Computer Vision (CS 419/619)
Course Introduction and Motivation

Dr. Puneet Gupta

Logistics

Instructor: Puneet Gupta
E-mail: puneet@iiti.ac.in, office: Room 411, POD-1A (Scandium Building)

Prefix email subject by CS619

Distribution:

- ✓ Midsem Exam: 30%
- ✓ Endsem Exam: 40%
- ✓ Term Project + Paper presentation+ quiz: 30%

Programming language for project: Python

LaTeX should be used for preparing reports.

Exams will be closed-book.

Attendance

Logistics

Project

- Form groups of size 3 or 4 students.
- A list of project ideas will be provided
- Can propose and work on your own project idea after discussing with me
- Better to perform all the action before the deadline to avoid last minute problems.

Similar action for paper presentation

Reading material from the relevant sources will be provided.

Kindly beware of the different notations.

Possible to create tutorials sessions into regular sessions.

Cheating and plagiarism without proper credit to the original source(s) will lead to strict punishments.

What is Computer Vision?

- Make computers understand images and videos.

- What kind of scene?
- Where are the cars?
- How far is the building?

What is Computer Vision?

- Make computers understand images and videos.

- Where are they?
- What are they doing?
- Why is this happening?
- What is important?
- What will I see?

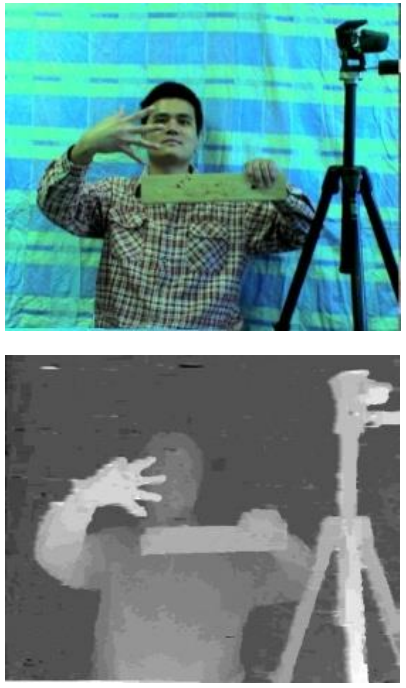
Computer Vision

- Automatic understanding of images and video
    1. Computing properties of the 3D world from visual data *(measurement)*

Slide credit: Kristen Grauman

1. Vision for measurement

Real-time stereo

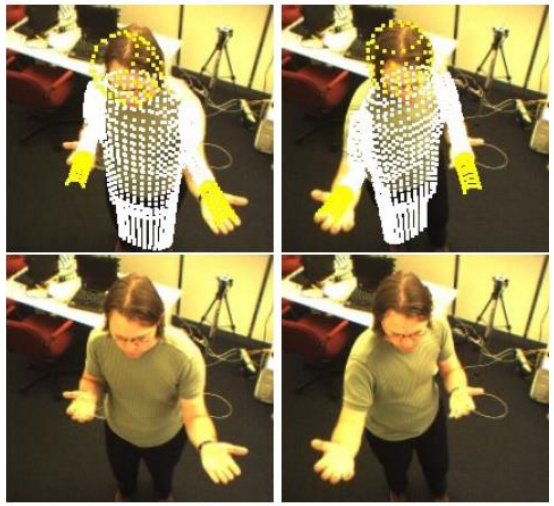


Wang et al.

Structure from motion

Snavely et al.

Tracking



Demirdjian et al.

