Prerequisites

- **Linear algebra**
- **Probability theory**
- **Statistics**
- **Calculus**
- **Basics in Machine Learning**
- **Programming (MATLAB/Python)**

What is Computer Vision?

- A field of study that seeks to develop techniques to help computers "see" and understand the content of digital images such as photographs and videos.

What kind of scene? Where is the river?

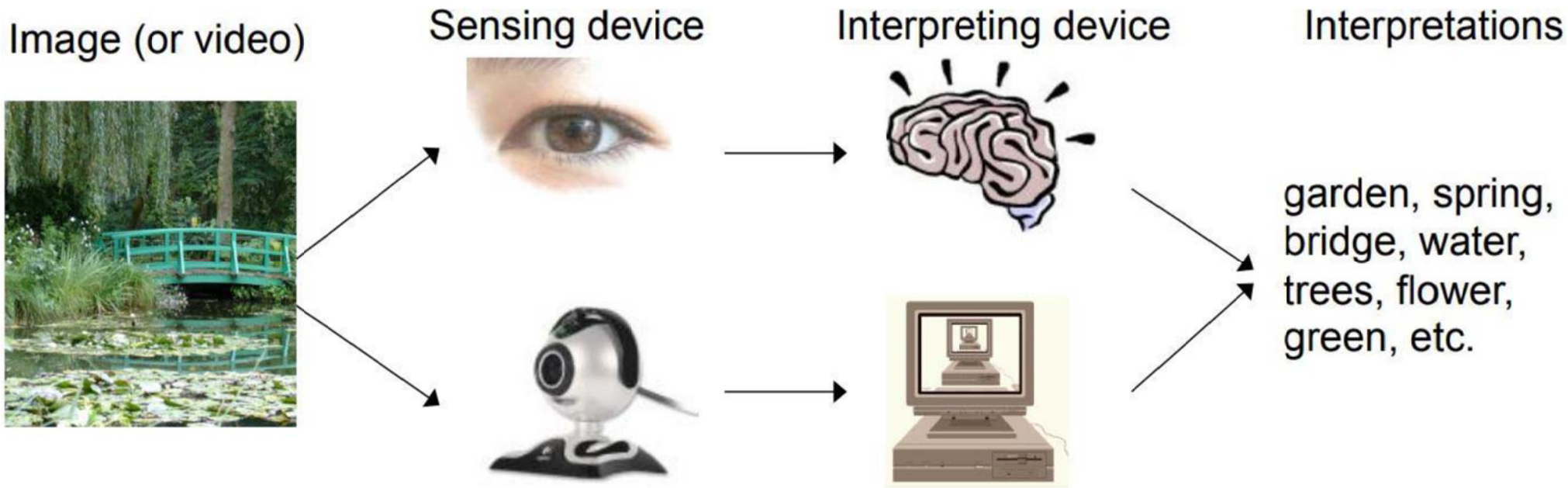
How far is the hill? and so on.

What is Computer Vision?

What is wrong with these images?

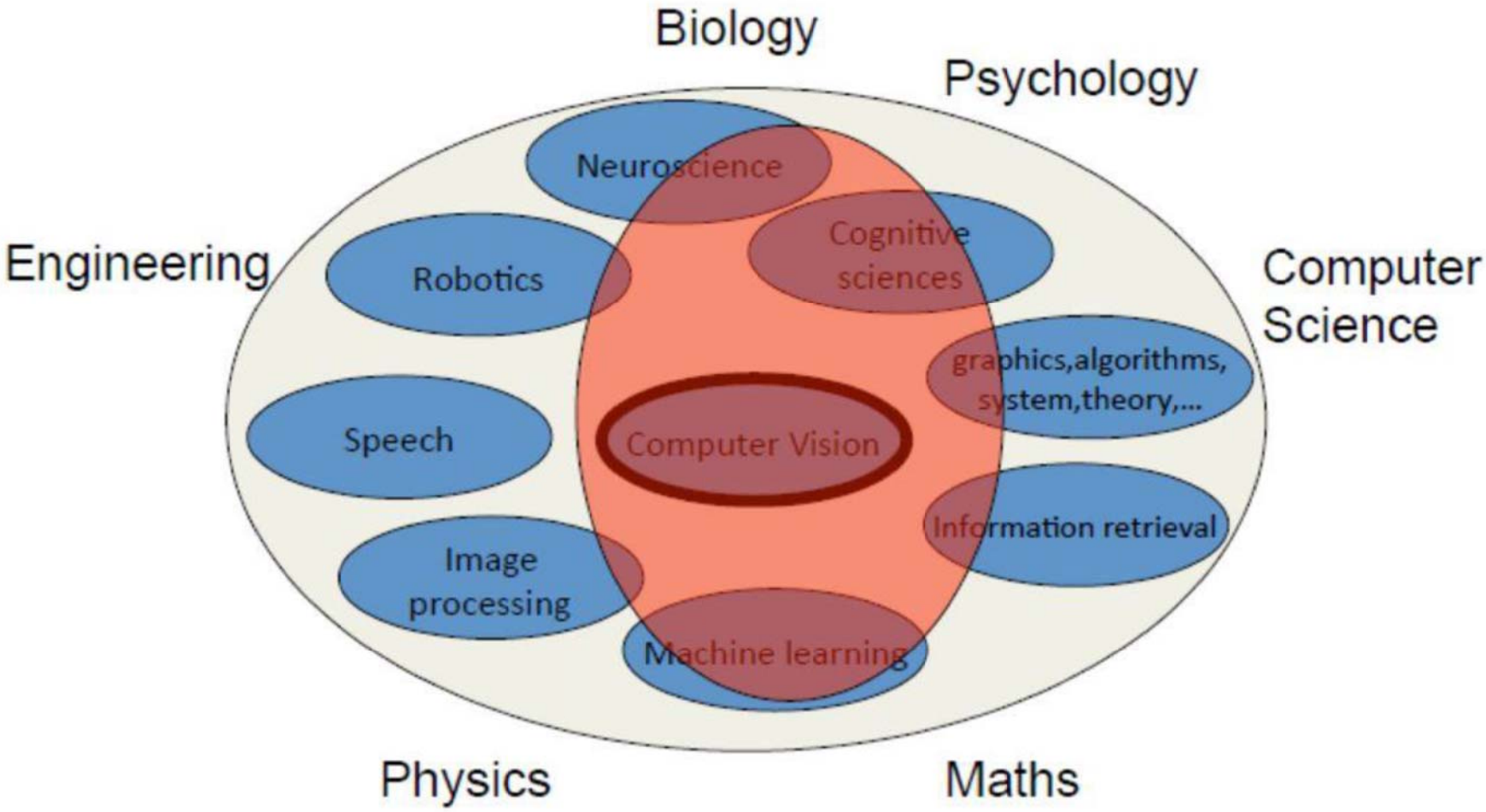
Can a machine answer these questions?

What is Computer Vision?



What it is related to?

An interdisciplinary field



Definition of Vision

Vision (human vision or animal vision) has two components:

Sensing device and

Interpreting device

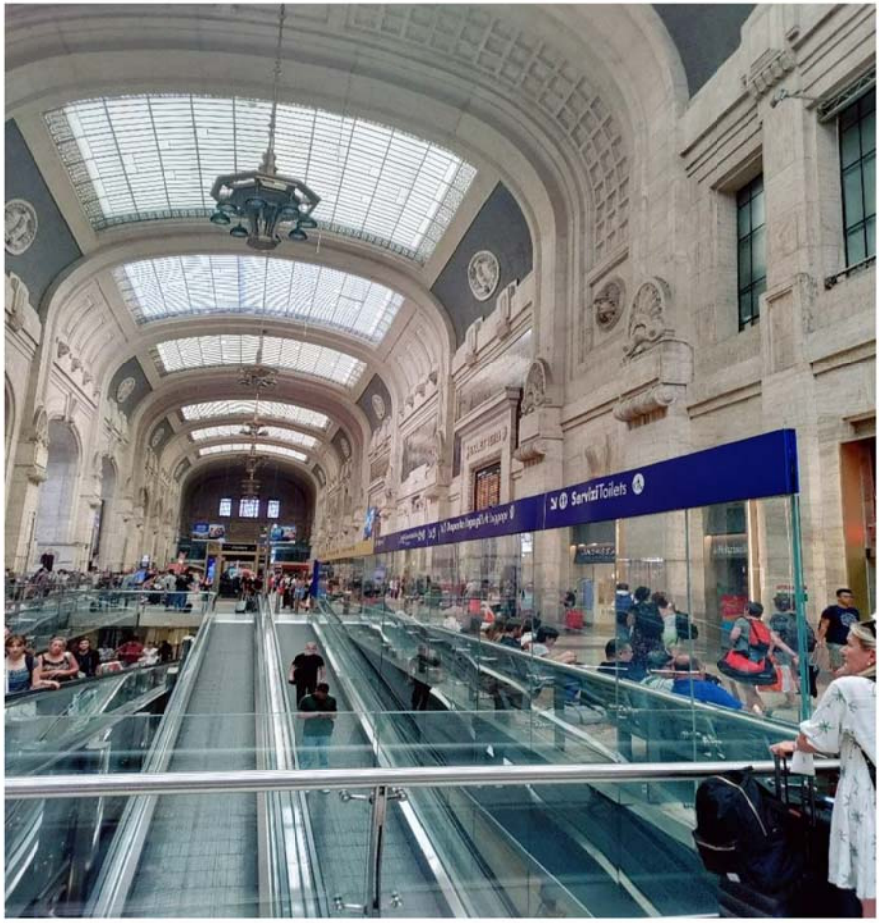
Every Picture Tells a Story

Goal of computer vision: perceive the "story" behind the picture

• Compute properties of the world
– 3D shape
- Number of persons
– Names of person or objects
– What happened?

