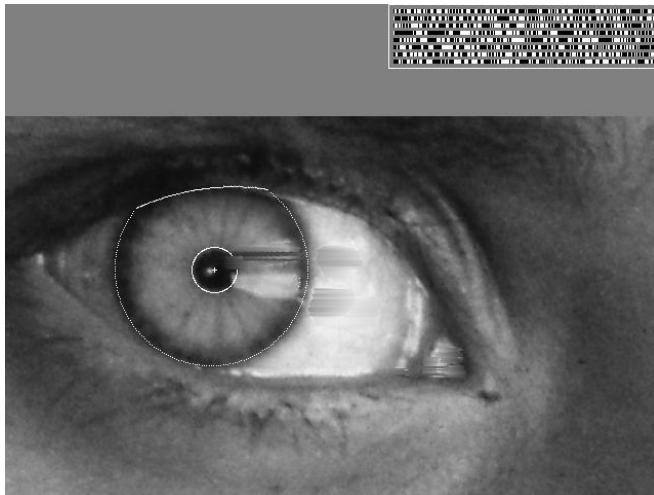
Vision-based biometrics

1984
Age 12



2002
Age 30

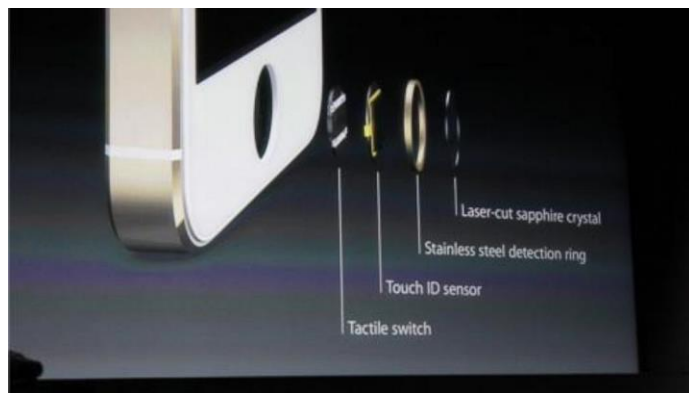
“How the Afghan Girl was Identified by Her Iris Patterns” Read the [story](#)



Login without a password...



iPhone X



Fingerprint scanners on smartphones, laptops, mice, other devices



As devastating diseases wipe out China's pigs, companies scramble to roll out facial and voice recognition to save them.



通过在猪栏安装摄像头，对群猪画面进行猪脸捕捉识别
capturing and recognizing on the pig face image, moreover

A system made by Yingzi Technology, a small Chinese company, scanning a barn to recognize pig faces. Yingzi uses video to capture faces because the pigs move a lot. "You can't take a single picture of a pig," the firm's chief executive said.

High-Tech Solutions Down on the Farm

By SUI-LEE WEE and ELSIE CHEN

A database of every pig's face. Voice scans that detect hogs with a cough. Robots that dispense just the right amount of feed. This could be China's pig farm of the future.

Chinese companies are pushing facial and voice recognition and other advanced technologies as ways to protect the country's pigs. In this Year of the Pig, many Chinese hogs are dying from a deadly swine disease, threatening the country's supply of pork, a staple of Chinese dinner tables.

So China's ebullient technology sector is applying the same techniques it has used to transform Chinese life — and, more darkly, that the Chinese government increasingly uses to spy on its own people — to make sure its pigs are in the pink of health.

"If they are not happy, and not eating well, in some cases you can predict whether the pig is sick," said Jackson He, chief executive officer of Yingzi Technology, a small firm based in the southern city of Guangzhou that has introduced its vision of a "future pig farm" with facial and voice recognition technologies.



China's biggest tech firms want to pamper pigs, too. Alibaba, the e-commerce giant, and JD.com, its rival, are using cameras to track pigs' faces. Alibaba also uses voice-recognition software to monitor their coughs.

Many in China are quick to embrace high-tech solutions to just about any problem. A digital revolution has transformed China into a place where nearly anything — financial services, spicy takeout, manicures and dog grooming, to name a few — can be summoned with a smartphone. Facial recognition has been deployed in public bathrooms to dispense toilet paper, in train stations to apprehend criminals and in housing complexes to open doors.

This pig push, however, may be a step too soon.

"I like the idea, I like the concept, but I need to be shown that it works," said Dirk Pfeiffer, a professor of veterinary epidemiology at the City University of Hong Kong. "Because if it doesn't work, it's counterproductive."

CONTINUED ON PAGE B2

But all is not so bright

One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority

In a major ethical leap for the tech world, Chinese start-ups have built algorithms that the government uses to track members of a largely Muslim minority group.