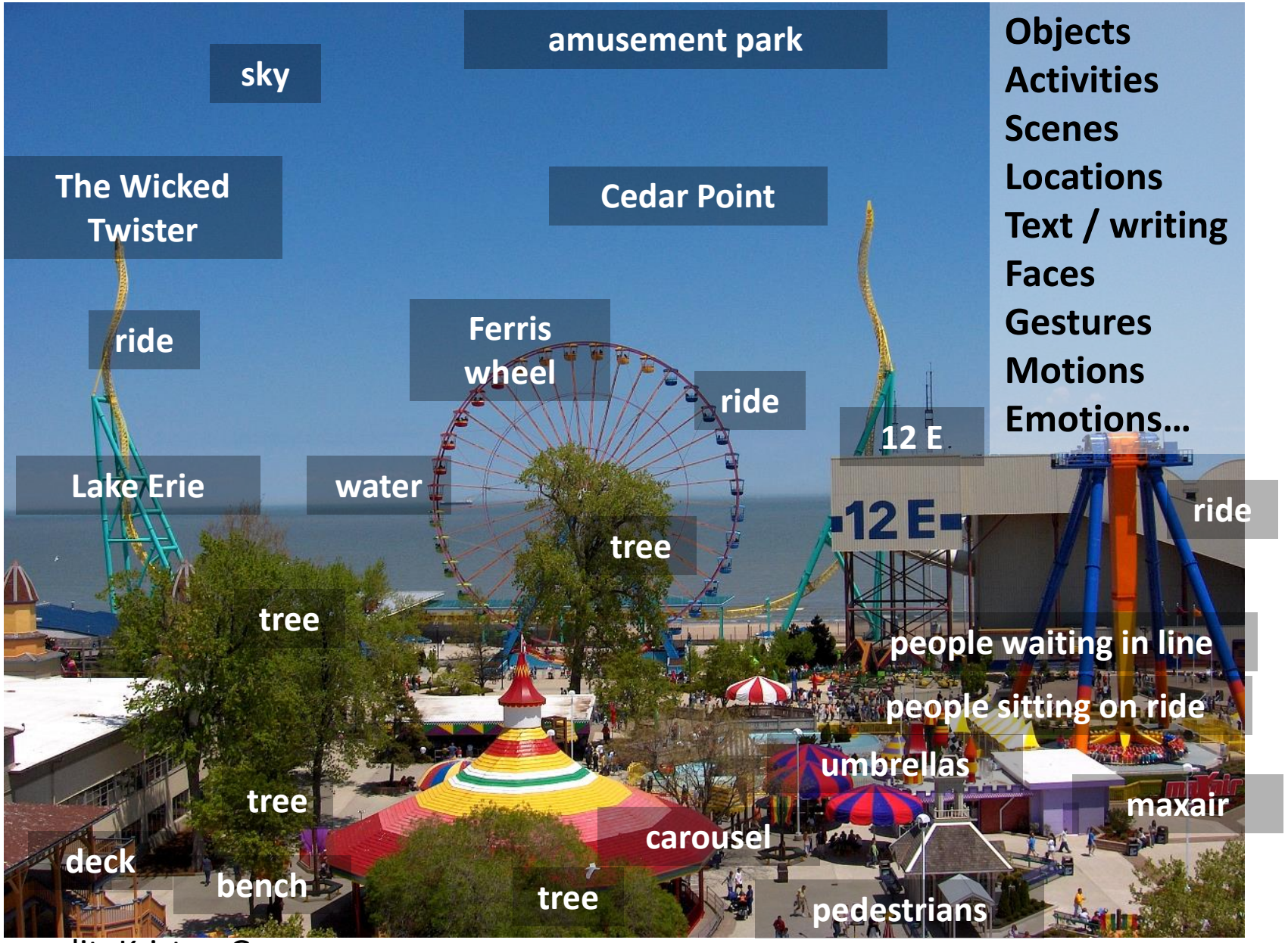
Slide credit: Kristen Grauman

Computer Vision

- Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data *(measurement)*
  2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. *(perception and interpretation)*