Goal of Computer Vision

To bridge the gap between pixels and "meaning"



What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Can Computer Match Human Perception?

- **Yes and no** (mainly no)

computers can be better at “easy” things

-humans are much better at “hard” things

Can Computer Match Human Perception?

Yes and no (mainly no)

computers can be better at "easy" things

-humans are much better at "hard" things

But huge progress has been made

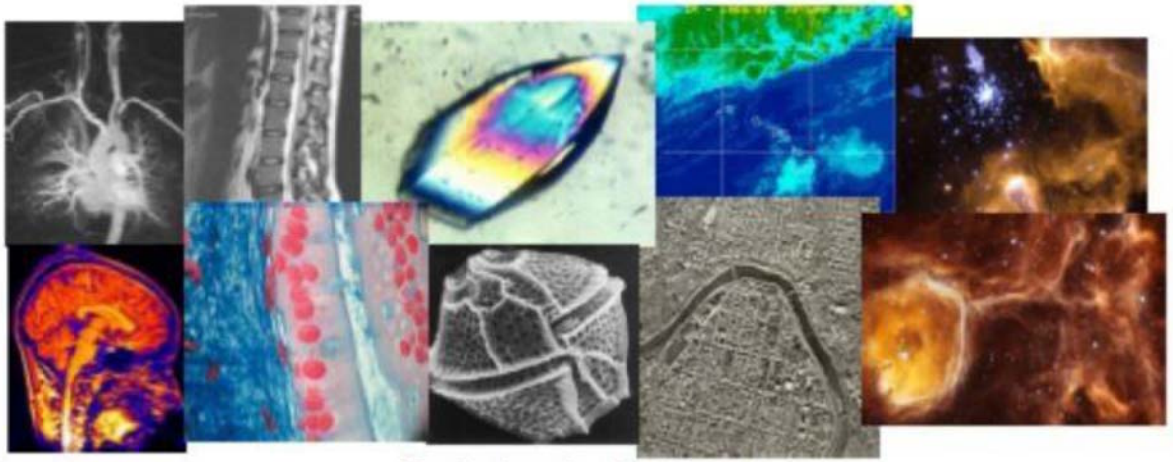
- Especially in the last 10 Years
- What is considered "hard" keeps changing

Is Computer Vision useful?

Vision is useful! Images and videos are everywhere!

Google Image Search Picasa™ flickr webshots picsearch YouTube

Surveillance and security



Medical and scientific images

Visual data on the Internet

- Flickr
  - 10+ billion photographs
  - 60 million images uploaded a month
- Facebook
  - 250 billion+
  - 300 million a day
- Instagram
  - 55 million a day
- YouTube
  - 100 hours uploaded every minute

Cisco Predicts That 90% of all Internet Traffic will be Visual

Why is Computer Vision Difficult?

Viewpoint variation

Illumination

Scale

Why is Computer Vision Difficult?

Intra-class variation

Motion (Source: S. Lazebnik)

Background clutter

Occlusion

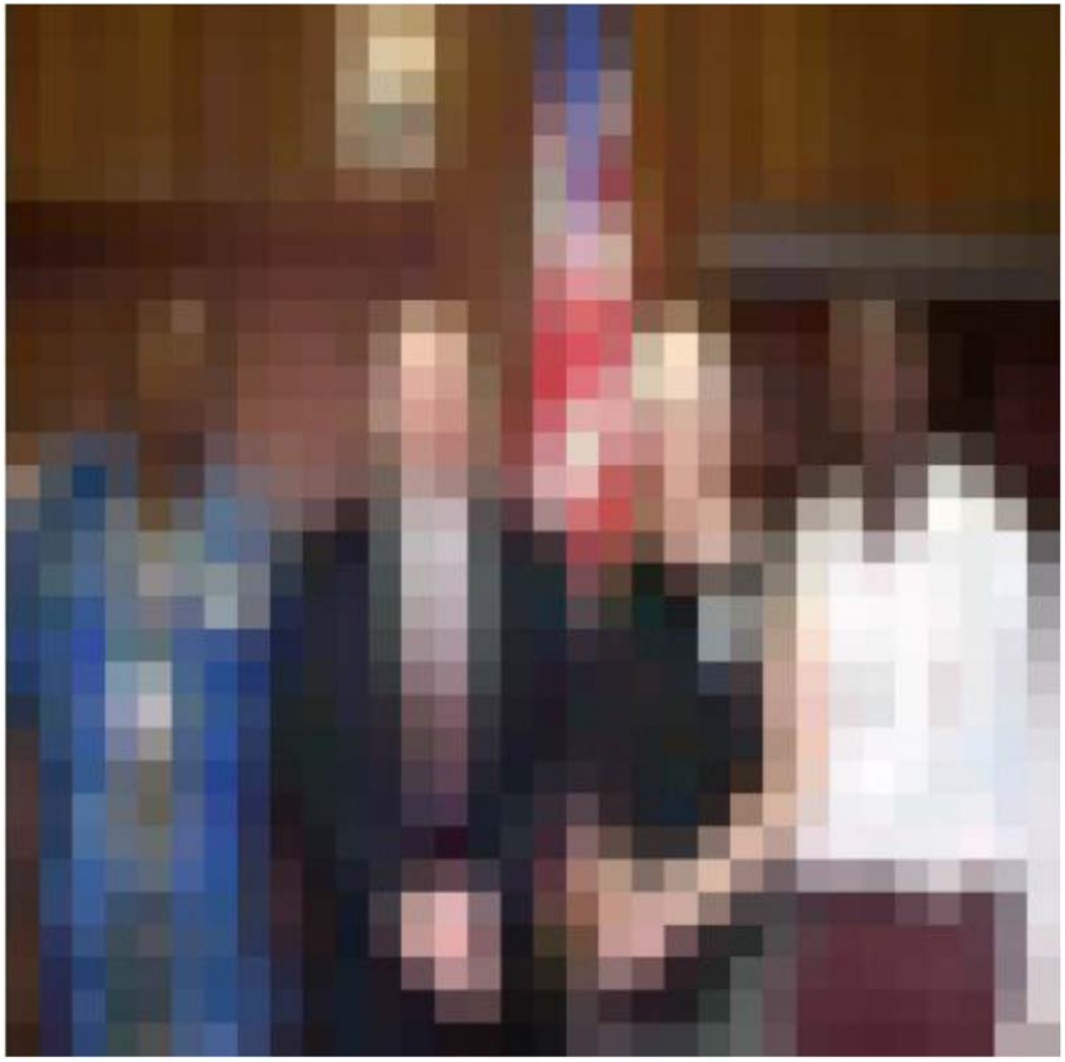
Why is Computer Vision Difficult?

Challenges: Local ambiguities

Human perception has its shortcomings

Sinha and Poggio, Nature, 1996
(The Presidential Illusion)

But humans can tell a lot about scene from a little information….



Source: "80 million tiny images" by Torralba, et al.

