



# Introduction to Compute Vision

# Today's Class

---

- ▶ What is Computer Vision?
- ▶ What are the applications of computer vision?
- ▶ Some topics.
- ▶ Why Vision+Language?



# Relationship with several nearby courses

---

- ▶ Computer Graphics: Models to Images

- ▶ Computer Vision: Images to Models

Computer vision I: Image formation,  
3D reconstruction

Computer vision II: Recognition



# Resources

---

## Online sources:

- ▶ Stanford CS231n: Convolutional Neural Networks for Visual Recognition
- ▶ Andrew Ng: Deep Learning

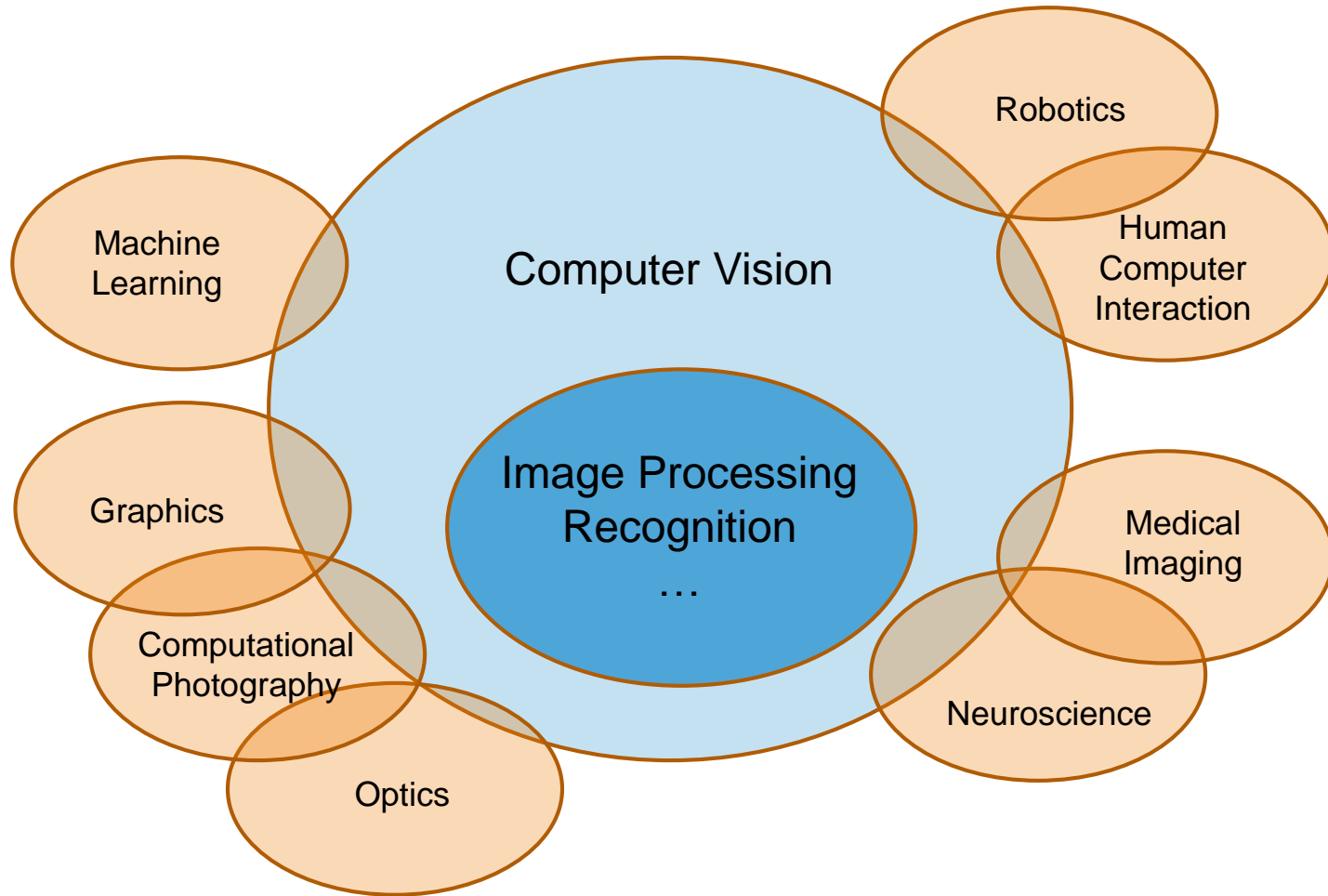
## Books:

- ▶ 1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2017
- ▶ 2. Richard Szeliski, CV: Computer Vision: Algorithms and Applications, 2010
- ▶ 3. [中文版]计算机视觉——算法与应用, 艾海舟[译], 清华大学出版社, 2012



# Relationship with several nearby Research Area

---



# What is Computer Vision?

---

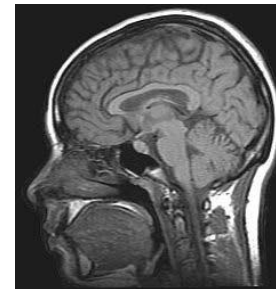
- ▶ What are examples of computer vision being used in the world?



# What is Computer Vision?

---

- ▶ **Computer vision** is the science and technology of machines that see.
- ▶ Concerned with the theory for building artificial systems that obtain information from images.
- ▶ The image data can take many forms, such as a video sequence, depth images, views from multiple cameras, or multi-dimensional data from a medical scanner



# Computer Vision

---

Make computers understand images and video.