Slide credit: Kristen Grauman

2. Vision for perception, interpretation



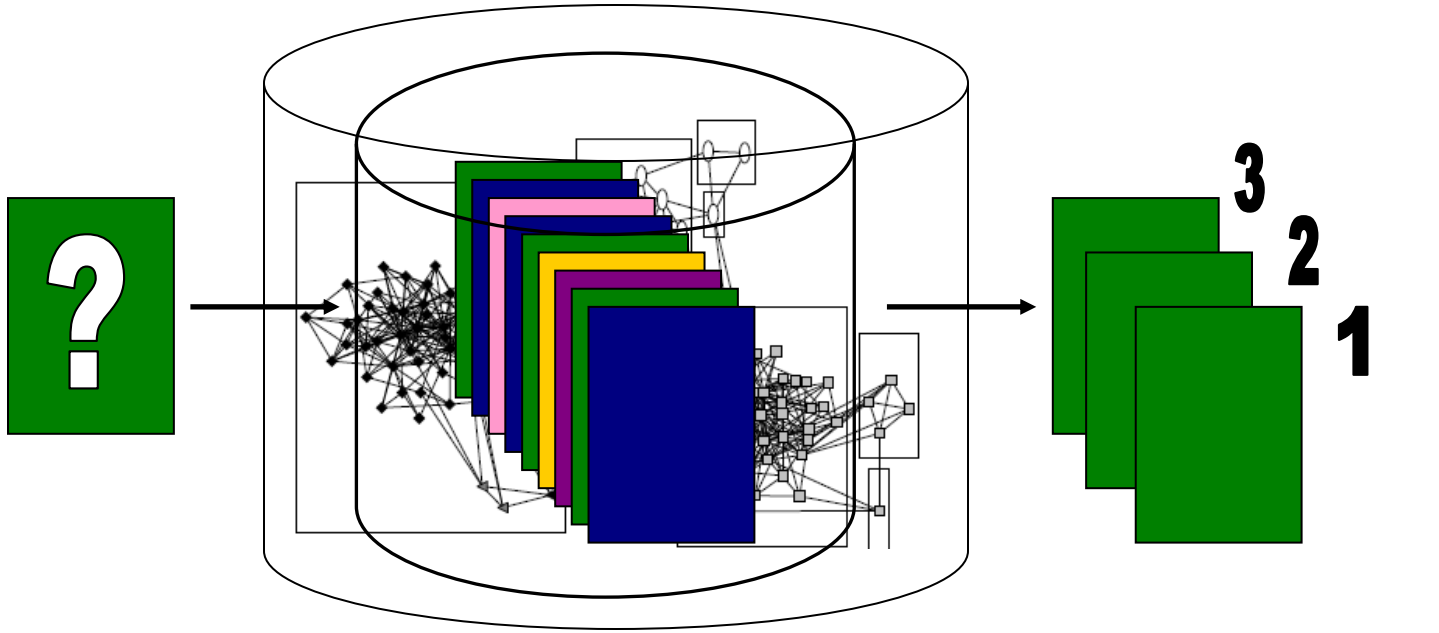
Slide credit: Kristen Grauman

Computer Vision

- Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data *(measurement)*
  2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. *(perception and interpretation)*
  3. Algorithms to mine, search, and interact with visual data (*search and organization*)

Slide credit: Kristen Grauman

3. Visual search, organization



Slide credit: Kristen Grauman

Computer vision

Jitendra Malik, UC Berkeley gave three ‘R’s of Computer Vision. “The classic problems of computational vision: reconstruction, recognition, and (re)organization.”

1. **Reconstruction** involves the **recovery of three-dimensional geometry** from images. More broadly, it can be interpreted as **"inverse graphics"**: estimating shape, spatial layout, reflectance, and illumination.
2. **Recognition** refers to the process of attaching *semantic category labels* to objects, scenes, events, and activities in images.
3. **(Re-)Organization** refers to the *grouping/segmentation* of visual elements. It is the computer vision analog of *perceptual organization* from Gestalt psychology.

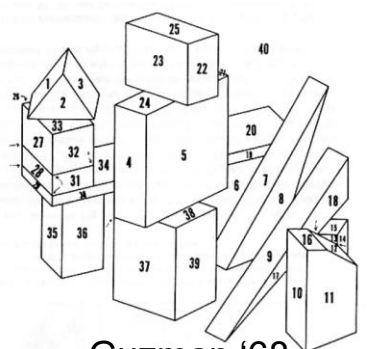
Computer vision is the capability of computers to perceive the natural world and extract information from it.
To move within it, to recognize and classify, and to make decisions upon that information and act.

A compelling idea: Vision comes so easily to most humans that it can’t be that difficult to program a computer to see. So it is that, in 1966, an unlucky intern at MIT’s AI lab was assigned ‘computer vision’ as a summer project.

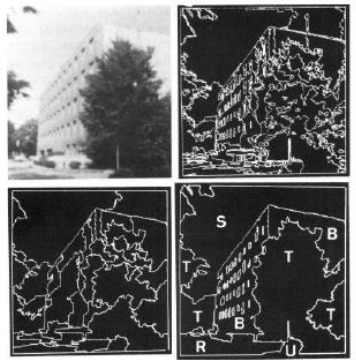
It turned out not to be so easy – even now academics struggled to make practical eye robots. Yet, in the past 10 years, we have seen significant progress and computer vision systems are now functional and deployable in masses – sometimes for better, and sometimes for worse.

Brief history of computer vision

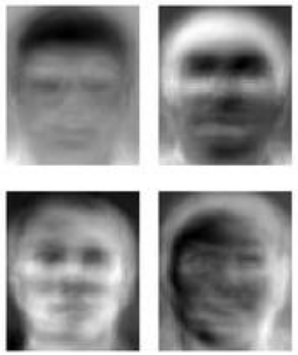
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1970’s: some progress on interpreting selected images
- 1980’s: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990’s: face recognition; statistical analysis in vogue
- 2000’s: broader recognition; large annotated datasets available; video processing starts
- 2010’s: Deep learning with ConvNets
- 2020’s: Widespread autonomous vehicles?



Guzman ‘68



Ohta Kanade ‘78



Turk and Pentland ‘91

Why is Computer Vision Hard?

What did you see?