Origin of Computer Vision

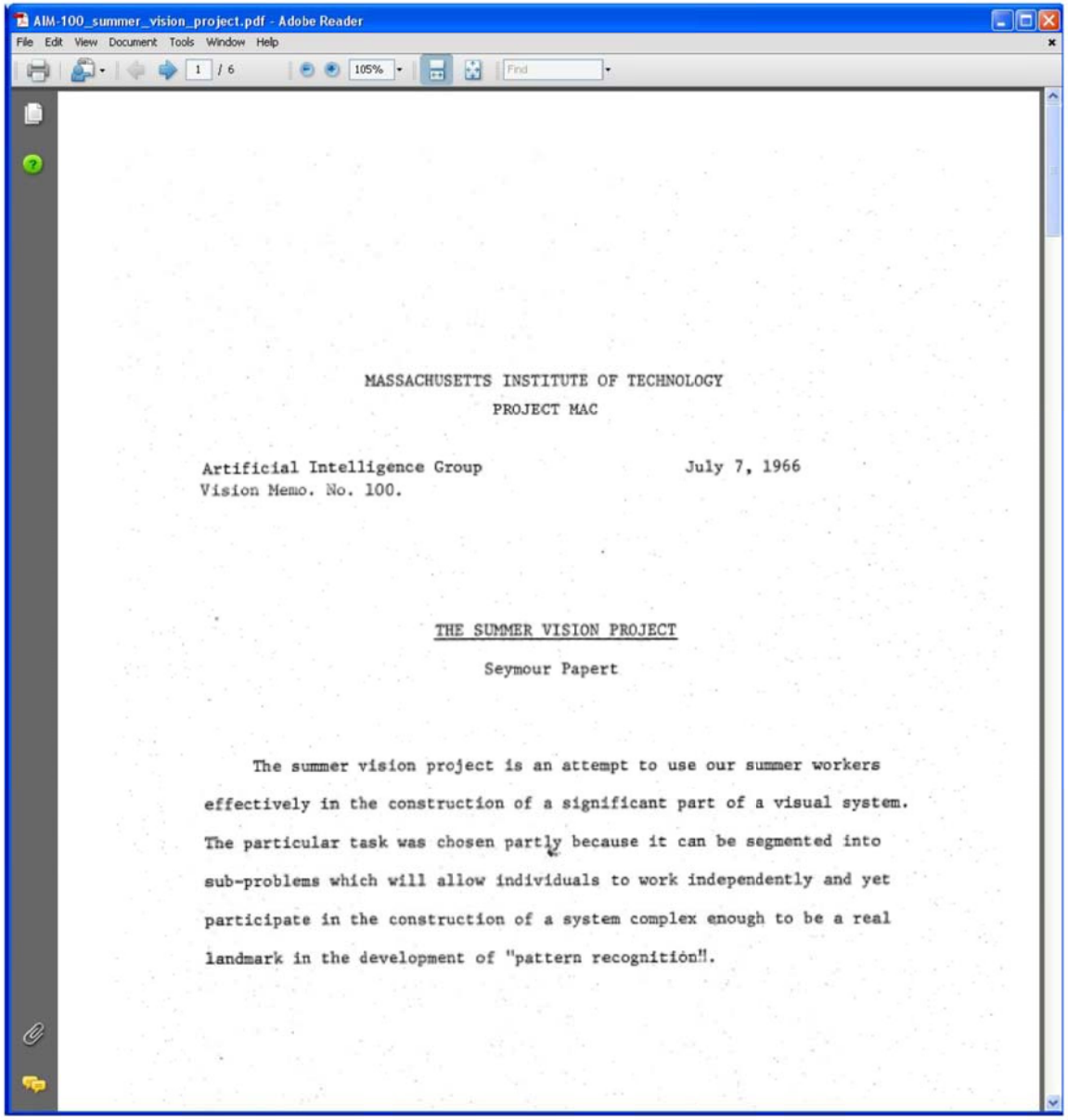
In 1966, Minsky hired a first year under graduate student and assigned him a problem to solve over the summer:

Connect a camera to a computer and get the machine to describe what it sees.

Marvin Minsky (Turing Awardee, 1969)