Pastors are getting in on Ring's ever-growing surveillance state

Share on Facebook

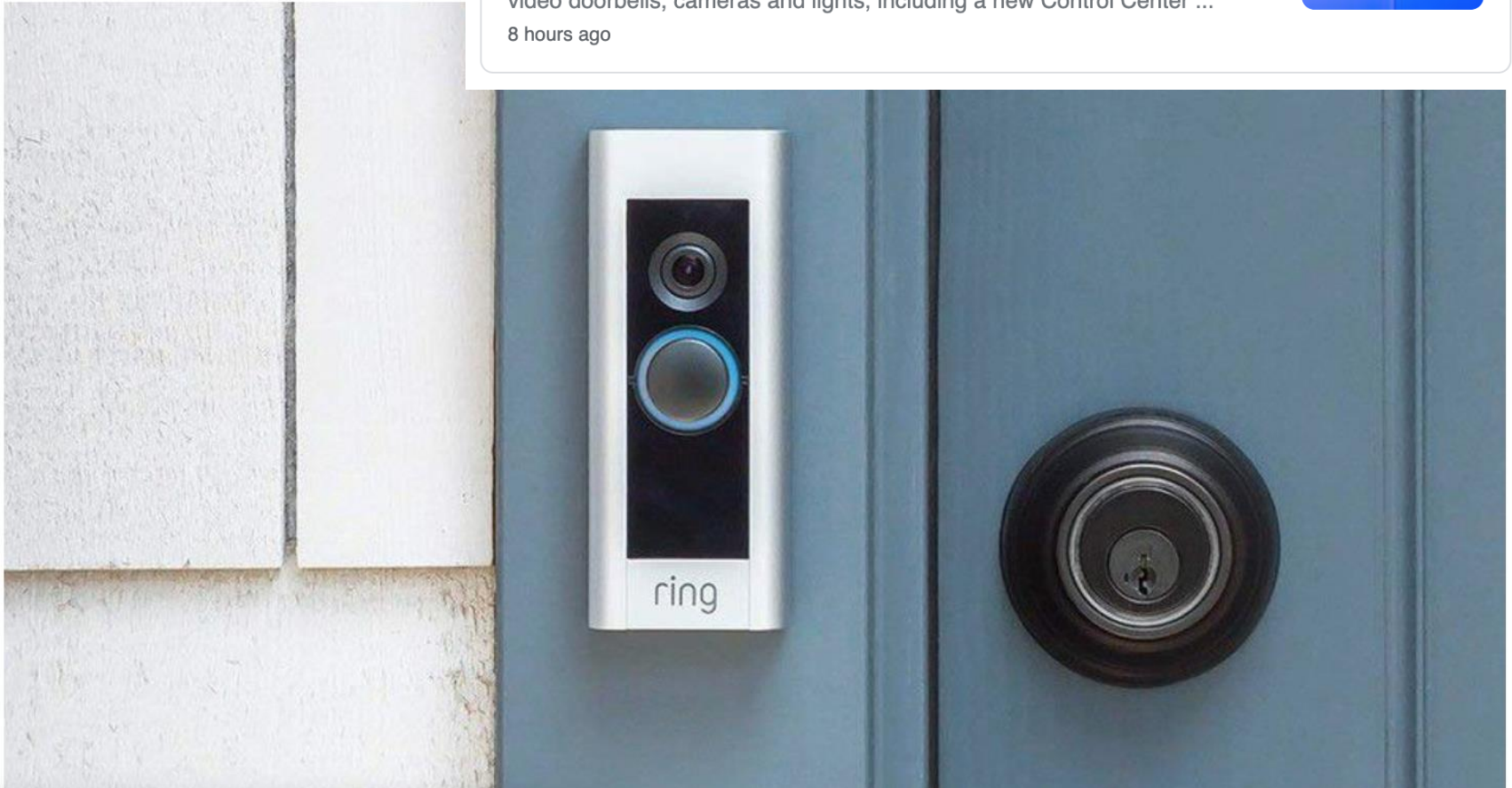
Share on Twitter

CNET

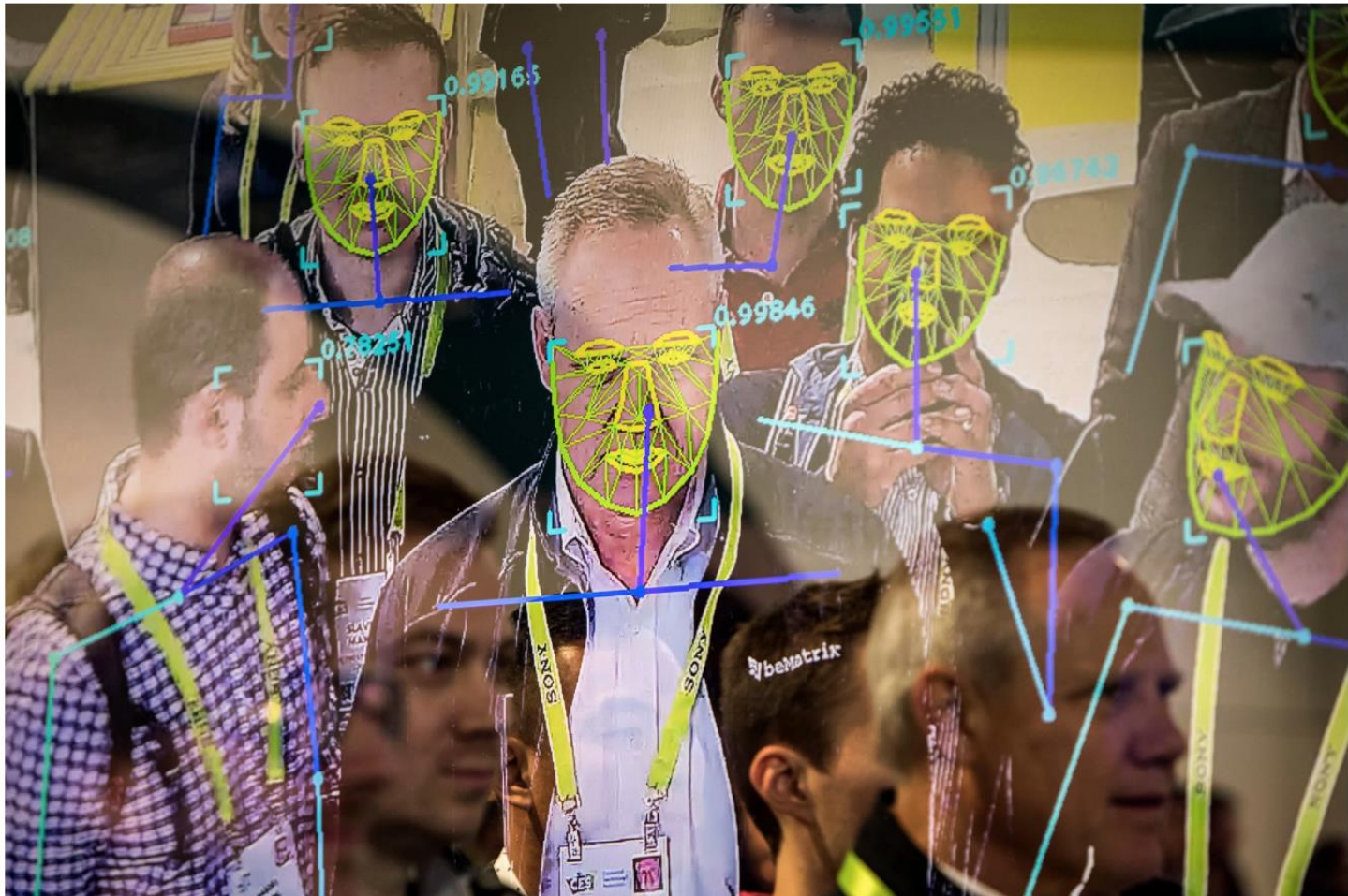
Ring CEO defends police partnerships as it rolls out new privacy tools