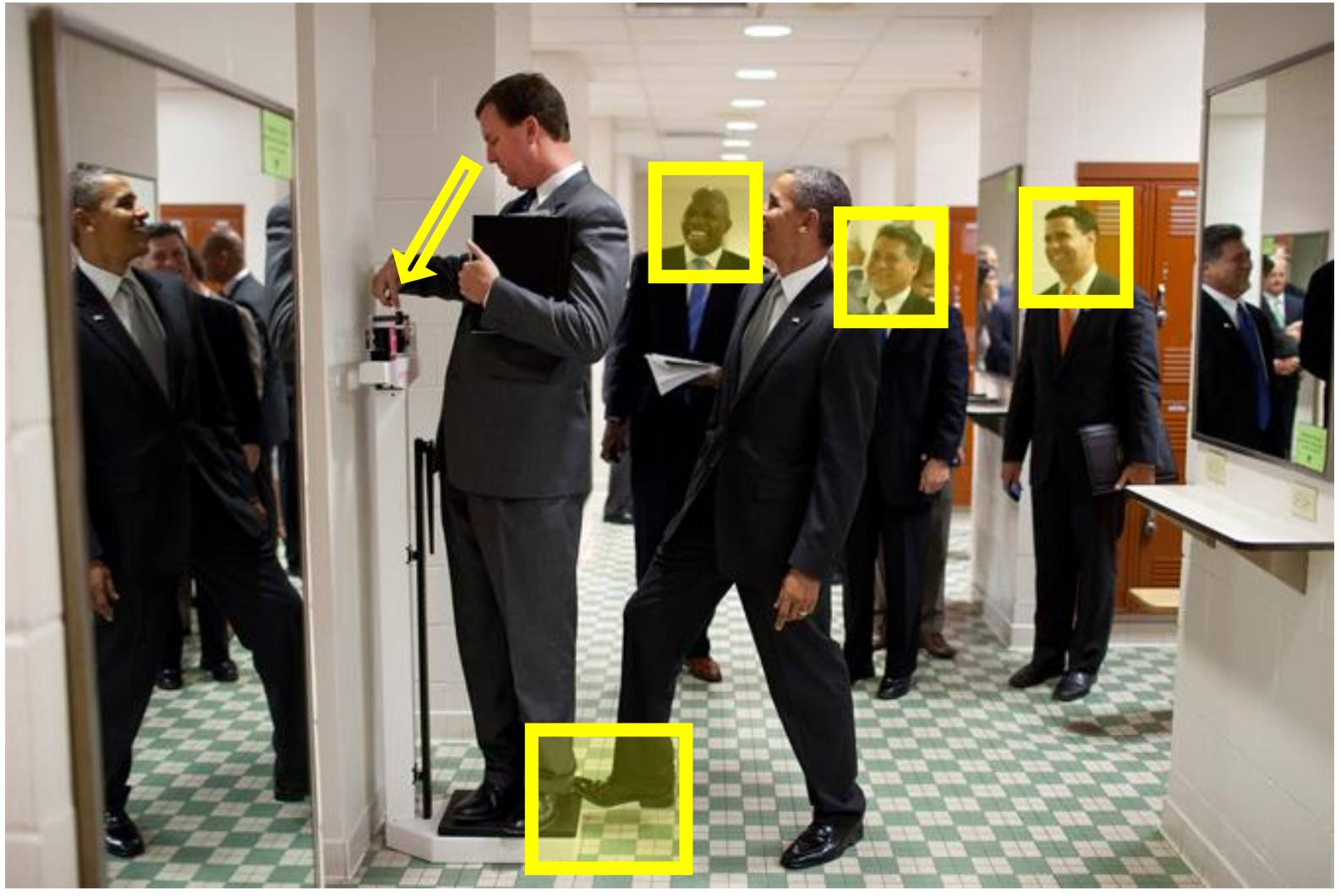
- Where this picture was taken?
- How many people are there?
- What are they doing?
- What object the person on the left standing on?
- Why this is a funny picture?

Why is Computer Vision Hard?

Why is Computer Vision Hard?

Why is Computer Vision Hard?

Why is Computer Vision Hard?



Why is Computer Vision Hard?

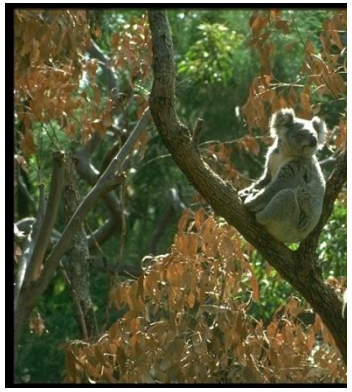
Why is Computer Vision Hard?