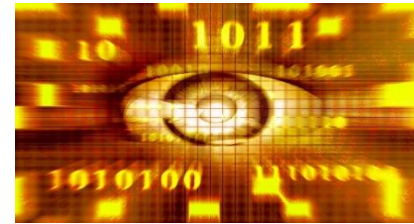


What kind of scene?

Where are the cars?

How far is the building?

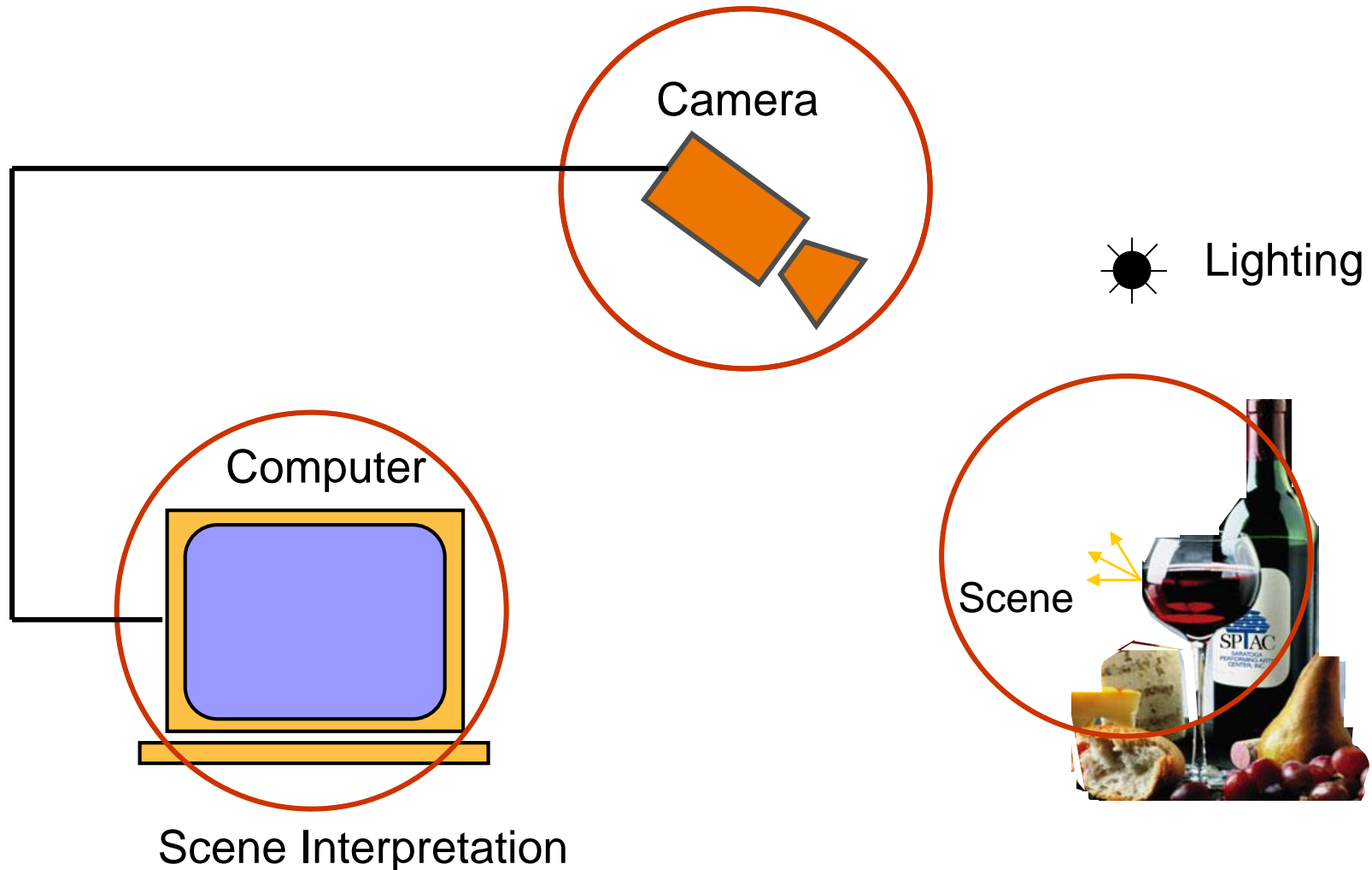
...





# Components of a computer vision system

---



# Computer vision vs human vision

---



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

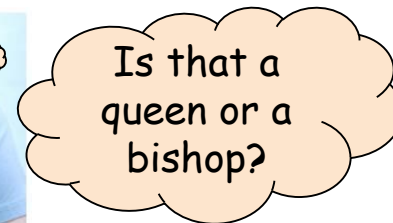
What a computer sees



# Vision is really hard

---

- ▶ Vision is an amazing feat of natural intelligence
  - ▶ Visual cortex occupies about 50% of the brain
  - ▶ More human brain devoted to vision than anything else