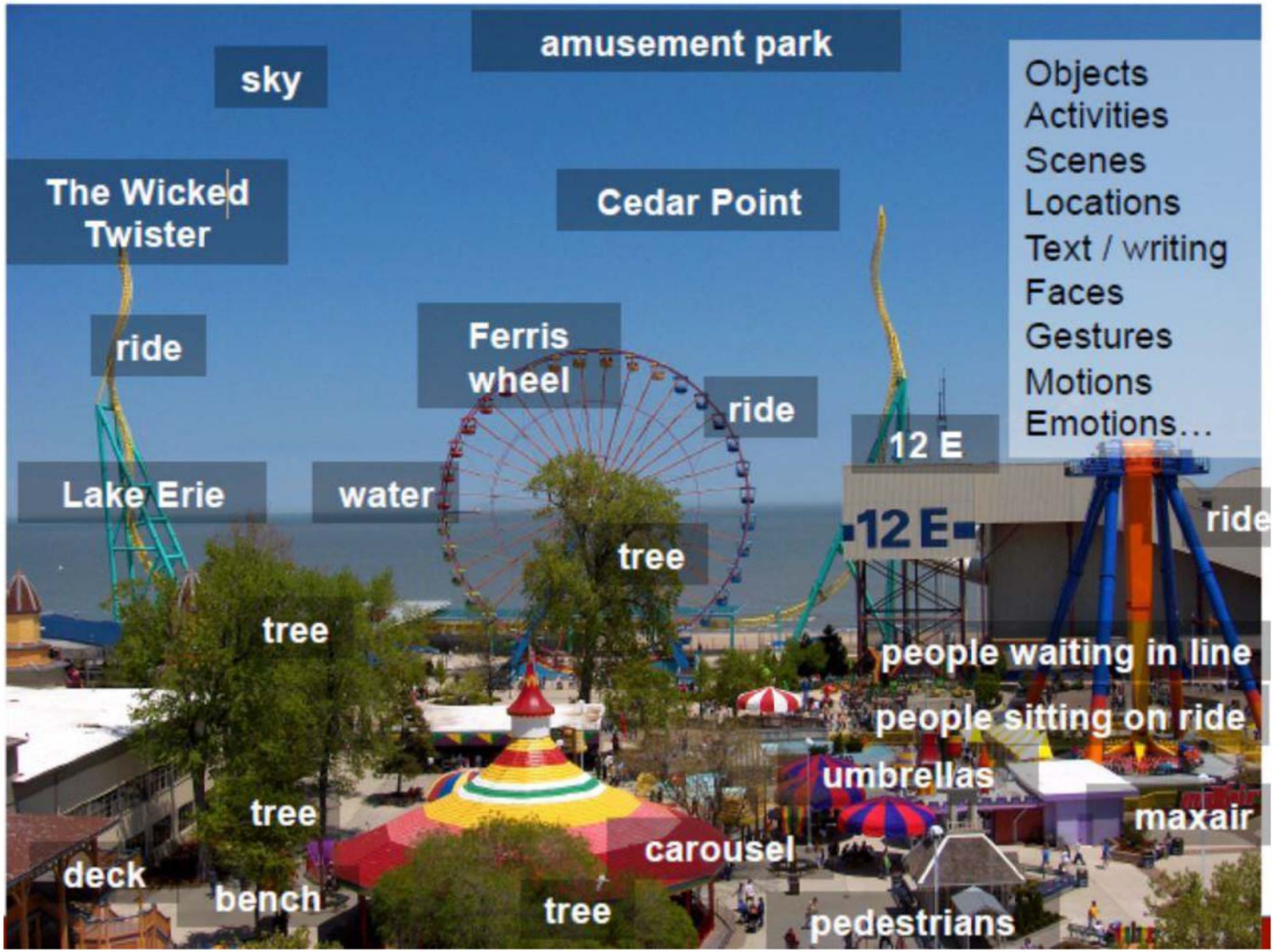
Origin of Computer Vision

An MIT Undergraduate Summer Project



Vision as a source of semantic information

Vision as a source of semantic information



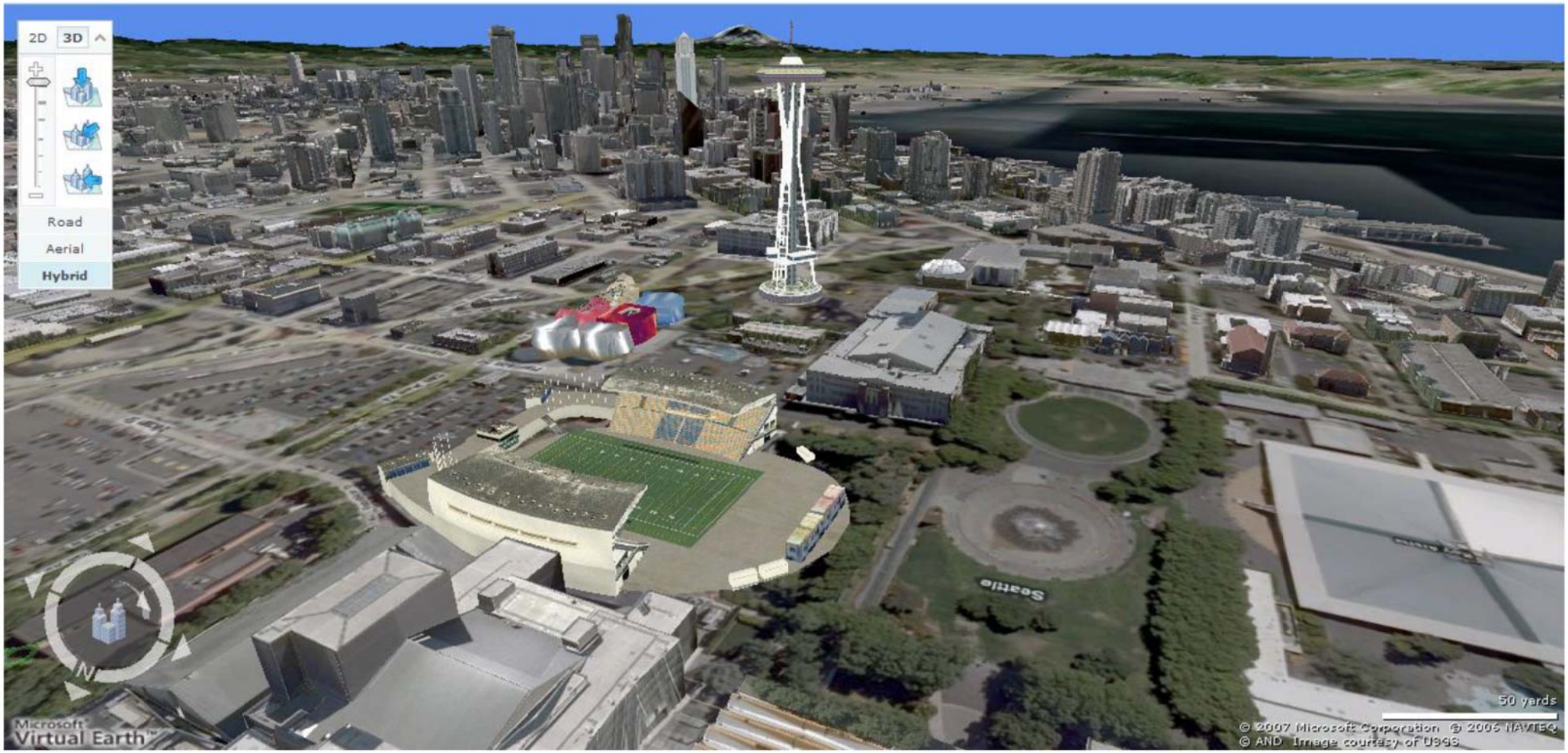
Computer Vision and Image Processing

- Computer vision is distinct from image processing.
- Image processing is the process of creating a new image from an existing image.
- A given computer vision system may require image processing to be applied to raw input, e.g. pre-processing images.
- The goal of computer vision is to understand the content of digital images.
- Typically, this involves developing methods that attempt to reproduce the capability of human vision.

Computer Vision: Text Books

- Richard Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010
- Simon Prince, *Computer Vision: Models, Learning, and Inference*, 2012
- David Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*, 2002
- Bishop, Christopher M, *Pattern Recognition and Machine Learning*, Springer, 2006
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, 2016

Earth Viewers (3D Urban modeling)

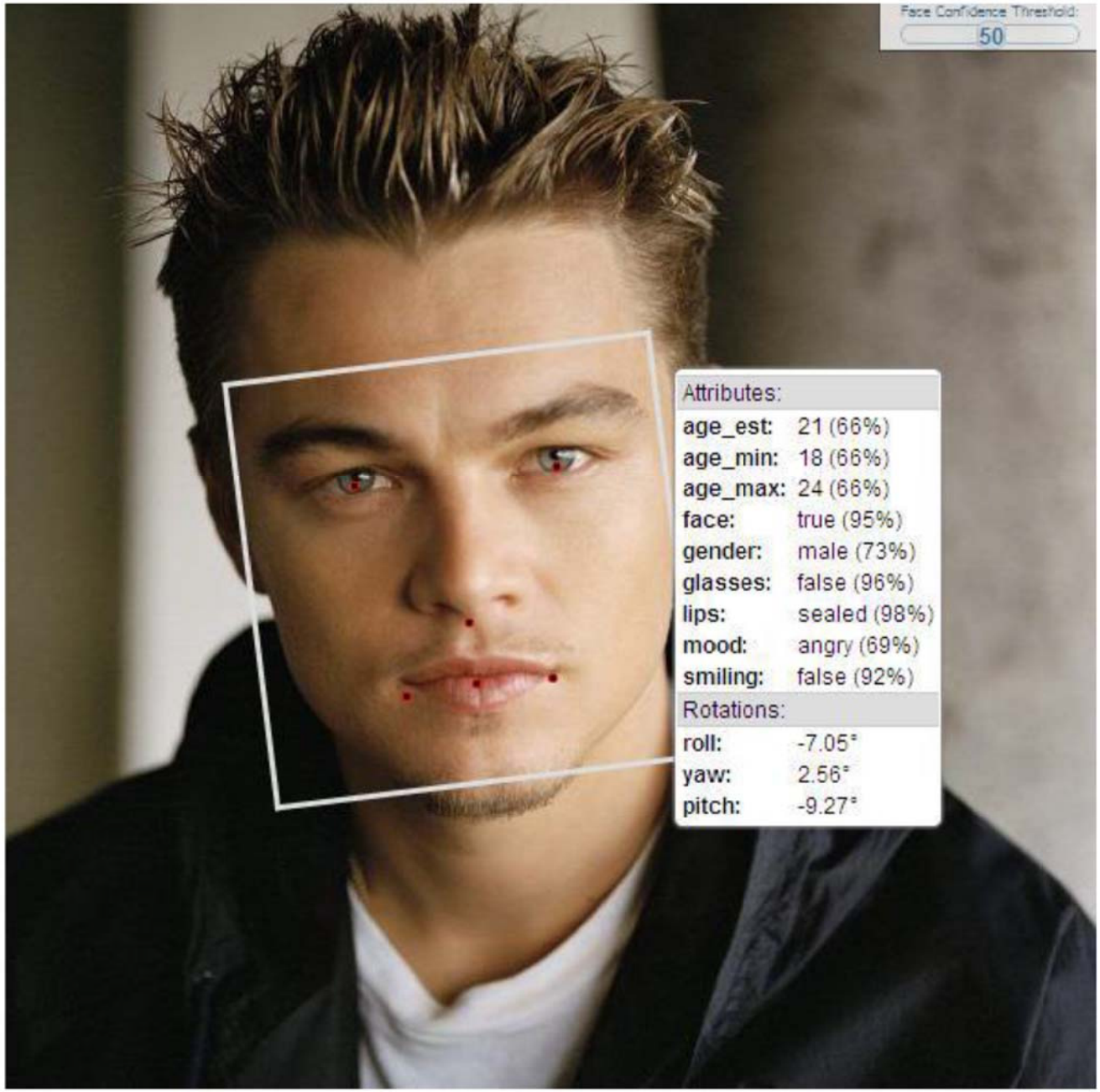


Face Detection

Many new digital cameras now detect faces

– Canon, Sony, Fuji,

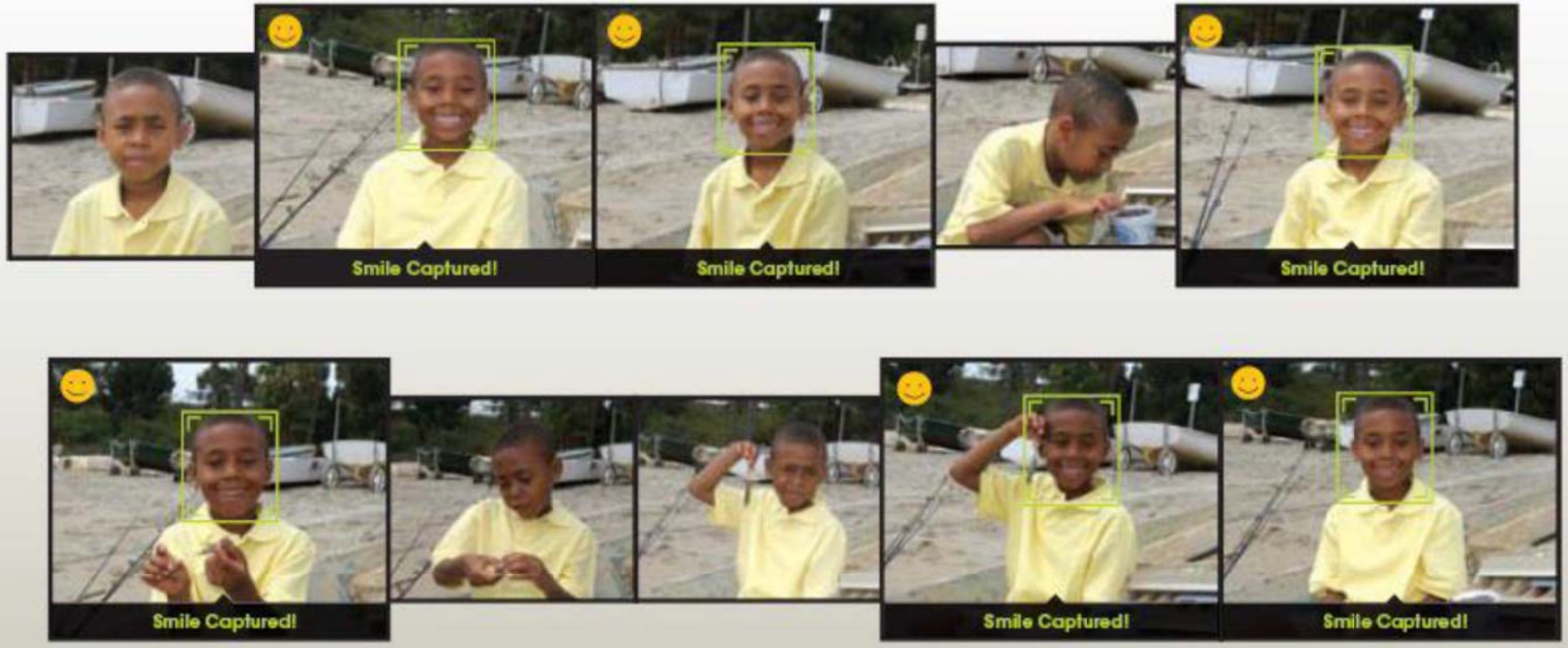
Face Analysis and Recognition



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.