Computer: okay, it’s a funny picture

Challenges: Many nuisance parameters

Illumination

Object pose



Clutter

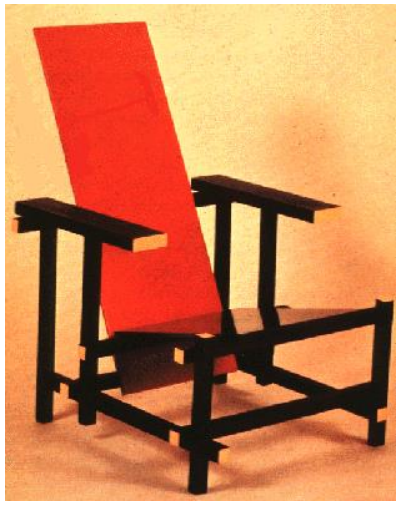
Occlusions

Intra-class appearance

Viewpoint

Slide credit: Kristen Grauman

Challenges: Intra-class variation



Slide credit: Fei-Fei, Fergus & Torralba

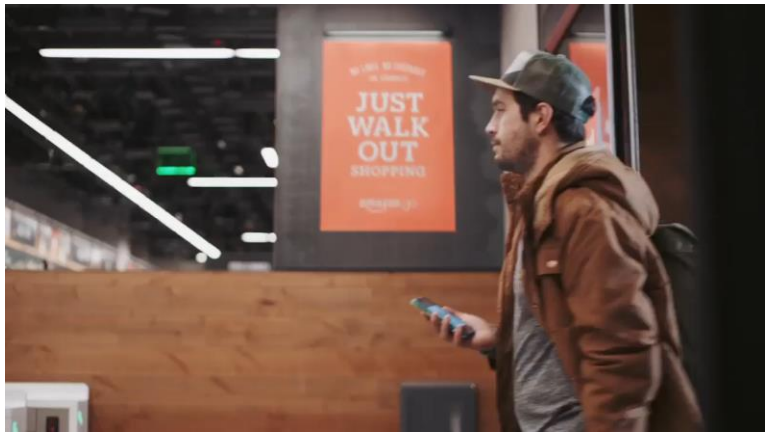
Some CV applications

Dawn of the Planet of the Apes

http://www.digitalspy.com/movies/oscars/feature/a584704/why-andy-serkis-deserves-an-oscar-nomination-for-planet-of-the-apes/?zoomable

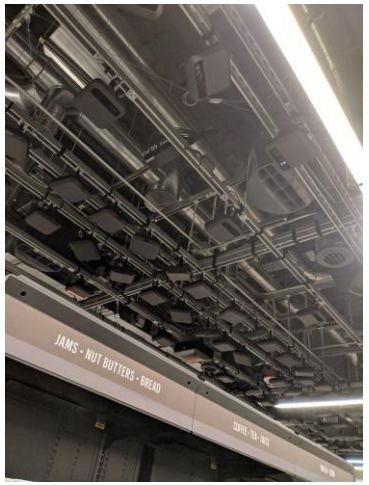
Amazon Go.
https://www.youtube.com/watch?v=NrmMk1Myrxc

Chandrayaan-3

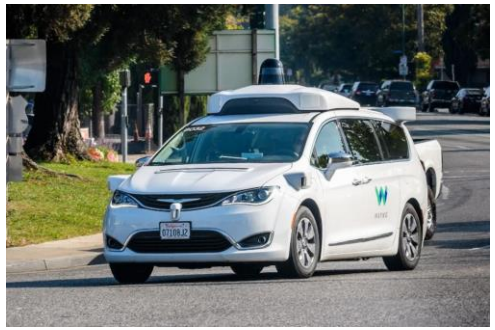


Cashier-less supermarkets are here. No need to bill your items. Just take them and walk out. Amazon Go is an example. Some Whole Foods branches also have this technology.

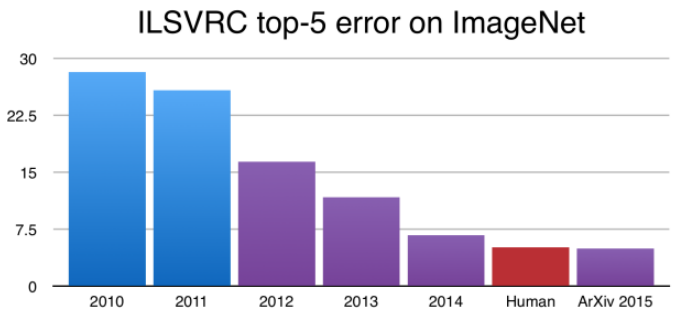
How does Amazon Go work? They use cameras (a lot of them – 100s) and computer vision. When you go these stores, look up. You will see the cameras.