Ring CEO defends police partnerships as it rolls out new privacy tools ... its video doorbells, cameras and lights, including a new Control Center ...

8 hours ago



San Francisco Bans Facial Recognition Technology



With great power comes great
responsibility

Vision for entertainment: shape capture



The Matrix movies, ESC Entertainment, XYZRGB, NRC

Vision for entertainment: motion capture



Facial
motion
capture

Vision for entertainment

- Football first down line



Sportvision first down line

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

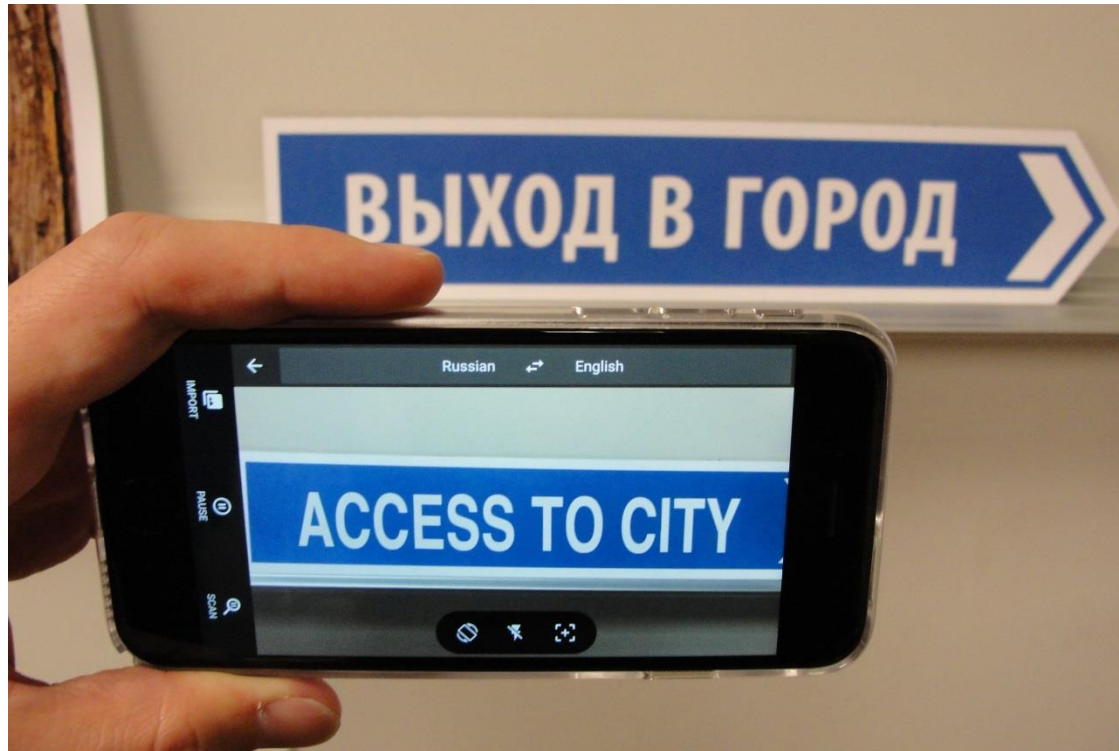
Vision for augmented reality

- AR Toolkit
- Blippar
- Magic Leap
- Microsoft HoloLens



Vision for augmented reality

- Text detection, localization, and translation, then render with similar font

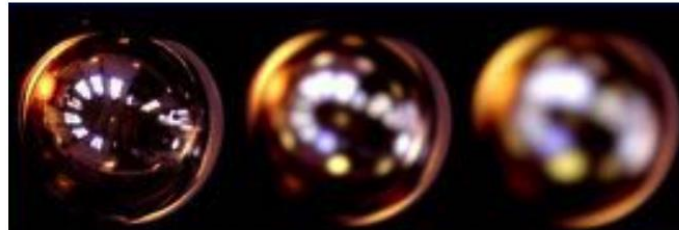


Vision for augmented reality

- Pokémon Go



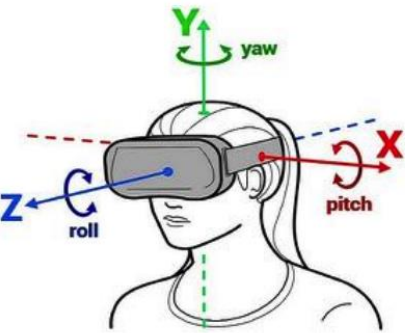
Vision in Immersive Augmented Reality



Material and lighting estimation



Gaze tracking



Head pose estimation



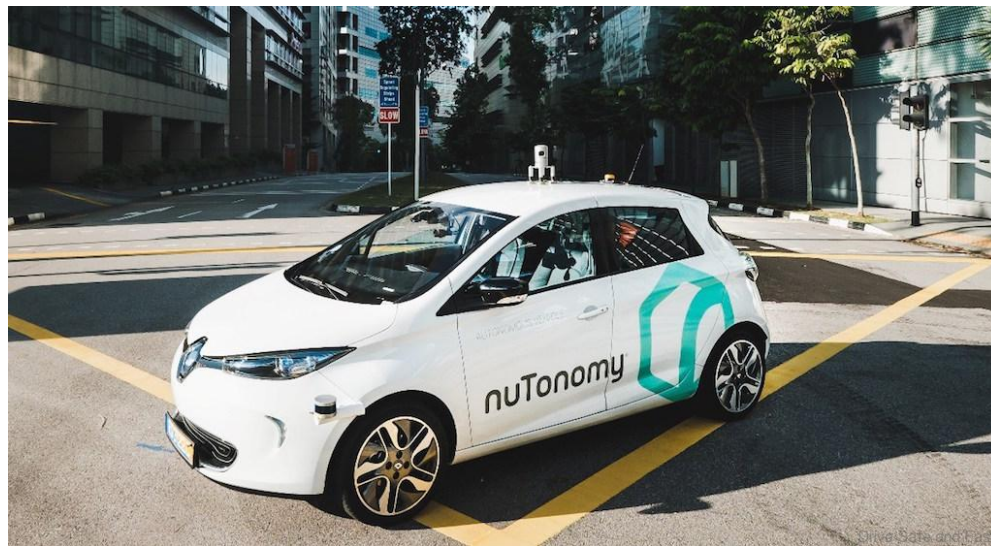
Depth estimation



Semantic segmentation



Vision for Autonomous Cars



Vision for smart cars

- The EU mandates from July 7, 2024, all vehicles must be equipped with intelligent-speed-assist (ISA) systems



Vision-based interaction (and games)



Nintendo Wii has camera-based IR tracking built in.



Digimask: put your face on a 3D avatar.