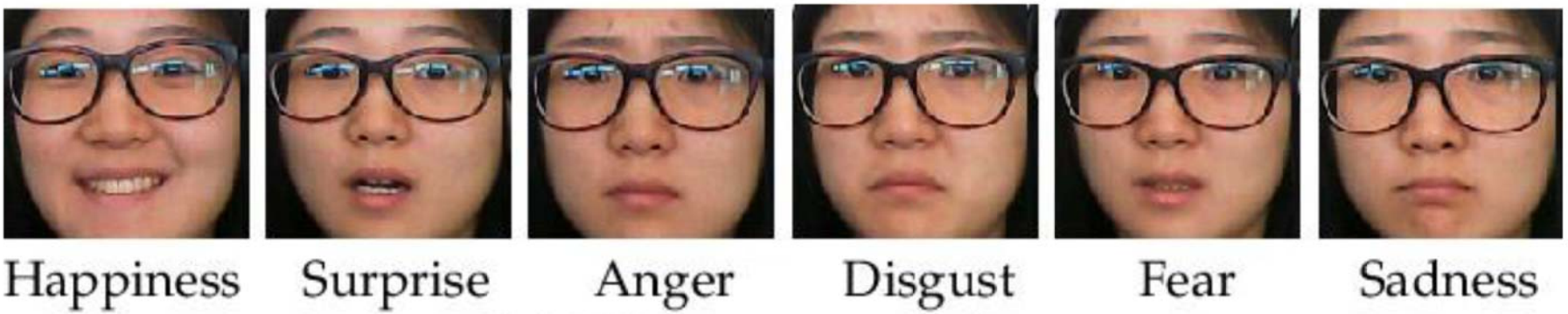


Sony Cyber-shot® T70 Digital Still Camera

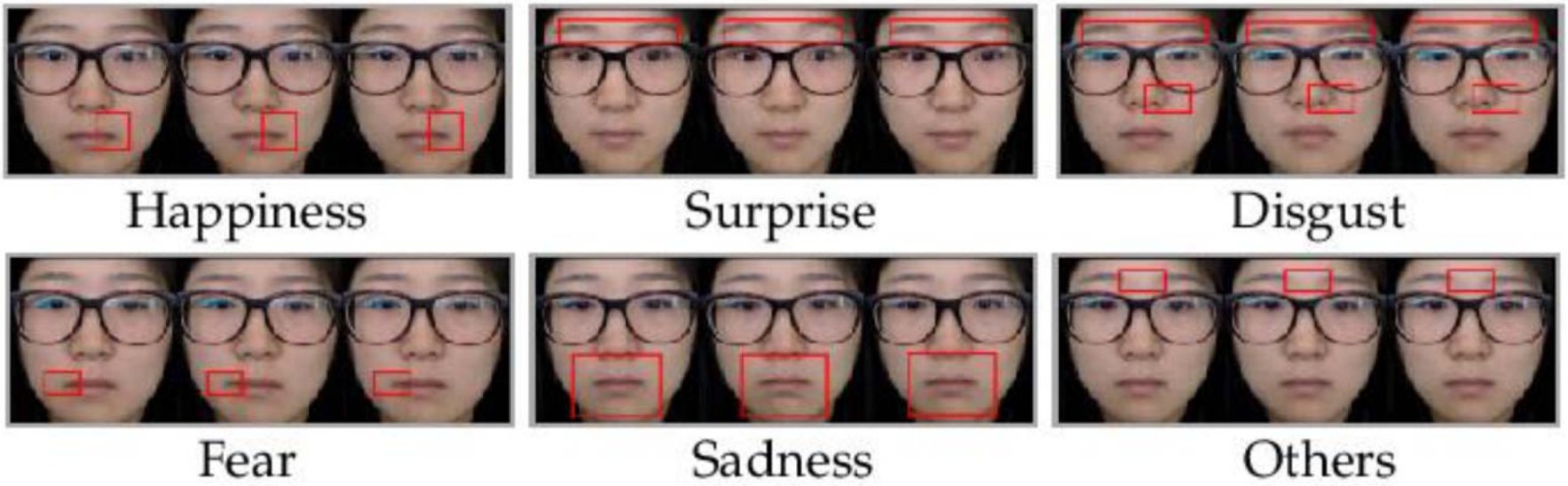
Facial Expression Recognition

Sony Cyber-shot® T70 Digital Still Camera

Macro vs Micro Facial Expression

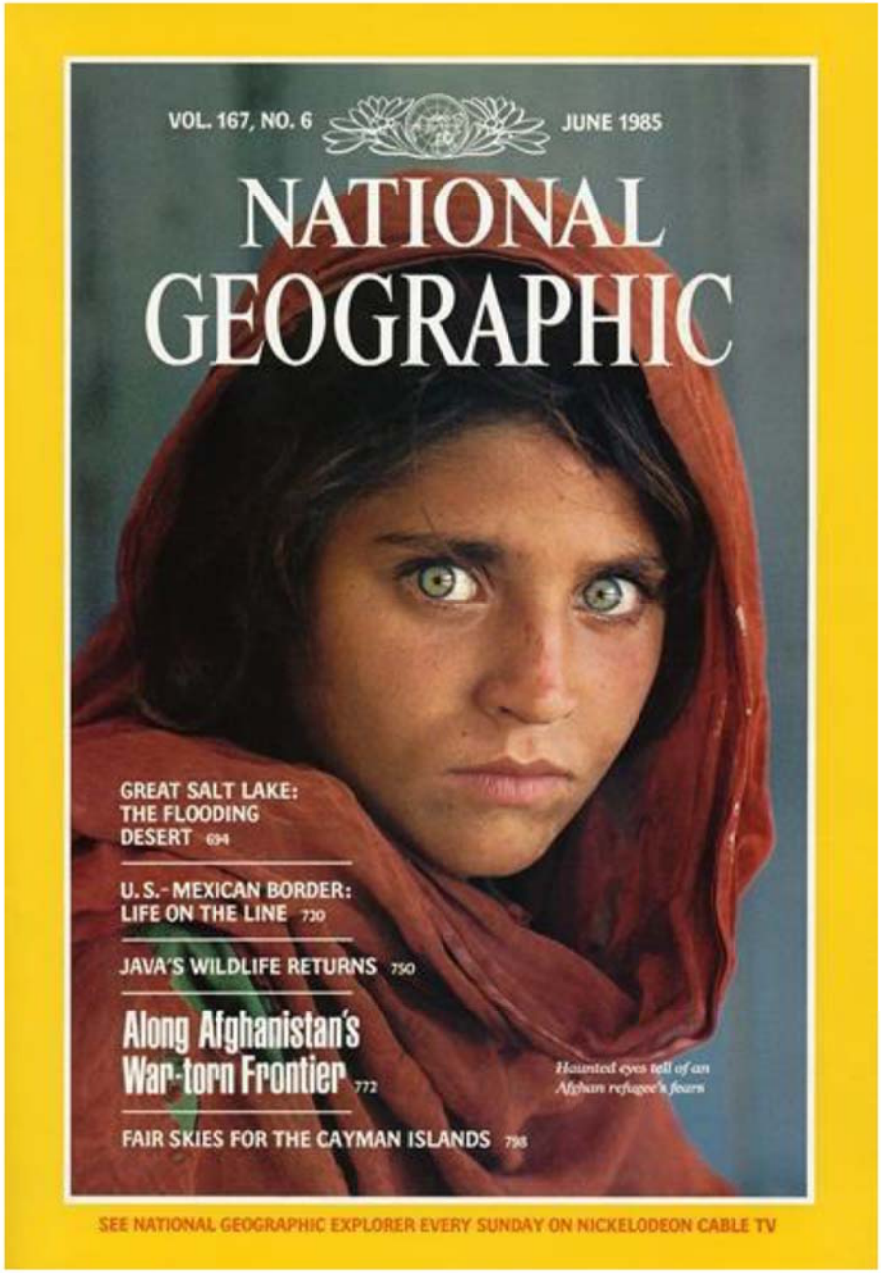


(a) Macro-expressions



(b) Micro-expressions

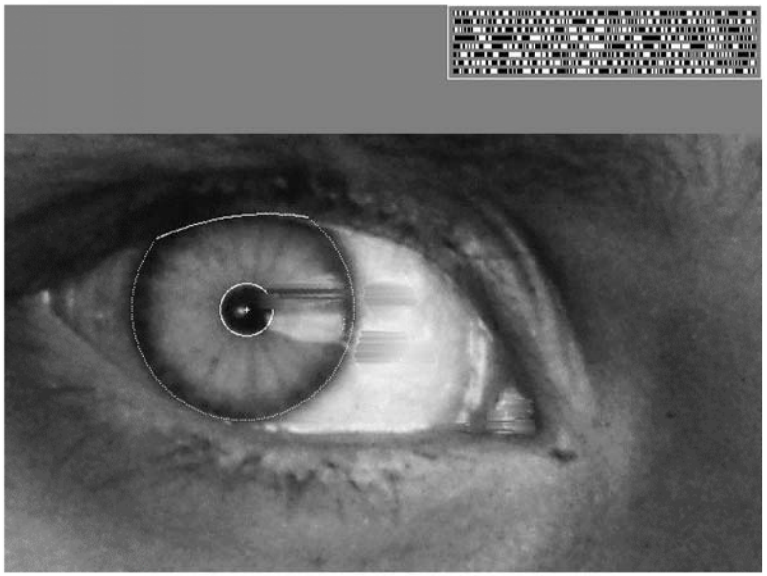
Vision-based Biometrics

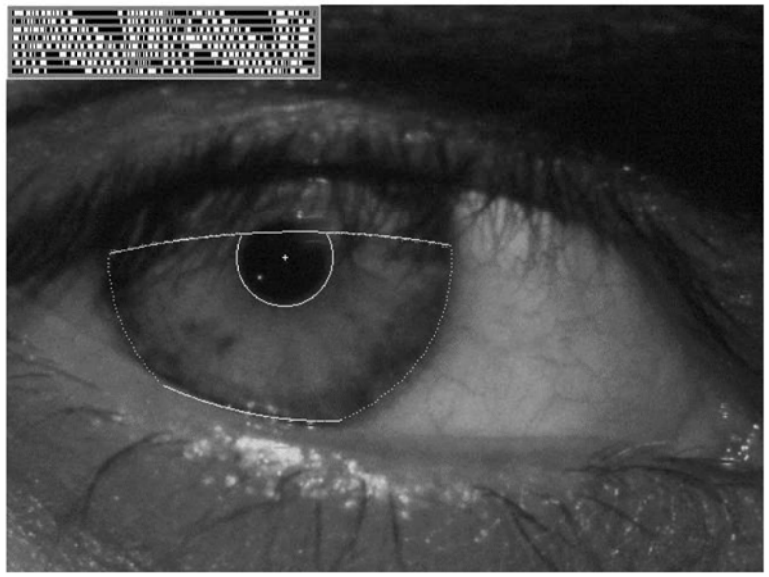