CV Application continues…



Robotics



Classification

Hawk-Eye

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

Image Captioning

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Visual QA

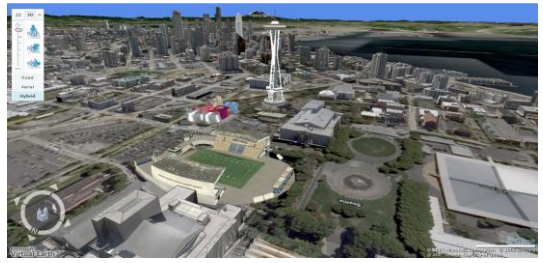


Image from Microsoft's Virtual Earth
(see also: Google Earth)

https://www.forbes.com/sites/jonmarkman/2018/12/21/self-driving-cars-are-speeding-into-view-thanks-to-google-and-nvidia/#a5016eaf80e4
https://in.pcmag.com/yuneec-typhoon-h-pro-with-intel-realsense-technology
https://srconstantin.wordpress.com/2017/01/28/performance-trends-in-ai/
https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2
https://visualqa.org/

Few more applications of CV….

Image Credit: James Tompkin

Text to Image Synthesis

Image Credit: James Tompkin

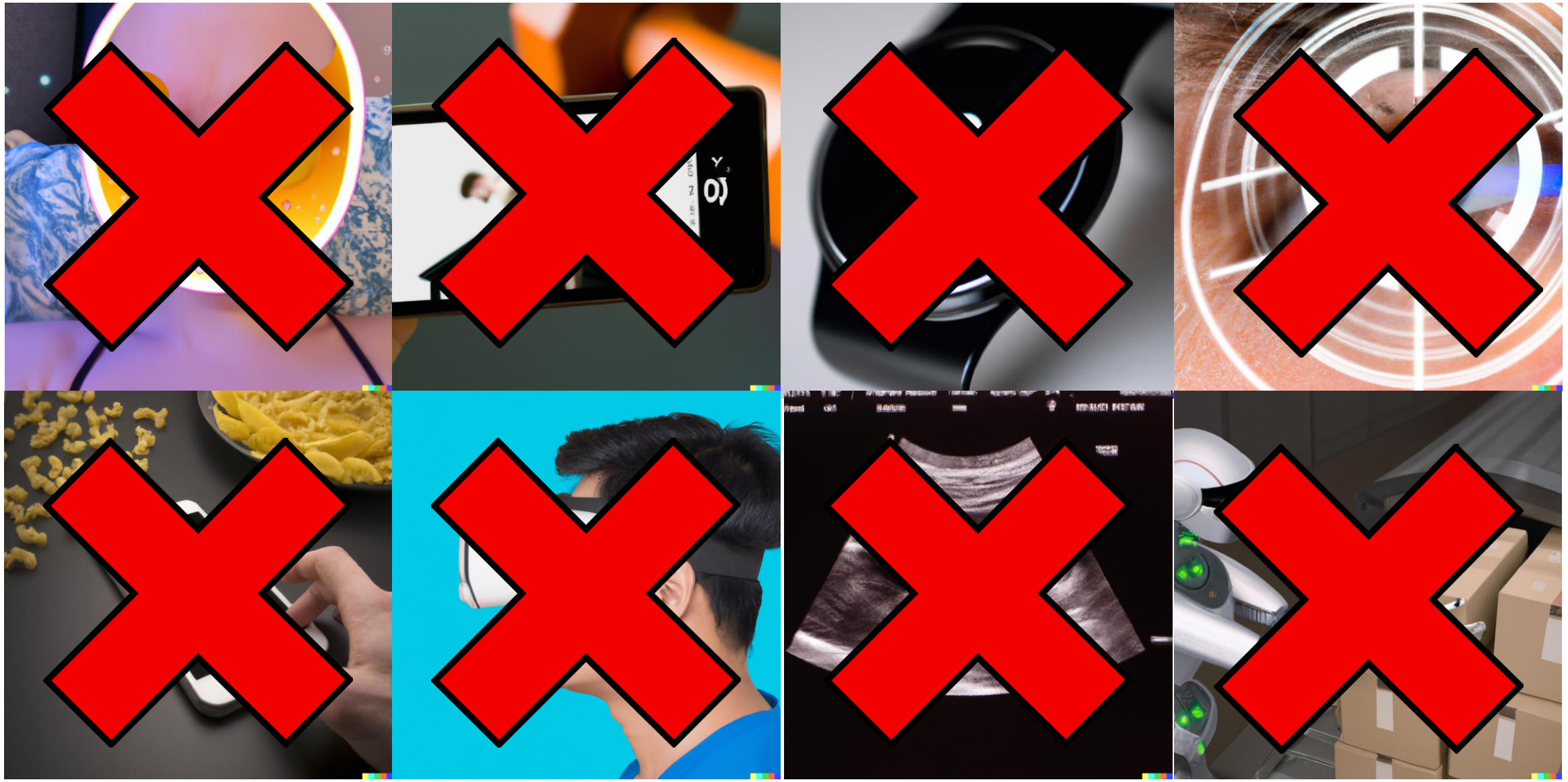
“Teddy bears working on new AI research underwater with 1990s technology”

“Cats playing chess”