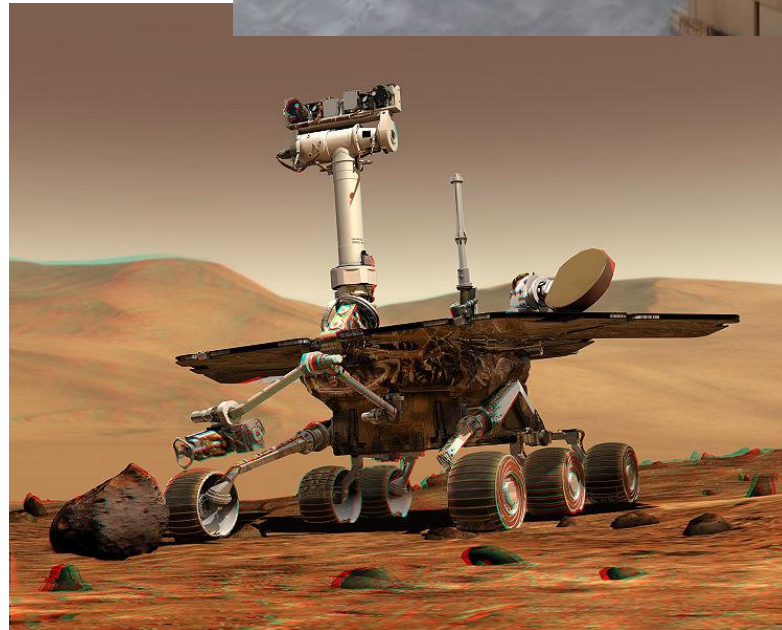
Playmotion game a Disney Epcot



Xbox
Kinect

Computer Vision I

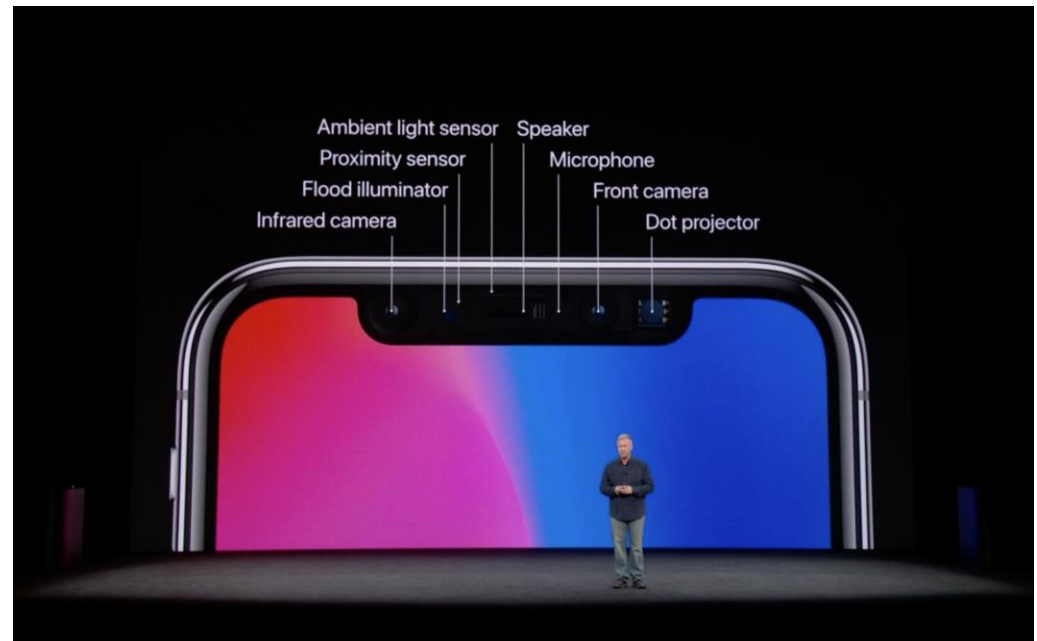
Vision for robotics



Vision for robotics