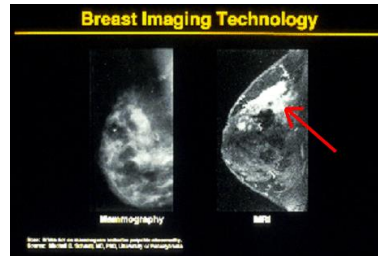


# Why computer vision matters

---



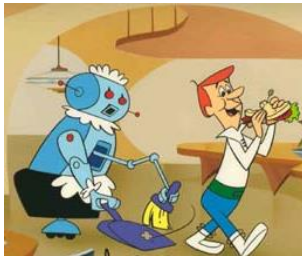
Safety



Health



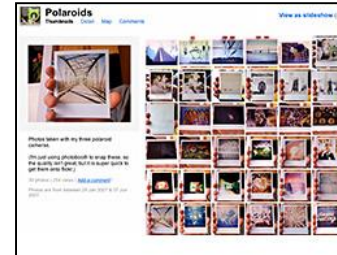
Security



Comfort



Fun



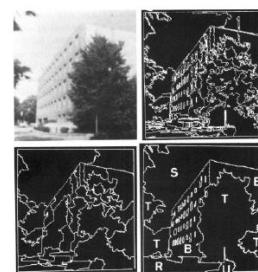
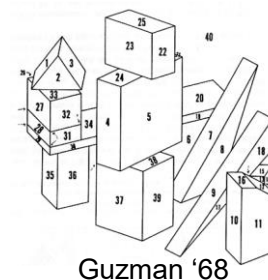
Access



# Brief history of computer vision

---

- ▶ 1966: Minsky assigns computer vision as an undergrad summer project
- ▶ 1960's: interpretation of synthetic worlds
- ▶ 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 greatly boosts the performance of computer vision.



# How Computer Vision is used now

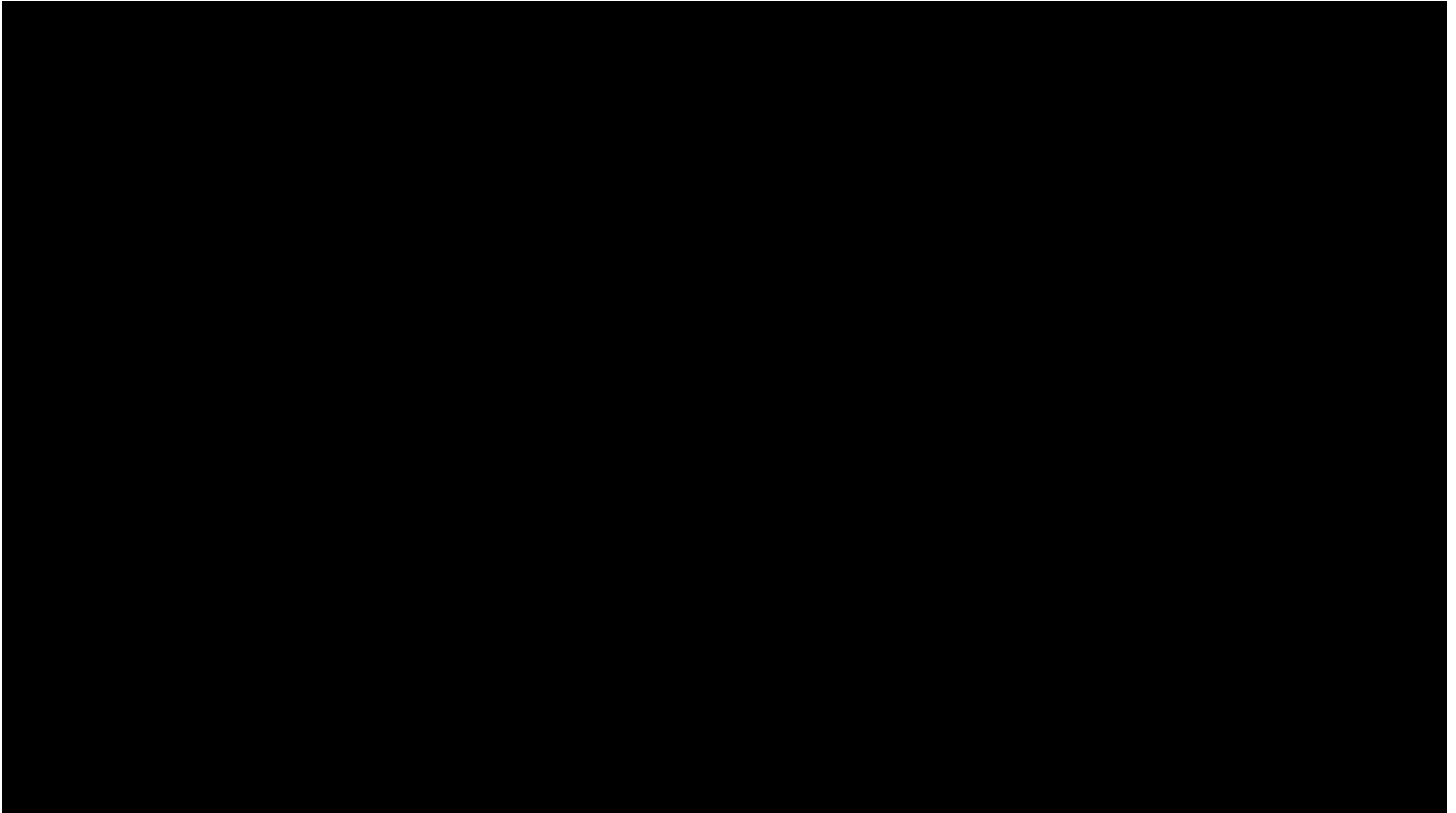
---

- ▶ Optical character recognition
- ▶ Face detection
- ▶ Image editing
- ▶ Medical image analysis
- ▶ . . .
- ▶ [https://www.bilibili.com/video/BV19a411t7vG/?vd\\_source=8275a2758ee7d2cac0897efeb47abf25](https://www.bilibili.com/video/BV19a411t7vG/?vd_source=8275a2758ee7d2cac0897efeb47abf25)