Who is she?

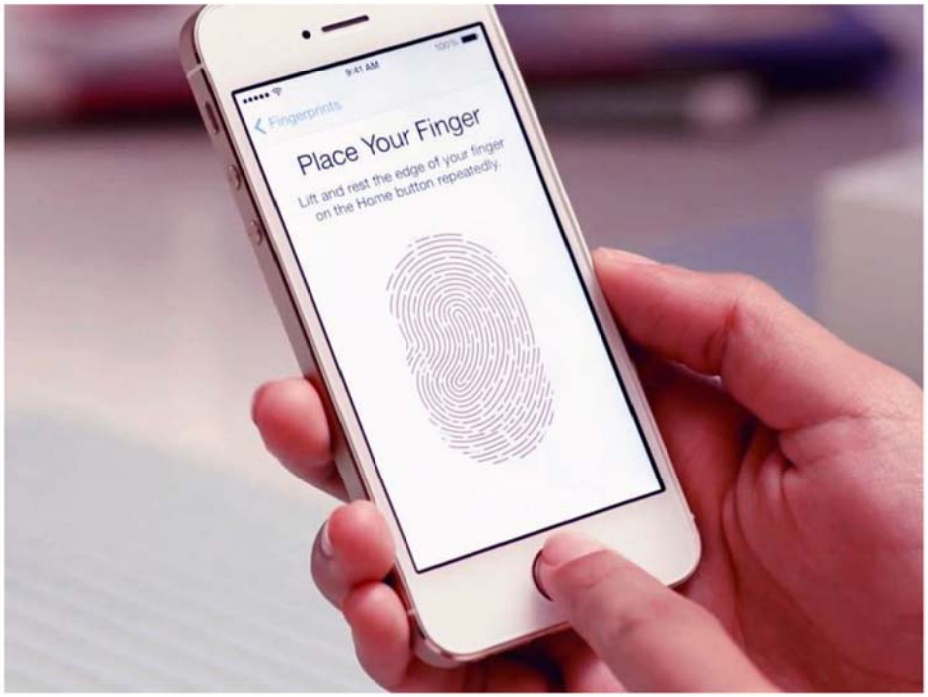
Vision-based Biometrics

"How the Afghan Girl was Identified by Her Iris Patterns"





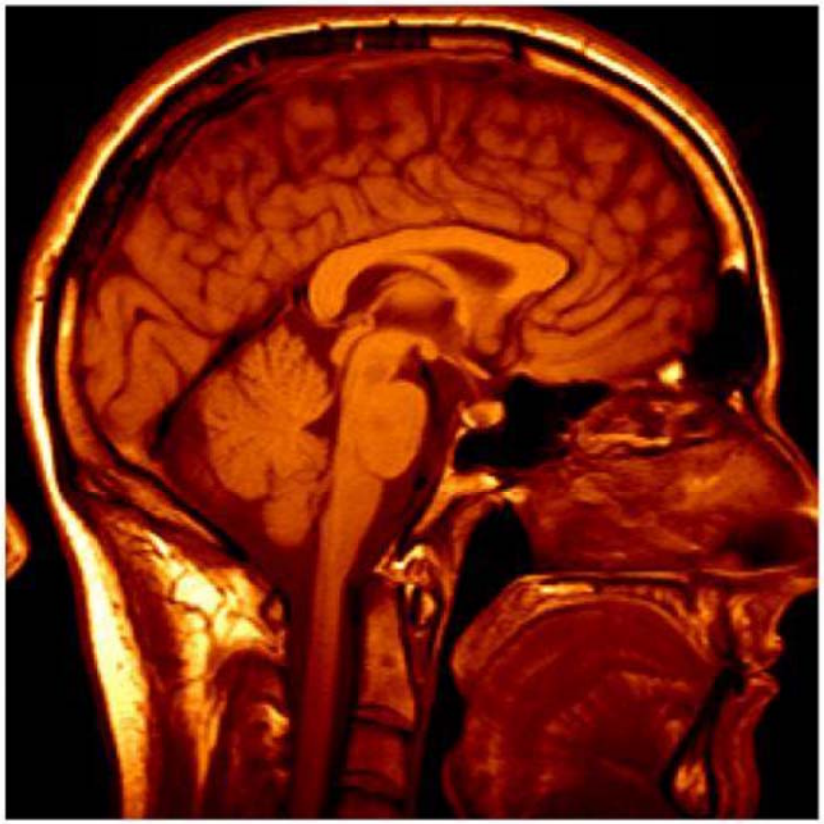
Login without a password



Fingerprint scanners in many new smartphones and other devices

Face unlock on Apple iPhone X

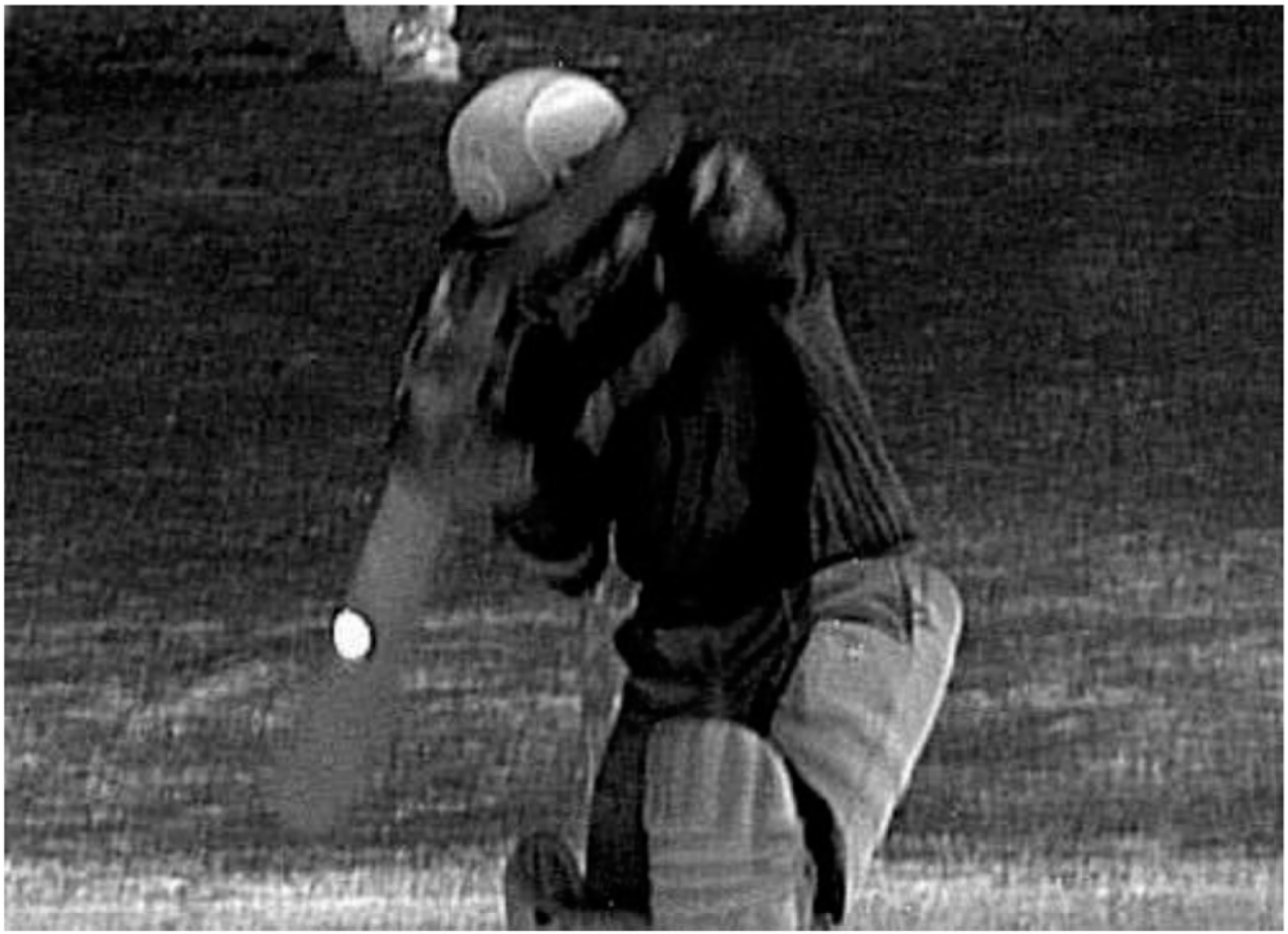
Medical Imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Sports