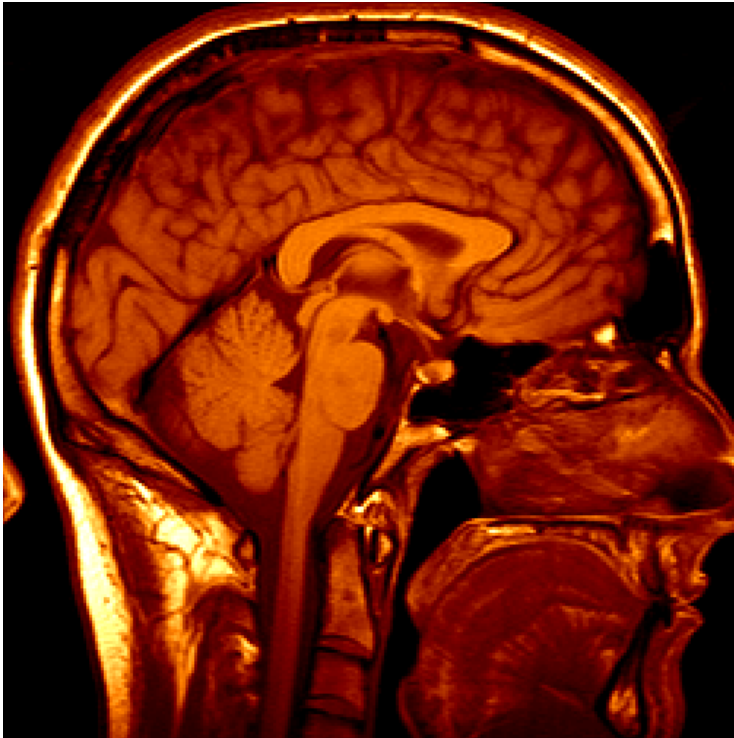
3D sensors



Vision from your perspective



Vision for medicine

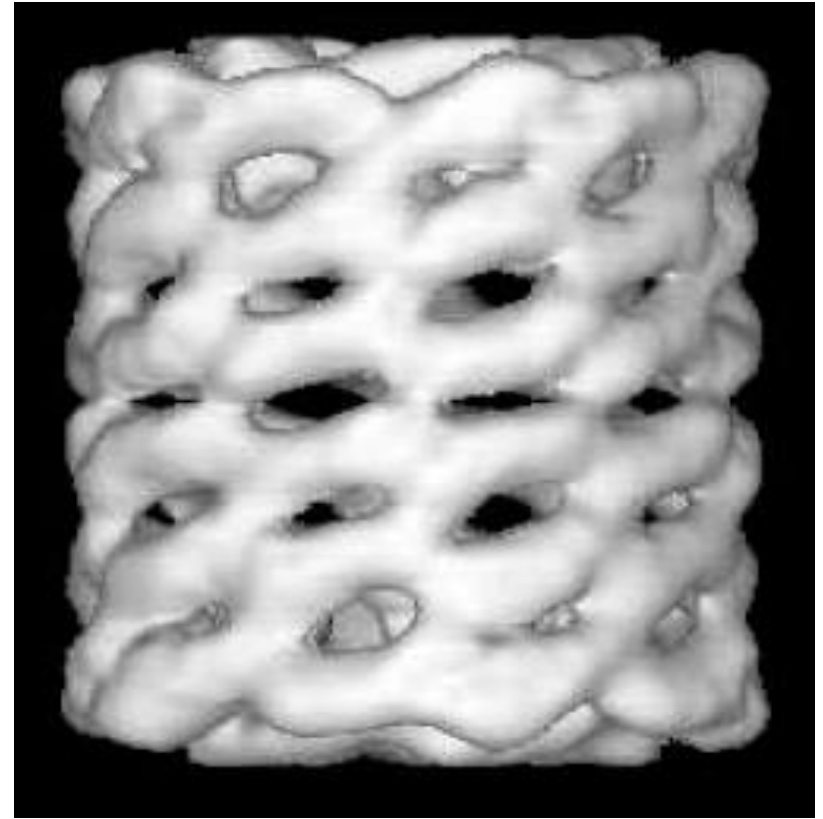
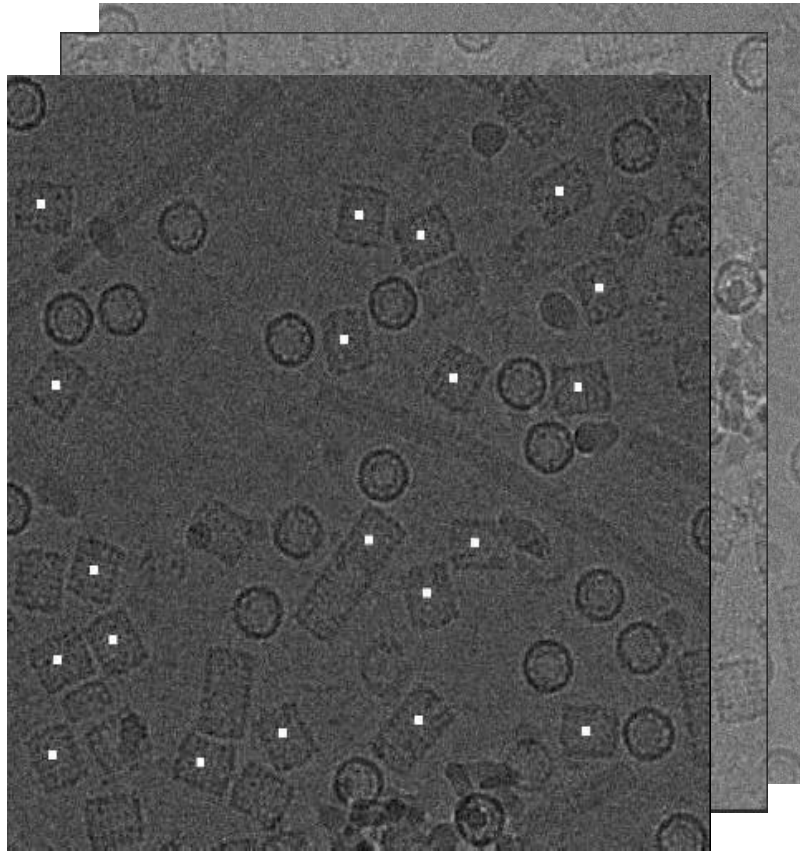