# How Computer Vision is used now

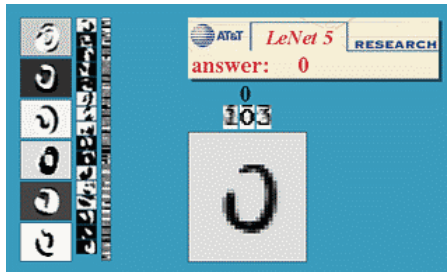
---



# Optical character recognition (OCR)

Technology to convert scanned docs to text

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



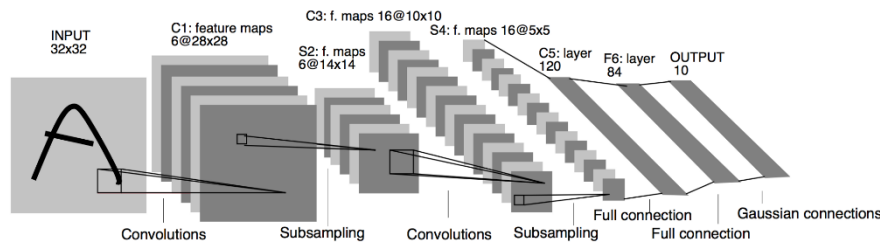
Digit recognition, AT&T labs

<http://www.research.att.com/~yann/>



License plate readers

[http://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition](http://en.wikipedia.org/wiki/Automatic_number_plate_recognition)



LeNet-5 [1998, [paper](#) by LeCun et al.]



# Face detection

---



- ▶ Many new digital cameras now detect faces
  - ▶ Canon, Sony, Fuji, ...

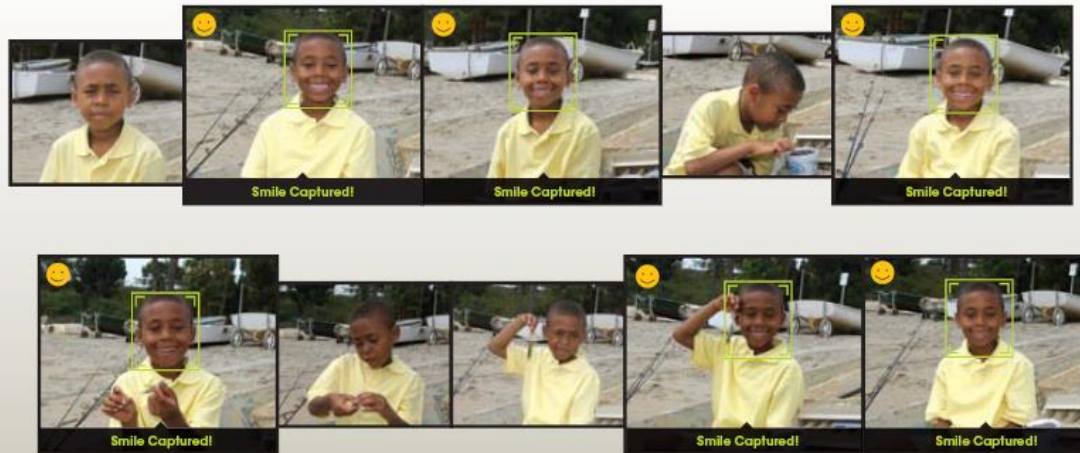


# Smile detection

---

## 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

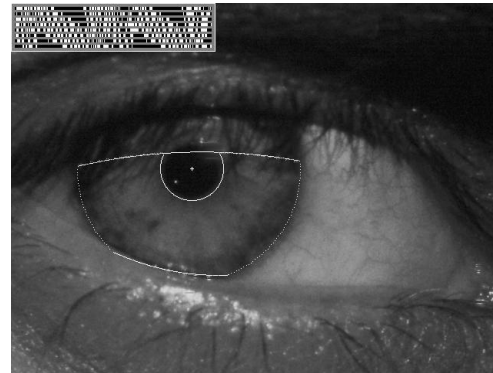
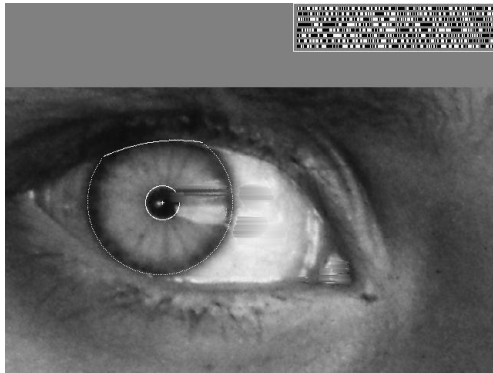


# Vision-based biometrics

---



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



# Login without a password...

---



Fingerprint scanners on many new laptops, other devices



Face recognition systems now beginning to appear more widely  
<http://www.sensiblevision.com/>