“a teddy bear at Brown University”

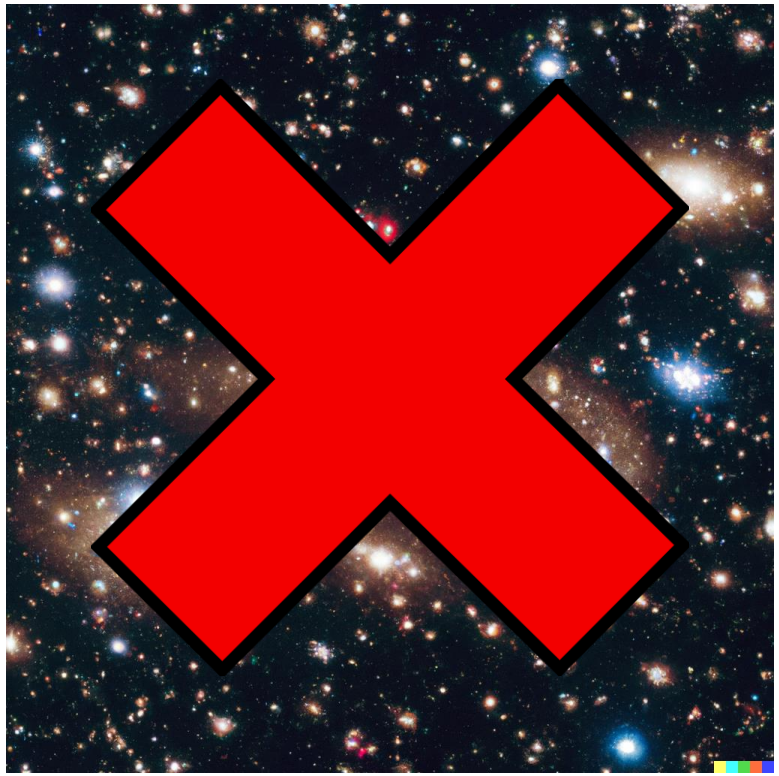
Real or Fake

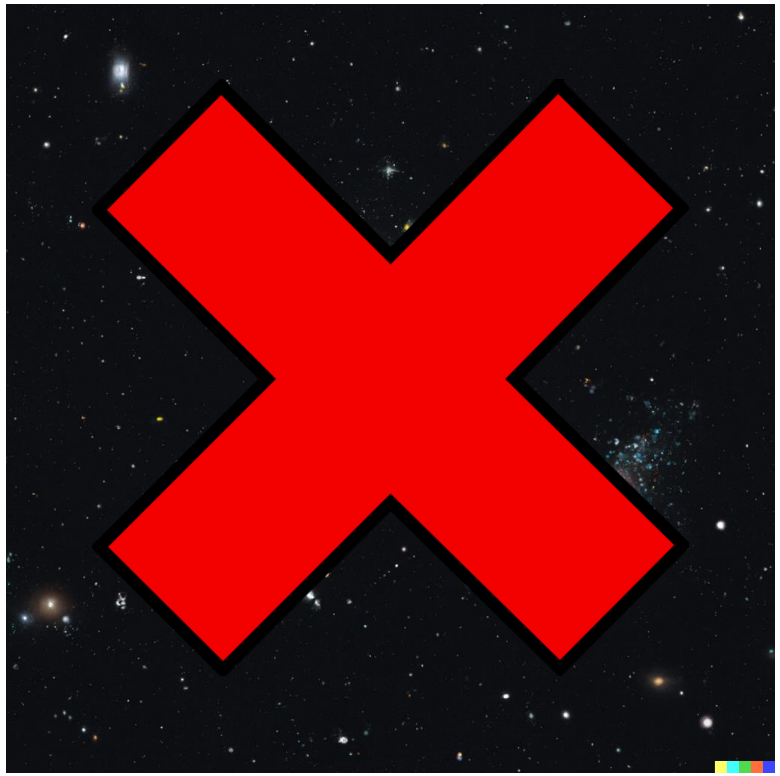
Image Credit: James Tompkin



Few More…

Image Credit: James Tompkin





Current status: Computer vision is ubiquitous…

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

Smartwatches: Heart rate detection, proximity detection

Security: Fingerprint/iris/face scanning (offices, airports), CCTV monitoring

Laptops/Desktops: Biometrics auto-login (face recognition, 3D)

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), Photoshop, PowerPoint (captioning, design suggestions)

Virtual Worlds: VR/AR head tracking (Oculus, HTC Vive), simultaneous localization and mapping, person tracking (Kinect), gesture recognition, virtual try-on, digital humans

Telepresence: Virtual backgrounds (Zoom, Google Meet), webcam person/face following

Media: Visual effects for film/TV, virtual sports replay, semantics-based auto edits