3D imaging
MRI, CT



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

Vision for Science

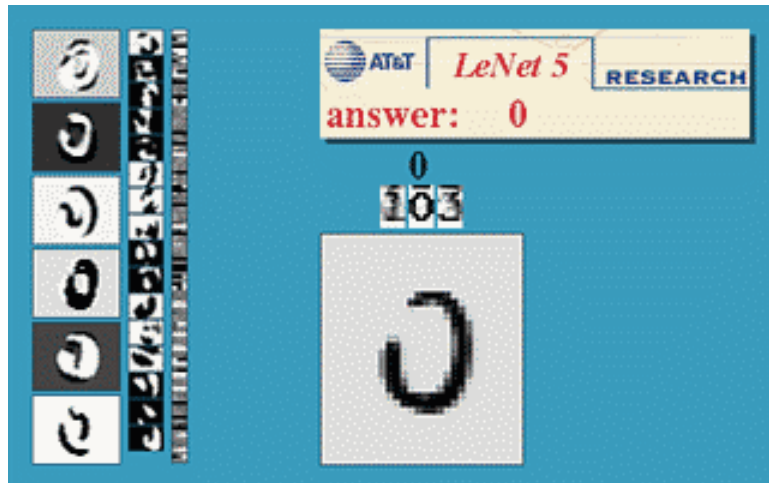


Molecular Reconstruction from
Cryo-electron Microscope Images

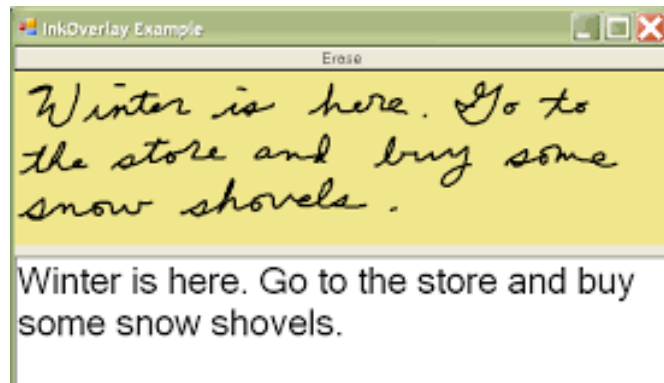
[Mallick, Zhu, Kriegman]

Computer Vision I

Vision to bridge communication between physical and digital worlds: Optical Character Recognition (OCR)



Digit recognition, AT&T labs



Handwriting recognition



License plate readers

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

Scene Text: Text Recognition in the Wild

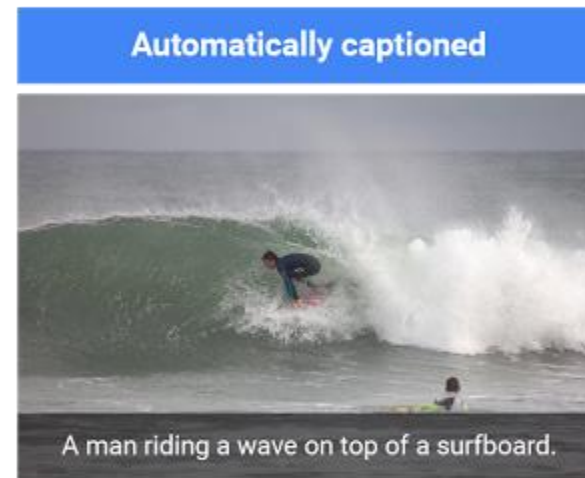


COCO-Text

A large-Scale Scene Text Dataset
<https://bgshih.github.io/cocotext/>



Automatic image captioning



Automatic image captioning



Video understanding

- Video classification
- Activity recognition
- Video segmentation
- Activity detection



Animal (97.76%)
Wildlife (92.16%)
Tiger (90.11%)
Terrestrial animal (68.17%)
Bengal tiger (64.77%)
Whiskers (63.30%)
Zoo (58.16%)
Roaring cats (56.41%)
Cat (44.12%)

▶ 0:29 / 1:38



Video url: https://storage.googleapis.com/cloudmleap/video/next/GooglePhotos_Animals.mp4

Computer Vision

- An interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos

Related Fields

Robotic Vision

Multi-variable SP

Non-linear SP

Computer Intelligence

Control Robotics

Artificial Intelligence

Signal Processing

Cognitive Vision

Machine Learning

Computer Vision

Machine Vision

Physics

Optics

Statistics

Mathematics

Image Processing

Imaging

Geometry

Optimization

Neurobiology

Smart Cameras

Computer Graphics

Computational Photography

Biological Vision

Related Fields

Robotic Vision

Multi-variable SP

Computer Intelligence

Control Robotics

Non-linear SP

Artificial Intelligence

Signal Processing

Cognitive Vision

Machine Learning

Deep Learning

Physics

Optics

Statistics

Mathematics

Image Processing

Imaging

Geometry

Optimization

Neurobiology

Smart Cameras

Computer Graphics

Computational Photography

Biological Vision

Deep Learning in Computer Vision

- First, understand the physics of the problem, then be wise about whether to use deep learning
 - If not, then you may “learn the bitter lesson that building in how we think we think does not work in the long run” and “general methods that leverage computation are ultimately the most effective, and by a large margin”
 - Richard Sutton, “The Bitter Lesson”
 - On the other hand, you may learn “we have to take into account the total cost of any solution, and that so far they have all required substantial amounts of human ingenuity”
 - Rodney Brooks, “The Better Lesson”

Four Rs of computer vision

- Reprojection
 - Rendering a scene from a different view, under different illumination, under different surface properties, etc.
- Reconstruction
 - Multiple view geometry, structure from motion, shape from X (where X is texture, shading, contour, etc.), etc.
- Registration
 - Tracking, alignment, optical flow, correspondence, etc.
- Recognition
 - Recognizing objects, scenes, events, etc.