# Object recognition (in mobile phones)

---



Point & Find, Nokia  
Google Goggles

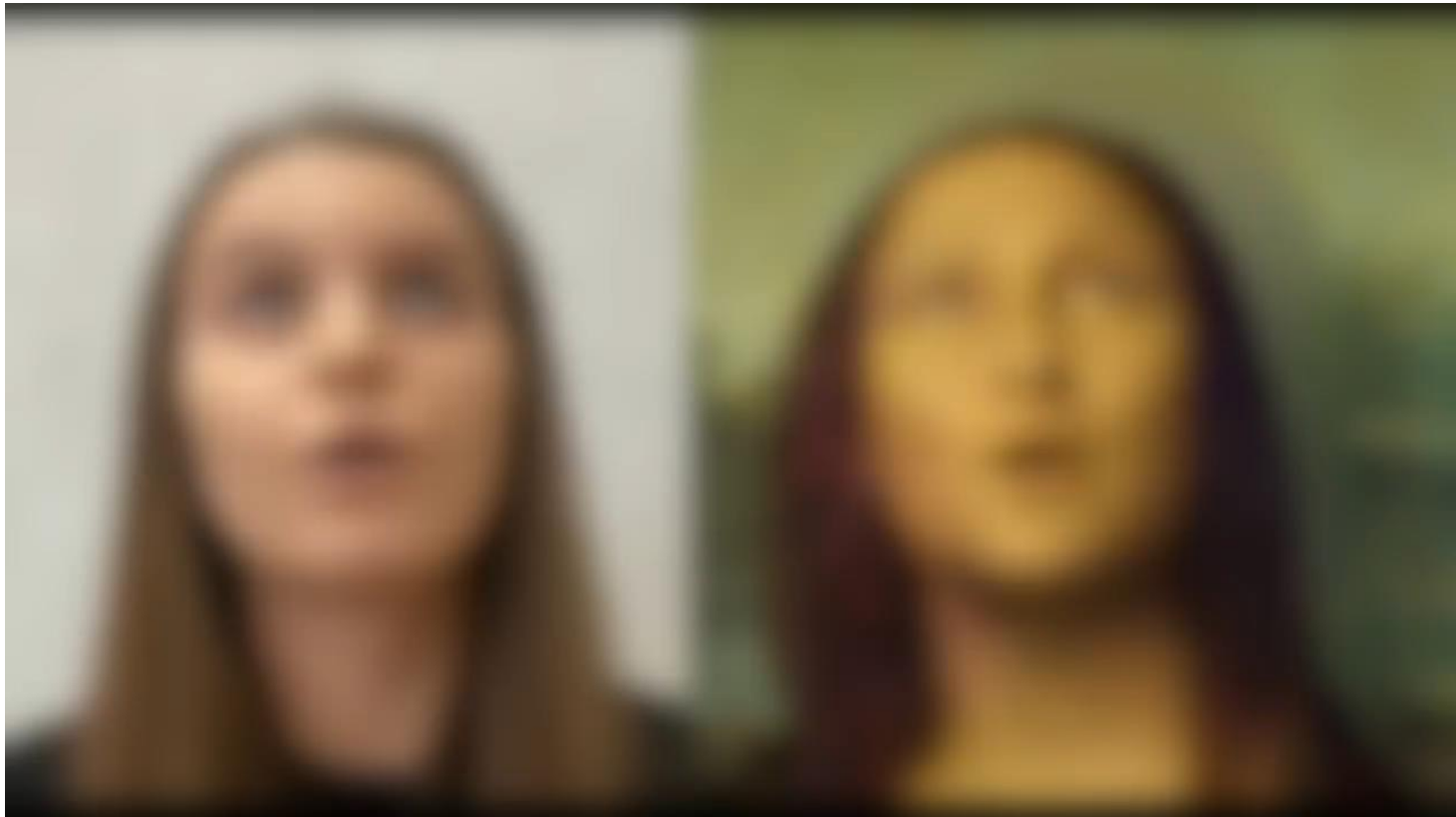


# Face expression transfer

---

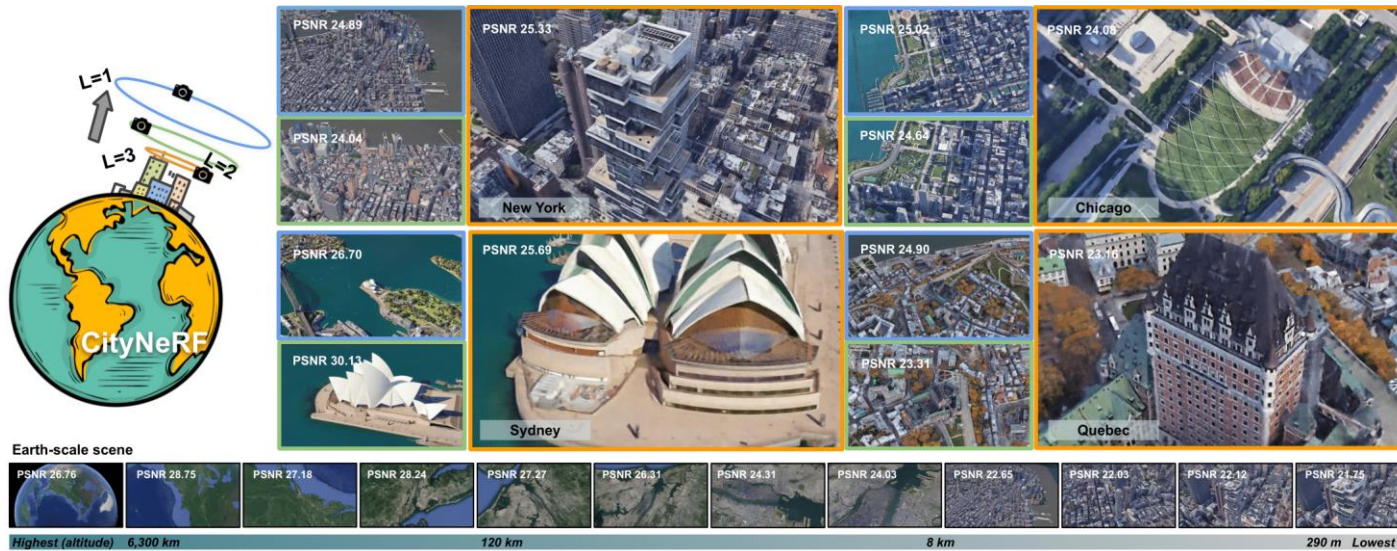
## Megapixel DeepFakes

<https://www.youtube.com/watch?v=JkUF40kPV4M>





# City Scale 3D Modeling



<https://www.youtube.com/watch?v=TaVOTvN8CpM>

-----

-----





# Special effects: shape capture

---



*The Matrix* movies, ESC Entertainment, XYZRGB, NRC



# Special effects: motion capture

---



*Pirates of the Caribbean*, Industrial Light and Magic



# Sports

---



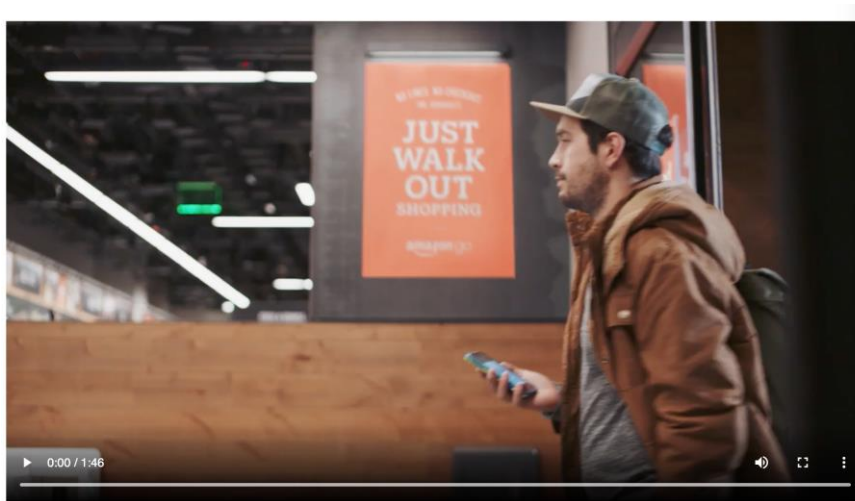
*Sportvision* first down line  
Nice [explanation](#) on [www.howstuffworks.com](http://www.howstuffworks.com)



# Case Study: Unmanned supermarket

---

- ▶ **Amazon Go** is a new kind of store featuring the world's most advanced shopping technology. No lines, no checkout – just grab and go! (Stage 3.0)



Ready to shop? Download the Amazon Go app ahead of time.

To enter Amazon Go and Amazon Go Grocery, open the app and hold the key on your phone to the gate's scanner.