Transportation: Assisted driving (cruise control, self-driving), face tracking/iris dilation for safety

Supermarkets: Cashier-less checkout, theft detection (Walmart), fruits/vegetables sorting, packaging, manufacture

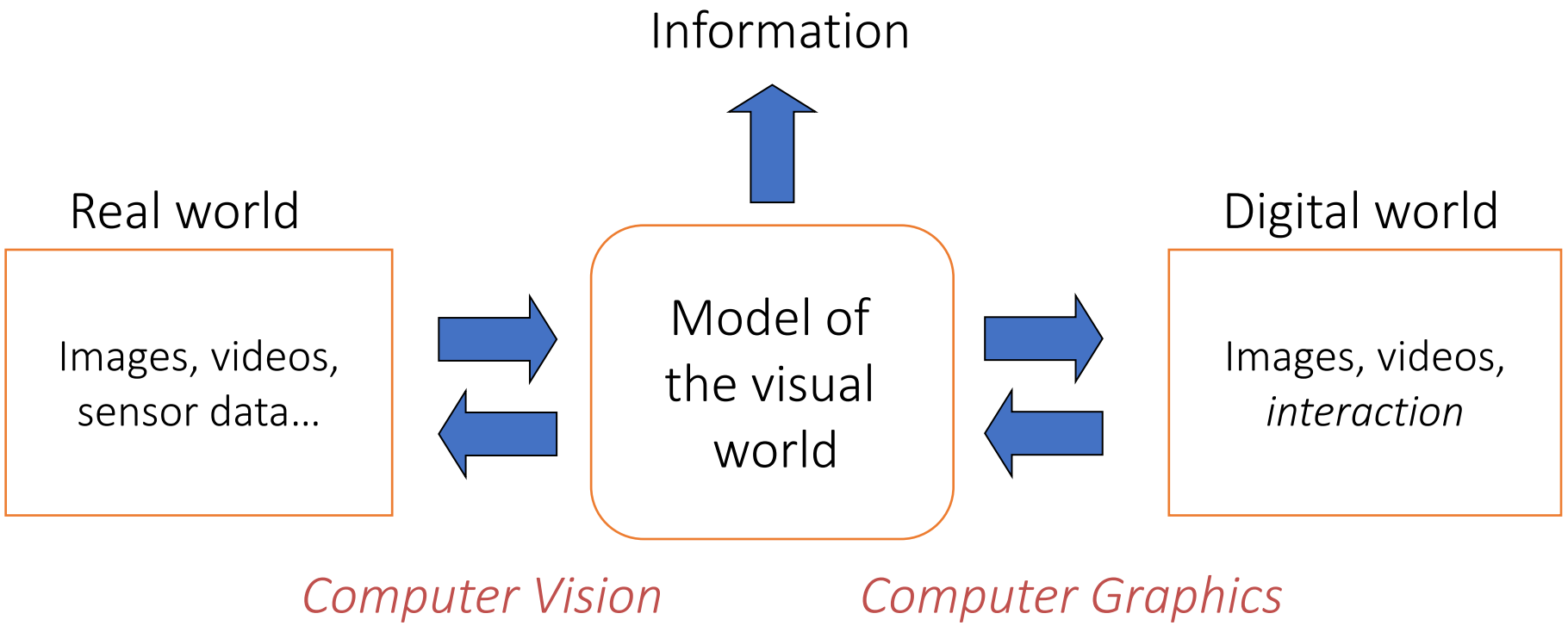
Medical imaging: CAT / MRI reconstruction, assisted diagnosis, automatic pathology, connectomics, endoscopy

Space Exploration: Mar rovers, space telescopes (Hubble, James Webb)

Industry: Vision-based robotics, online shopping (Amazon, Walmart), machine-assisted router (jig), OCR (USPS), ANPR (number plates for tolls), drones

Creative: Photoshop, vision-language models for image generation (Dall-E), video editors

CV and computer graphics



Computer Vision and Computer Graphics are the inverse of each other.

Future ahead….

Derogatory current summary of computer vision: Machine learning applied to visual data

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

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

WHAT’S Missing?... ‘world-sense’.

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

Needs understanding of possible failures + Responsible approach + Techniques to overcome bias.

Course Outline

Basics of image processing, including:

- Filtering in pixel and frequency domains,
- Linear and non-linear filters
- Feature extraction
- Feature matching using RANSAC
- Edge detection

Basics of ML, including Dimensionality reduction and Ensemble learning

Basics of Deep Learning and CNN

- Image classification
- Image segmentation
- Improving trustworthiness of CNN, using adversarial examples, out of distribution sample detection, and Interpretability of CNN

Basics of Transformers and its improvements

Applications of CV in medical imaging, like Remote-PPG and Medical image segmentation