Others may have slightly different Rs

Rudiments: The implied fifth R

- Image filtering
- Edge detection
- Interest point detection
- Probability
- Statistics
- Linear algebra
- Projective geometry
- Optics
- Fourier analysis
- Sampling
- Algorithms
- Photometry
- Physics of color
- Human vision
- Psychophysics
- Performance evaluation

CSE 252A topics

- Geometric image formation
- Photometric image formation
- Photometric stereo
- Image filtering
- Edge detection and corner detection
- Calibrated stereo
- Feature matching
- Uncalibrated stereo
- Feature extraction
- Structure from motion
- Robust model fitting
- Optical flow and motion
- Recognition, detection, and classification
- Neural networks
- Convolutional neural networks
- Color
- Perception

Syllabus

- Instructor: Ben Ochoa
- TAs: Karan Santhosh, Rishikanth Chandrasekaran, and Mehmet Simsek
- Public course website
 - <https://cseweb.ucsd.edu/classes/fa23/cse252A-a/>
- Course is on Canvas
 - Piazza for discussion
 - Gradescope for submitting assignments
- 19 lecture meetings
 - No lecture on Thanksgiving Eve (Wed, Nov 22)

Syllabus

- Grading

- 5 homework assignments (60% of grade)

- By hand (typed) and programming using Python

- Late policy: 15% grade reduction for each 12 hours late

- Will not be accepted 72 hours after the due date

- 4% for assignment 0; 14% for assignments 1-4

- Final exam (40% of grade)

- Piazza

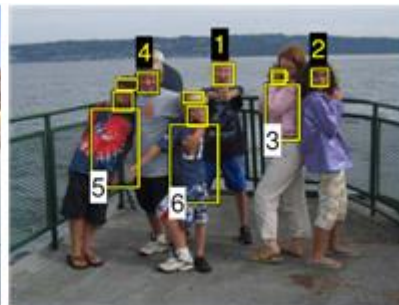
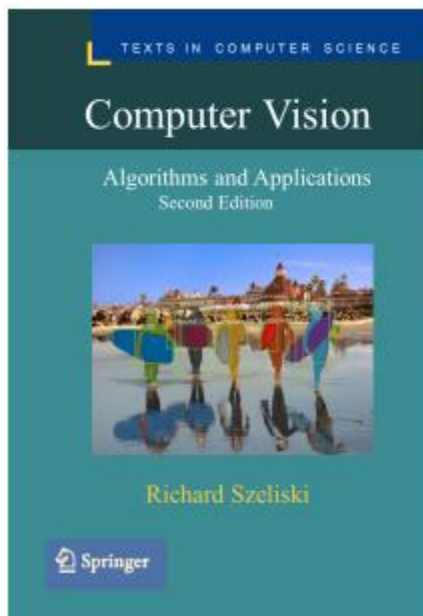
- Ask (and answer) questions using Piazza, not email

- Extensive, nontrivial participation could raise your grade (e.g., raise a B+ to an A-)



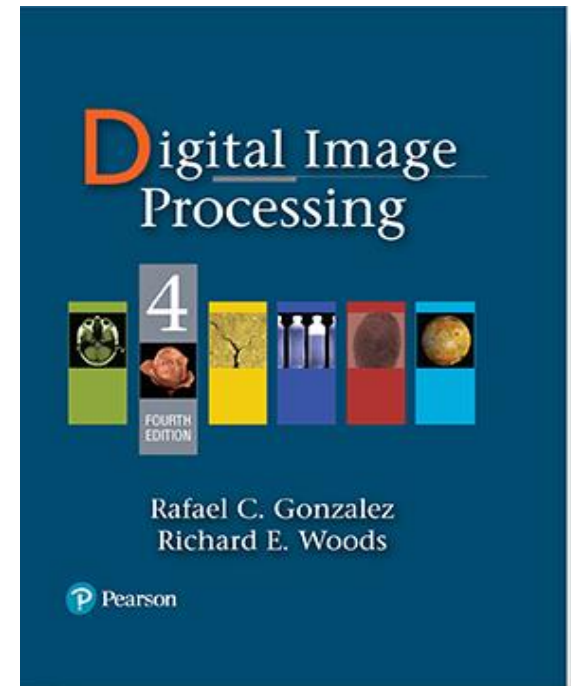
Textbook (optional)

- Computer Vision: Algorithms and Applications, 2nd edition
 - Richard Szeliski
 - <http://szeliski.org/book>



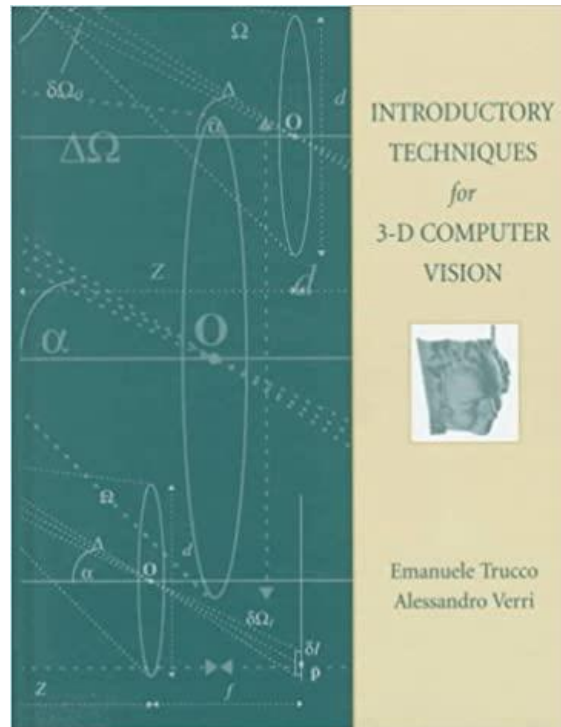
Textbook (optional)