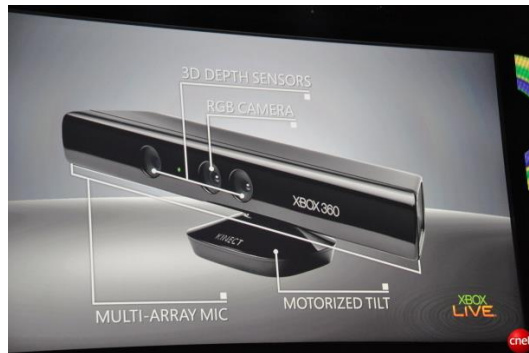
<https://www.amazon.com/b?ie=UTF8&node=16008589011>



# 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>



# Horizon Workrooms

---



<https://www.youtube.com/watch?v=lgj50IxRrKQ>  
<https://www.youtube.com/watch?v=Rncz85tVt5I>



# Horizon Workrooms

---





# Industrial robots

---



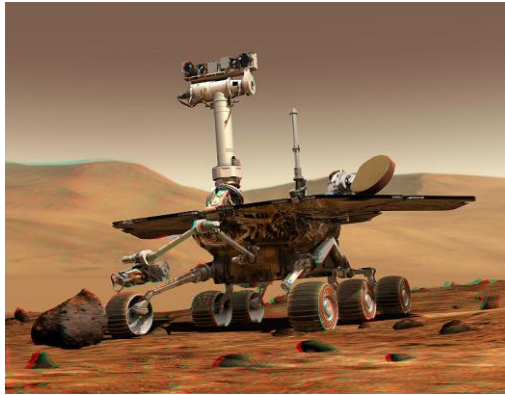
Vision-guided robots position nut runners on wheels





# Mobile robots

---



NASA's Mars Spirit Rover

[http://en.wikipedia.org/wiki/Spirit\\_rover](http://en.wikipedia.org/wiki/Spirit_rover)



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



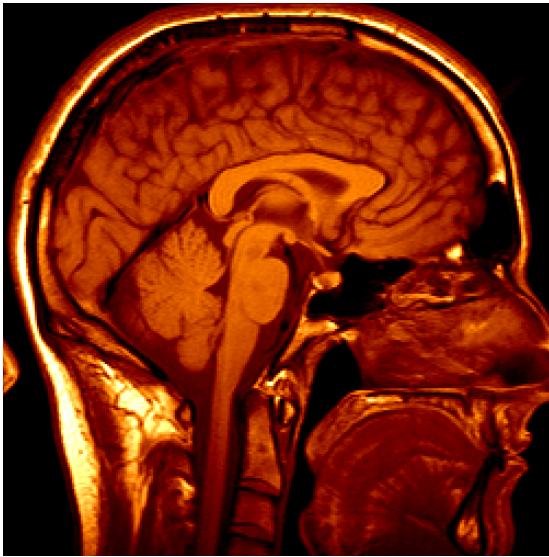
Saxena et al. 2008

[STAIR](#) at Stanford



# Medical imaging

---



3D imaging  
MRI, CT



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

# Case Study: self-driving cars

---



- ▶ Google Driverless Car
- ▶ Tesla Model S
- ▶ Minieye
- ▶ [https://www.youtube.com/watch?v=PRg5RNU\\_JLk](https://www.youtube.com/watch?v=PRg5RNU_JLk)