Hotspot Technology in Cricket
(Infrared imaging system)

Sports

Hotspot Technology in Cricket
(Infrared imaging system)

Sports

Sportvision first down line

Optical Character Recognition (OCR)

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

Digit recognition, AT&T labs (1990's)
http://yann.lecun.com/exdb/lenet/

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

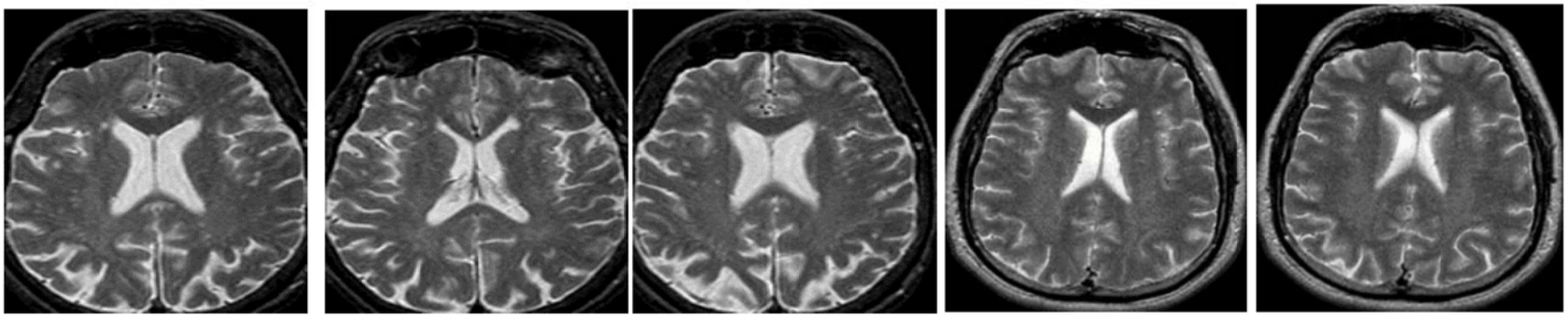
Automatic check processing

Sudoku grabber
http://sudokugrab.blogspot.com/

Scene Image Classification

Medical Image Classification

Normal brain MR image



Abnormal brain MR image

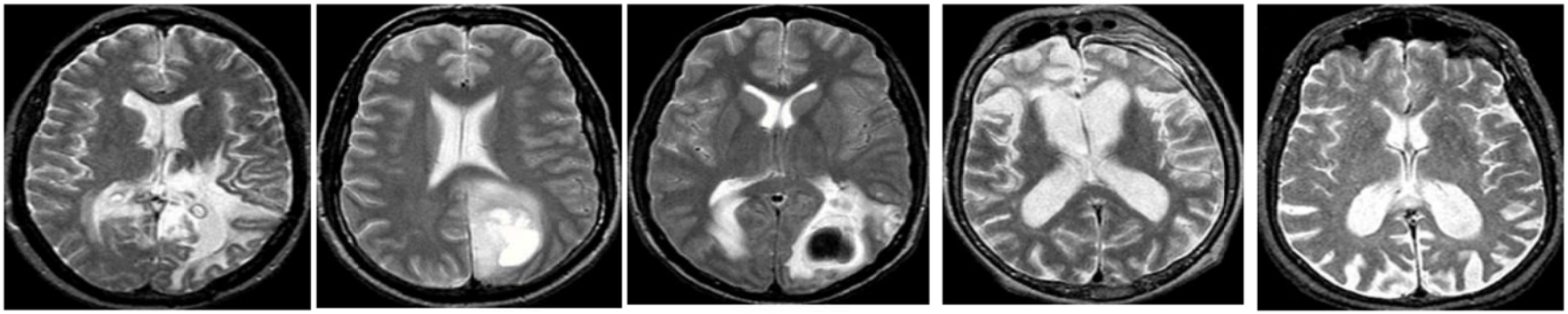


Image Captioning

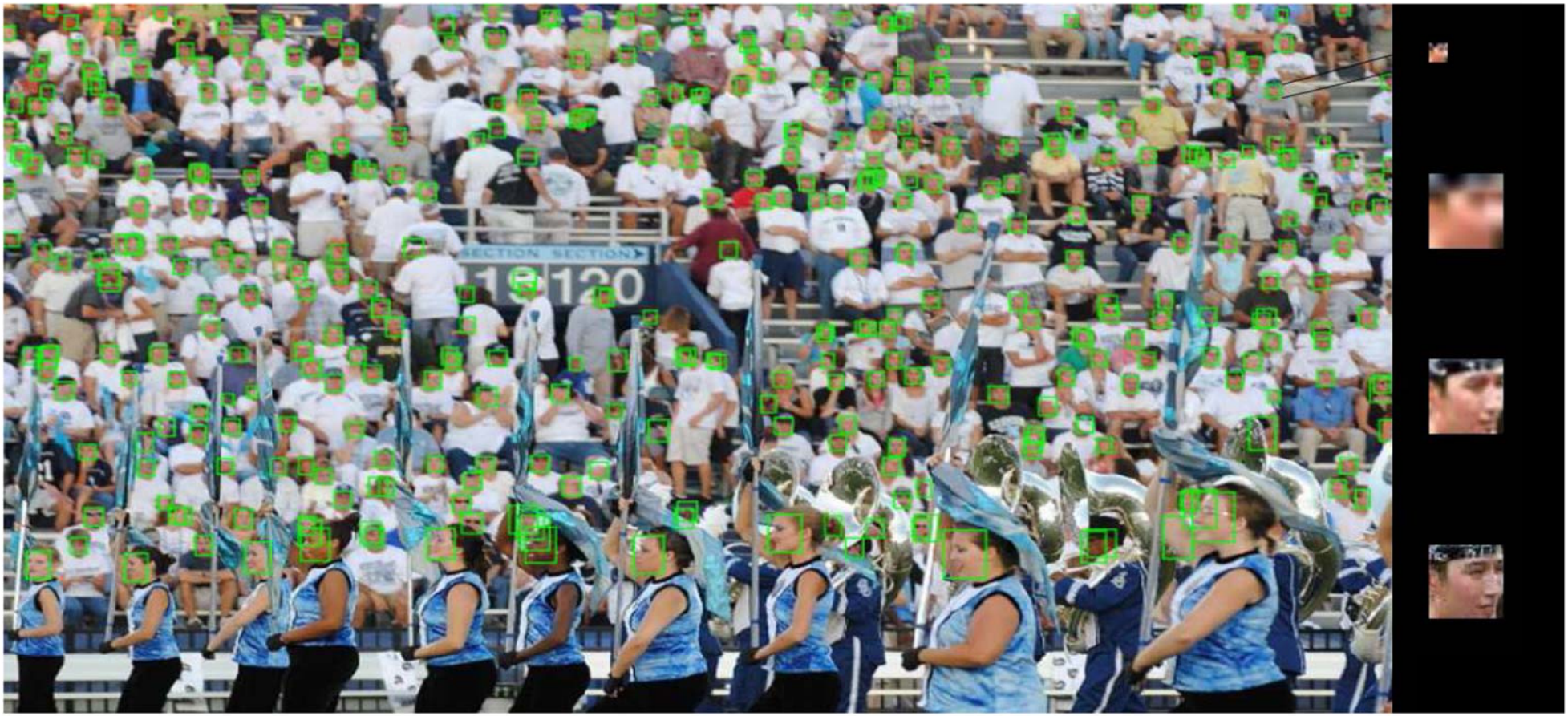
"man in black shirt is playing guitar."

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

"two young girls are playing with lego toy."