- Digital Image Processing, 4th edition
 - Rafael C. Gonzalez and Richard E. Woods
- See book website
 - Corrections and clarifications
 - Review material
 - Linear systems
 - Matrices and vectors
 - Probability



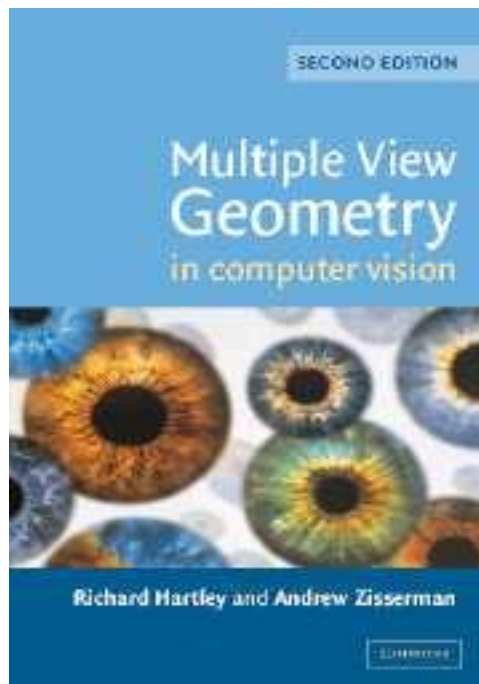
Textbook (optional)

- Introductory Techniques for 3-D Computer Vision
 - Emanuele Trucco and Alessandro Verri



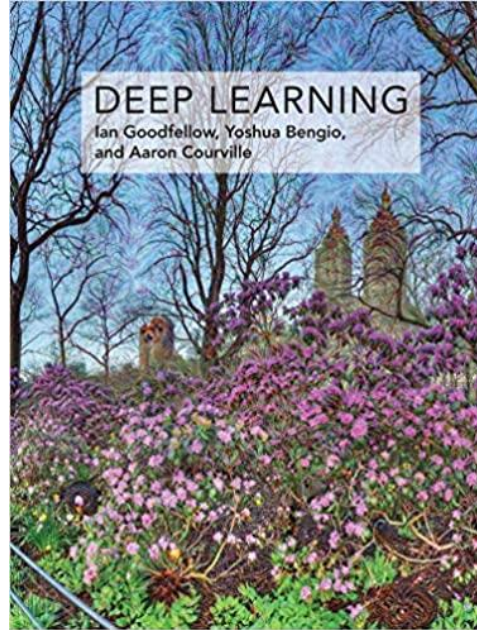
Textbook (optional)

- Multiple View Geometry in Computer Vision, 2nd edition
 - Richard Hartley and Andrew Zisserman
- Download the corrections and errata



Textbook (optional)

- Deep Learning
 - Ian Goodfellow, Yoshua Bengio, and Aaron Courville
 - <https://www.deeplearningbook.org/>



Collaboration Policy

- Ask and answer questions on Piazza, not email
- Post **publicly** (optionally anonymously)
 - Conceptual questions and high-level questions about assignments
- All other posts must be **private** to “Instructors” (includes instructor and instructional assistants)
 - Low-level, detailed assignment questions (e.g., implementation details)
 - Assignment-specific code
 - Results (intermediate or final; e.g., numerical values, images, figures)
 - **Posting such items publicly is an academic integrity violation**
- If you are unsure, then post privately to “Instructors”
 - If suitable, then it will be changed to a public post
- Piazza is the official, University-sanctioned discussion forum
 - **Do not use Piazza to solicit others to an alternative forum**

Collaboration Policy

It is expected that you complete your academic assignments on your own and in your own words and code. The assignments have been developed by the instructor to facilitate your learning and to provide a method for fairly evaluating your knowledge and abilities (not the knowledge and abilities of others). So, to facilitate learning, you are authorized to discuss assignments with others; however, to ensure fair evaluations, you are not authorized to use the answers developed by another, copy the work completed by others in the past or present, or write your academic assignments in collaboration with another person.

Academic Integrity Policy

Integrity of scholarship is essential for an academic community. The University expects that both faculty and students will honor this principle and in so doing protect the validity of University intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind.

Academic Integrity Violation

If the work you submit is determined to be other than your own, you will be reported to the Academic Integrity Office for violating UCSD's Policy on Integrity of Scholarship. In accordance with the CSE department academic integrity guidelines, *students found committing an academic integrity violation will receive an F in the course.*

Student Conduct Policy

Maintaining an academic community free from disruption/obstruction, physical abuse, harassment, and other conduct inconsistent with UCSD's Principles of Community is each of our responsibility and essential to promoting mutual respect and academic rigor. Non-academic student misconduct will be reported to the Center for Student Accountability, Growth, and Education for violating the University Standards of Conduct.

Wait List

- Number of enrolled students is limited by
 - Size of room
 - Number of TAs
- General advice
 - Wait for as long as you can
- Concurrent enrollment (Extension) students have lowest priority

- And, if you are going to drop the class, please officially drop it to make room for others

Certification of Commencement of Academic Activity

- Every course at UC San Diego, per the US Department of Education, is required to certify whether students have commenced academic activity for a class to be counted towards eligibility for Title IV federal financial aid. This certification must be completed during the first two weeks of instruction.
- For CSE 252A, this requirement will be fulfilled via an ungraded prior knowledge quiz, which will assist the instructional team by providing information about your background coming into the course
 - In Canvas (<https://canvas.ucsd.edu>), go to the CSE 252A course and navigate to Quizzes, then click on **First Day Survey: Prior Knowledge #FinAid**

Next Lecture

- Geometric image formation