# Case Study: self-driving cars

---



# Topics

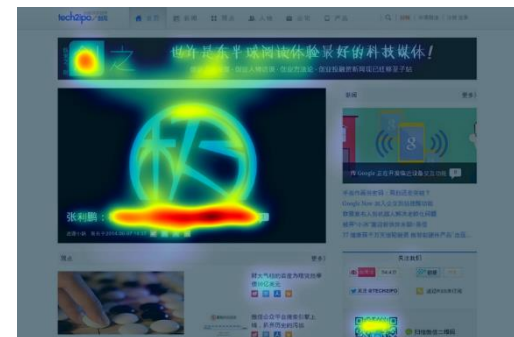
---

- ▶ **saliency detection**
  - ▶ **unsupervised saliency detection/salient object discovery)**
- ▶ **segmentation**
  - ▶ **(unsupervised image segmentation, semantic segmentation)**
- ▶ **object detection**
  - ▶ **(face detection/general object detection, supervised methods)**
- ▶ **object recognition**
  - ▶ **(face recognition/general image classification, (un)supervised methods)**
- ▶ **image generation**
  - ▶ **Generative adversarial network, style transfer, diffusion**
- ▶ **video processing**
  - ▶ **(tracking/event classification(supervised))**
- ▶ ...



# Saliency detection

---



# Application of computer vision

---

## Content based image resizing (Image Retargeting)

<https://www.youtube.com/watch?v=6NclJXTlugc>



Website Design

Image/Video Compression

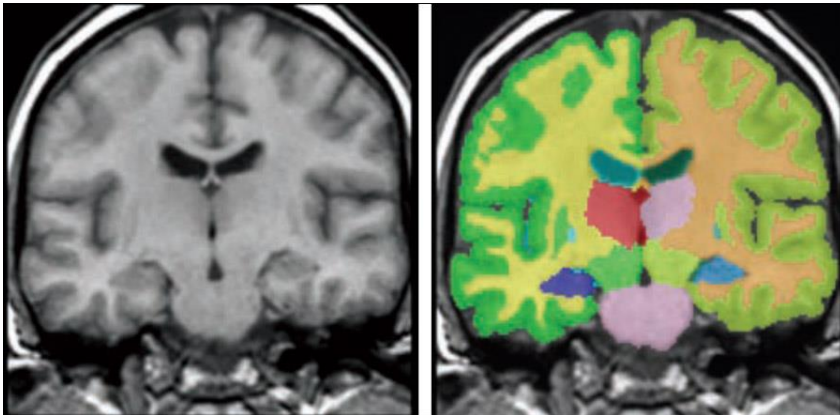
---





# Image Segmentation

---



(a) Color Labels (ACA)



(b) Texture Classes



(c) Crude Segmentation



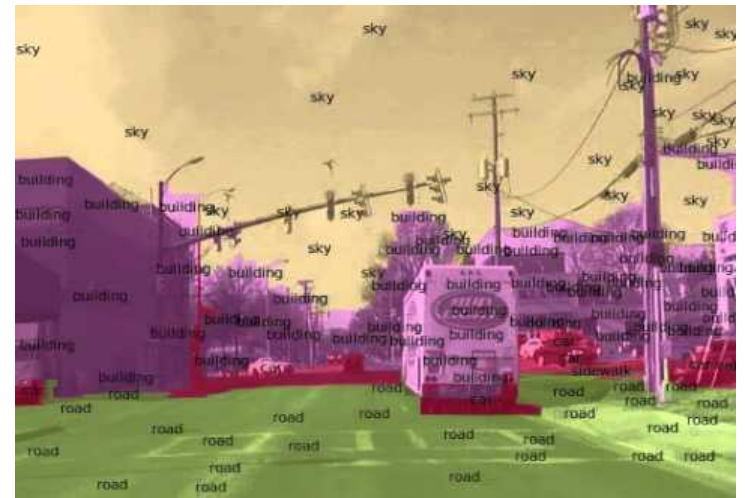
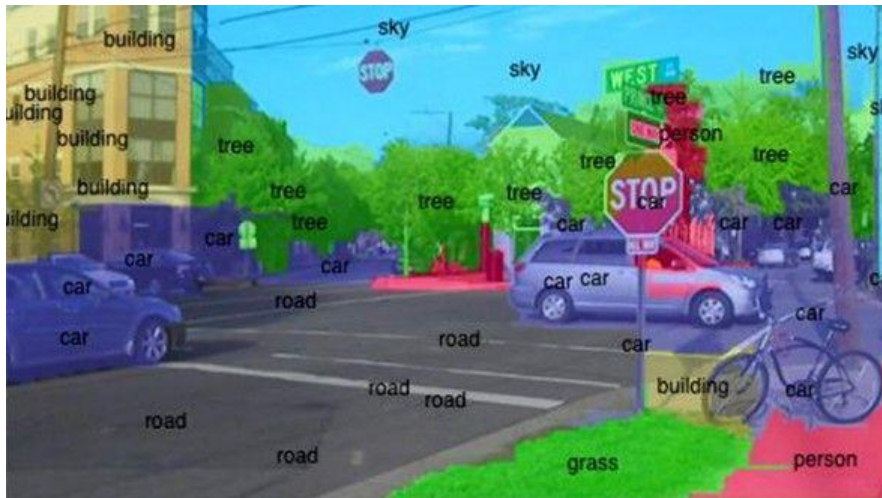
(d) Final Segmentation