Computer Vision: Challenge

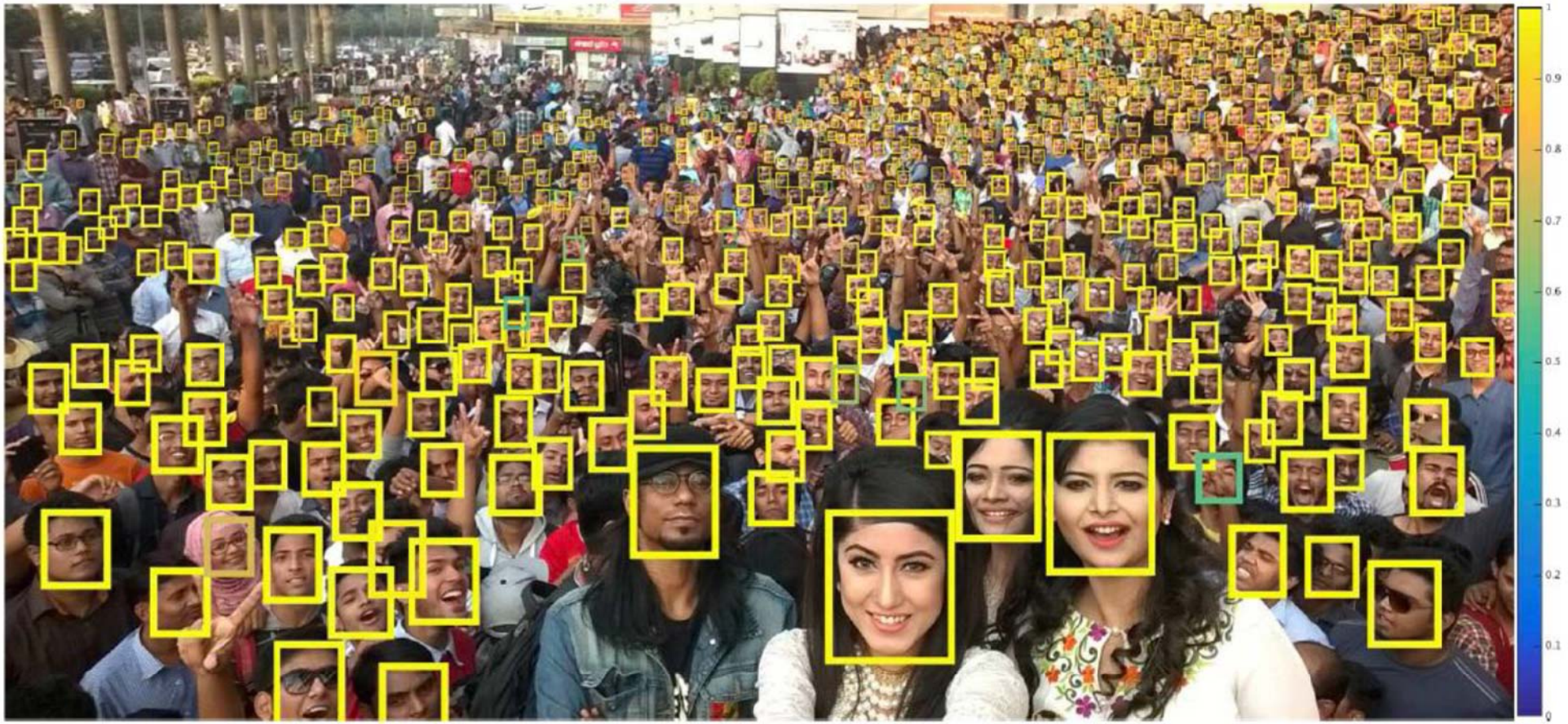
Finding Tiny Faces in the Wild



Bai et al., CVPR 2018

Computer Vision: Challenge

Finding Tiny Faces in the Wild



Hu et al. (CVPR 2017) detector that can find around 800 faces out of the reportedly 1000 present

Computer Vision: Challenge

Image Captioning

LSTM: a man and a woman in a suit and tie
CNN: a black and white photo of a man and woman in a suit
GT: A man sitting next to a woman while wearing a suit.

LSTM: a cat is laying down on a bed
CNN: a polar bear is drinking water from a white bowl
GT: A white polar bear laying on top of a pool of water

LSTM: a bear is standing on a rock in a zoo
CNN: two bears are walking on a rock in the zoo
GT: two bears touching noses standing on rocks

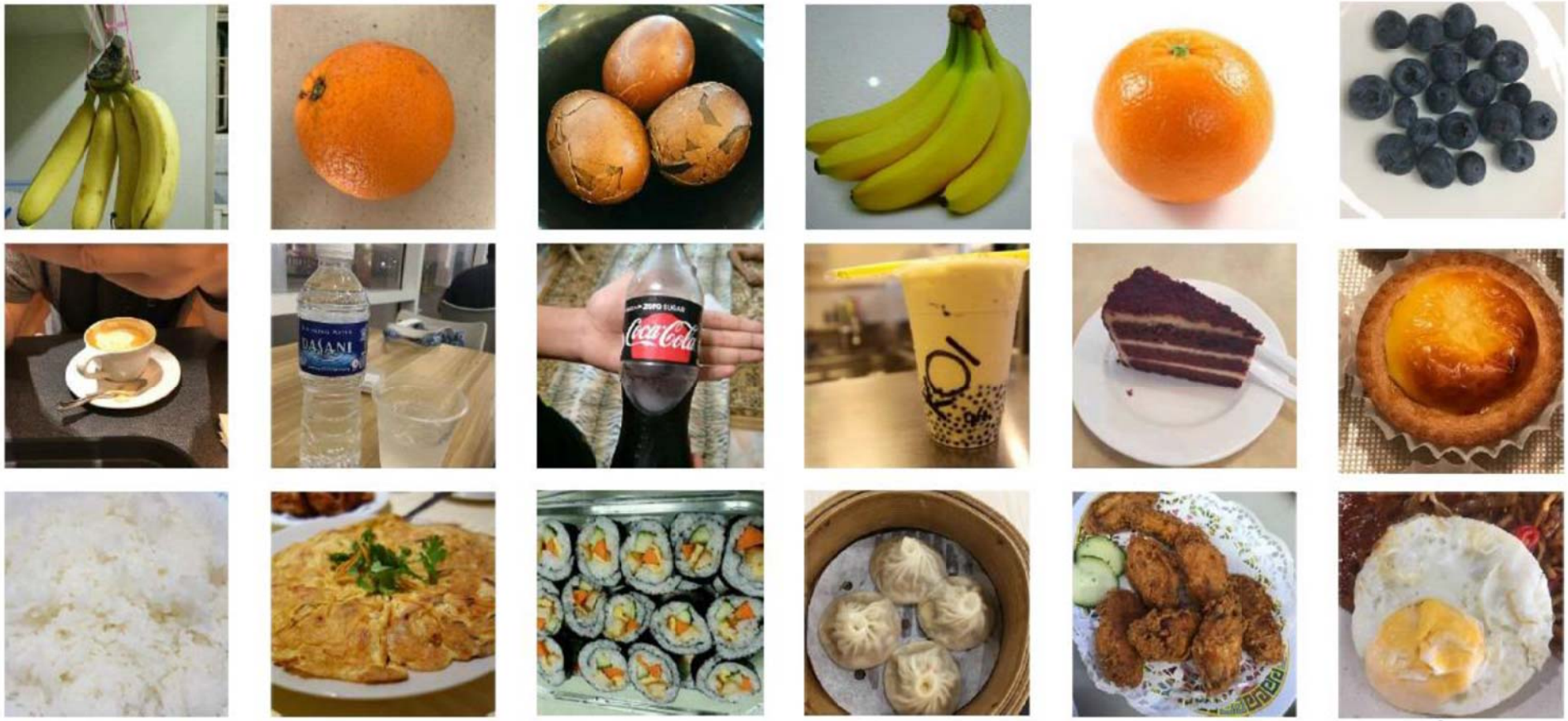
LSTM: a box of donuts with a variety of toppings
CNN: a box of doughnuts with sprinkles and a sign
GT:A bunch of doughnuts with sprinkles on them

LSTM: a dog and a dog in a field
CNN: two cows are standing in a field of grass
GT: A dog and a horse standing near each other

Aneja et al. (CVPR 2018) detector that can find around 800 faces out of the reportedly 1000 present

Computer Vision: Challenge

Food image classification: Easy cases



Computer Vision: Challenge

Food image classification: Hard cases

Large inter-class similarity (e.g., drinks)

Kopi O

Americano

Computer Vision: Challenge

Food image classification: Hard cases

Large inter-class similarity (e.g., drinks)

Computer Vision: Challenge

Food image classification: Hard cases

Incomplete Food










Computer Vision: Challenge

Food image classification: Hard cases

Poorly taken photos (illumination, rotation, occlusion, etc.)

Computer Vision: Challenge

Food image classification: Hard cases

Multiple food items










Computer Vision Tasks

- Optical character recognition (OCR)
- Machine inspection
- Retail (e.g. automated checkouts)
- 3D model building (photogrammetry)
- Medical imaging
- Automotive safety
- Match move (e.g. merging CGI with live actors in movies)
- Motion capture (mocap)
- Surveillance
- Fingerprint recognition and biometrics