# Semantic segmentation

---



# Application

---

Medical Image Analysis: MRI/CT/PET

Image retrieval

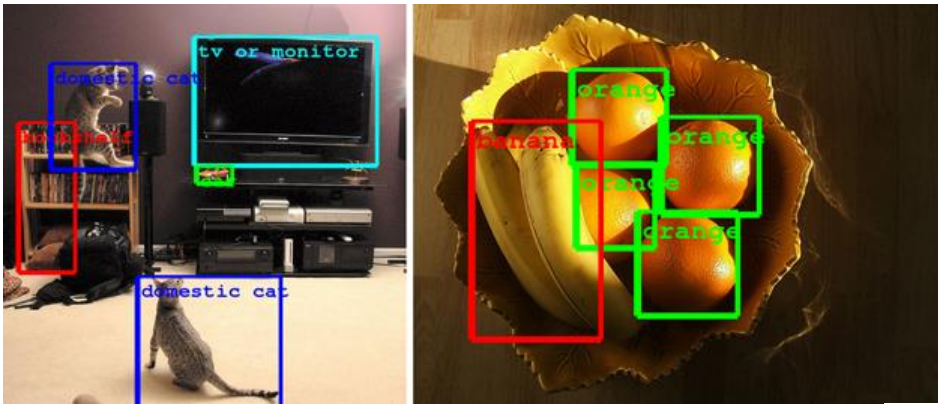
Finding Images by Sketching





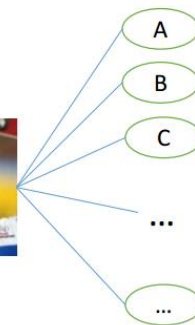
# Object Recognition

---



## Face recognition

Face identification



Multiclass  
classification



# Object Identification

---



Query



Database Images

Face verification



Same person or not.

Binary Result

# Generative Adversarial Network

---

- ▶ Face generation
- ▶ Style transfer
- ▶ Scene generation
- ▶ <https://www.youtube.com/watch?v=OGGjXG562WU>



# Generative Adversarial Network

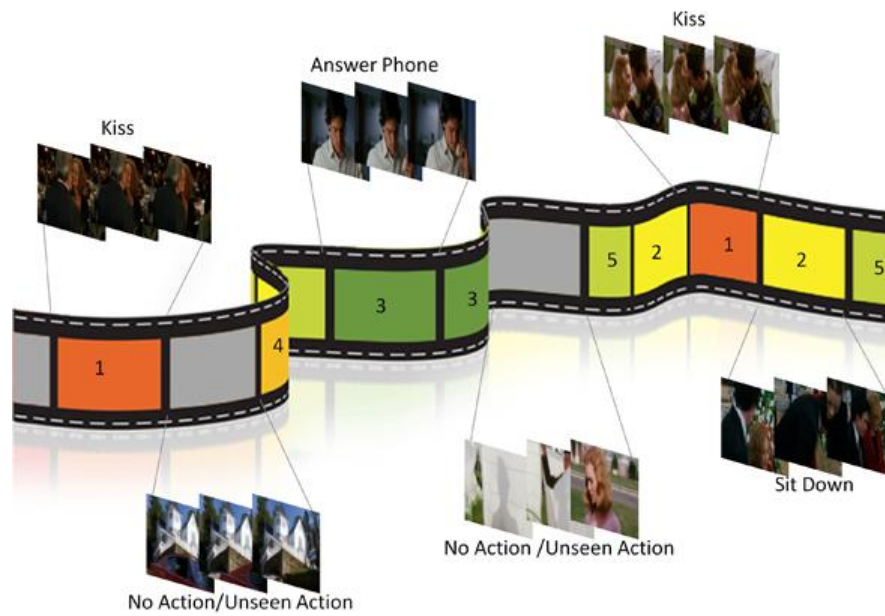
---





# Video Classification

---





# References

---

- ▶ 1. Computer Vision: A Modern Approach (2nd Edition) , David A. Forsyth , University of California, Berkeley Jean Ponce
  - ▶ 2. Computer Vision: Algorithms and Applications (Texts in Computer Science) , by [Richard Szeliski](#) (Author), Springer, 2010
  - ▶ 3. Hartley and Zisserman, [Multiple View Geometry in Computer Vision](#), Cambridge University Press, 2004
  - ▶ 4. [Stephen E. Palmer](#), Vision Science: Photons to Phenomenology, MIT Press, 1999
  - ▶ 5. Koller and Friedman, [Probabilistic Graphical Models: Principles and Techniques](#), MIT Press, 2009
- 



# References

---

## Online sources:

- ▶ Stanford CS231n: Convolutional Neural Networks for Visual Recognition
- ▶ Andrew Ng: Deep Learning

## Books:

- ▶ 1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2017
- ▶ 2. Richard Szeliski, CV: Computer Vision: Algorithms and Applications, 2010
- ▶ 3. [中文版]计算机视觉——算法与应用, 艾海舟[译], 清华大学出版社, 2012

